Molecular mechanisms of virus assembly – how new technology shapes our understanding of infectious agents

Presented by

Roman Tůma, Ph.D.

Faculty of Science

University of South Bohemia

České Budějovice

Czech Republic

and

Astbury Centre for Structural Molecular Biology

Faculty of Biological Science

University of Leeds

United Kingdom

September 2019

Acknowledgements

The results presented here would not be possible without the contribution of numerous lab members, colleagues and collaborators with whom it was a joy to work, share ideas and discuss results. This volume is dedicated to my mentors, doc. Peter Mojzeš, Prof. Josef Štěpánek, Prof. George Thomas, Jr., Prof. Peter Prevelige and Prof. Dennis Bamford, who guided me through early stages of my scientific career.

Introduction

In science, discoveries come in two basic flavors, on one hand those serendipitous, seemingly appearing with little assistance from new technologies. On the other hand, there are those which result from step changes in our analytical tools and methodologies. The first class is epitomized by fundamental theories and major discoveries such as that of penicillin while the second kind may be less obvious but equally, if not more, important for the advancement of science. The latter approach often takes advantage of advance in one branch of science or technology and finds a new application in another discipline. The now classical example is structural biology which is firmly rooted in physics (X-ray and electron diffraction, electron optics) and takes advantages of advances in computer technology (new ways to interpret and handle extensive data sets) and semiconductor manufacturing (new, more sensitive detectors for X-rays and electrons, more computer memory and faster processors). Biophysics is likewise advancing thanks to a wide variety of new technologies, such as pulsed lasers, fast and sensitive detectors, precision micromanipulation and manufacturing and of course the underpinning ever increasing computational power and speed of electronics. But unlike structural biology, which usually considers individual systems, e.g. particular enzyme or a virus, biophysics is more concerned with general principles which are behind biological processes, e.g. what are the general rules of self-assembly, protein folding or what might be the fundamental limitations of reaction rates in biological systems (Katen and Zlotnick, 2009). Here, we will limit our discourse to virus assembly process, which provides enough complexity from chemical as well as physical perspective, yet remains tractable both experimentally as well as computationally (Moisant et al., 2010). Virus assembly is also important from medical and agricultural point of view since it enables us to understand life cycles of major human, animal and crop pathogens, e.g. hepatitis viruses, bluetongue virus and potyviruses.

Viruses are ultimate cellular parasites, relying on the host cell machinery for reproduction. Biochemically they differ in composition and are broadly classified into DNA and RNA viruses according to the type of their genetic material. These classes are further subdivided into viruses with dsDNA and ssDNA genomes and those with positive sense (+)ssRNA, negative sense (-)RNA and dsRNA genomes. The RNA genomes can consist of one or more molecules, which are called segments. The genomes of dsDNA viruses can be large, consisting from tens to several hundred kilobase pairs. On the other hand, RNA genomes or segments are generally smaller, ranging from less than a kilobase to about 20 kb (Knipe and Howley, 2013).

Since the coding capacity of viruses is limited their capsids consist of multiple copies of few types of subunits (Crick and Watson, 1956). In the simplest case just one type of subunit is sufficient to build a capsid. This leads to capsids of two basic shapes: spherical and rod-like. The repetitive nature of these structures often leads to symmetrical arrangements, either of icosahedral or helical type (Klug and Caspar, 1960). However, not all viruses are ordered and some viruses, albeit exhibiting a roughly spherical or rod-like shape, are highly disordered and exhibit pleomorphic capsids. An example of this class is the core of human immunodeficiency virus (Knipe and Howley, 2013). Most enveloped viruses, which exhibit outer lipid bilayer, also lack overall symmetry due to the fluidity and deformability of the membrane. The “infamous”

influenza virus falls within this category (Knipe and Howley, 2013). However, many enveloped viruses still possess symmetrical core within which the genome is packaged. From early characterization by electron microscopy (EM) it became obvious that chemically identical subunits in larger viruses are in different structural environments, thus must adopt slightly different conformations based on the assembly context. These so called quasi equivalent conformations were originally proposed to result from slight deformations of the subunits (Caspar and Klug, 1962). Later it became clear that they often result from alternative folding of domains within capsid proteins (Tuma et al., 1996a). Because the intrinsic propensity of capsid proteins to assemble or aggregate they exhibit low solubility and hence are not amenable to structural studies. Hence, alternative biophysical techniques need to be used to characterize structure of assembly precursors and early intermediates and to follow conformational changes during assembly.

Initial structural characterization of viruses and the apparent high symmetry led to a notion that it might be possible to assemble simple capsids *in vitro* using purified subunits and nucleic acids (e.g. RNA). This had been subsequently proved experimentally for simple RNA viruses which readily self-assemble in the test tube (Bancroft et al., 1967; Bancroft et al., 1968). Later, other *in vitro* systems have been developed to follow assembly of more complex dsDNA viruses (Prevelige and King, 1993; Prevelige et al., 1993). This paved way for study of the assembly process in detail using biophysical and structural techniques. However, the challenge remains to characterize early assembly intermediates, which play key role in the control and regulation of the process. At steady state, they are found in low abundance and thus time-resolved, pre-steady state techniques are necessary. For example, the mechanism of specific recognition of cognate RNA by subunits of small ssRNA viruses remained elusive until recently when single molecule techniques enabled to examine assembly reactions at low RNA concentrations with exquisite sensitivity and under pre-steady state conditions.

Here I summarize methodological advances over the past two decades which we (myself with numerous colleagues) have contributed to the field of virus assembly. We will first consider *in vitro* assembly of simple dsDNA viruses and then continue with the more complex dsRNA viruses and finish with the role of RNA in the assembly of various RNA viruses.

Virus assembly and protein folding

The principle of quasi equivalency requires subunits to acquire different conformations according to the structural context within the growing capsid. An immediate question arises as to what conformation the subunit exists prior to assembly. Is it one of the conformations observed within the assembled capsid or a mixture of all the conformations required for assembly, or yet another conformation which is not seen in the capsid at all? Is the precursor subunit fully folded or does it represent a folding intermediate or even an ensemble of partially folded states? To answer these questions, it is necessary to structurally and thermodynamically characterize the unassembled state of the subunit. However, in most but few cases it is impossible to crystallize the subunit in such a state since it has tendency to assemble or aggregate at higher concentrations that are required during crystallization. In

addition, it is quite likely that the subunit may adopt multiple discrete or even a continuous ensemble of partially folded conformational states, a situation incompatible with most high-resolution techniques. However, similar set of problems has been encountered when studying protein folding and optical spectroscopic techniques, such as circular dichroism and fluorescence spectroscopy, were found useful in delineating folding intermediates. These techniques enable to monitor formation of secondary and tertiary structure with millisecond temporal resolution but lack high spatial resolution details. In similar vein Raman spectroscopy (RS) can be used to detect conformational changes, both secondary and tertiary, albeit with time resolution of minutes (Tuma, 2005). Raman difference spectroscopy is particularly sensitive and minute, few percent changes in secondary structure content can be detected (Tuma and Thomas, 1997). The sensitivity and quantitative aspects of RS in determining secondary structure have been exploited to follow conformational changes during virus assembly as illustrated in several earlier publications considering assembly of dsDNA bacteriophages (Tuma et al., 1996a; Tuma et al., 1998; Tuma et al., 1996b). The main finding was that assembly is coupled to folding and stabilization of small helical domains, both within the capsid coat protein as well as in the so-called scaffolding helper protein.

Folding of a polypeptide chain, especially formation of tertiary and quaternary contacts leads to the exclusion of aqueous solvent from the interior of the protein and from subunit interfaces. This has a profound effect on the isotope exchange of peptide NH groups which are pseudo-acidic and become readily deuterated (within milliseconds) in unfolded form when exposed to deuterated water at room temperature and neutral pH (Woodward et al., 2004). On the other hand, NH groups that are buried within the folded protein core exchange extremely slowly, sometimes taking weeks to months to complete the reaction (Li and Woodward, 1999). Thus, hydrogen-deuterium exchange (HDX) kinetics is a good indicator of degree of folding. For smaller proteins HDX can be conveniently measured with residue resolution by nuclear magnetic resonance, however this technique is unusable for larger assemblies like viruses. This limitation can be overcome by RS since it follows vibrational energy levels that are sensitive to atomic masses. In particular, collective vibrational modes associated with the peptide linkage, the so-called Amide I, II and III modes, encompass the NH group and thus are sensitive to deuteration (Thomas and Agard, 1984). While RS lacks residue resolution it enables to selectively follow exchange of different types of secondary structure. We have exploited this feature and applied RS and HDX to characterize folding of viral coat protein [paper 1 in the appendix] and virus assembly intermediates in dsDNA bacteriophage P22 [2]. The somewhat unexpected result was that the unassembled subunit is partially unfolded, i.e. exchanging rather rapidly [1], but a surprising discovery was that folding of a substantial portion of the polypeptide chain continues within the assembled, icosahedral capsid precursor (the procapsid) [2]. In contrast, the highly alpha helical nucleocapsid shell of dsRNA bacteriophage phi6 did not exhibit any substantial folding upon assembly (Tuma et al., 1999) suggesting that this type of behavior may be specific to a particular type of coat protein fold which undergoes further maturation and capsid stabilization.

In the mid 90ties advances in protein mass spectrometry (MS) led to the development of a new methodology for mapping the exchanging sites onto the polypeptide sequence (Zhang and Smith, 1993). Hence, we have adopted this approach, albeit using the then available

MALDI (matrix assisted laser desorption ionization) TOF (time of flight) instrumentation that offered the necessary monoisotopic (1 Da) resolution and applied it to follow folding of the polypeptide chain within the assembled P22 procapsid and capsid [3]. Using this method we could then refine RS results and map the structural changes and folding to a hinge which is flexible and protease sensitive in the procapsid (Lanman et al., 1999) but becomes folded and buried within the mature capsid.

Later, HDX-MS became a useful tool to map subunit interfaces [4] in more complex dsRNA bacteriophages of the *Cystoviridae* family, in which four proteins (major core structural protein P1, RNA polymerase P2, hexameric packaging motor P4 ATPase and assembly factor P7) first assemble into an empty polymerase core (also called a procapsid, PC) and then package three segments of RNA into the PC [5]. The P4 hexamer associates with the PC using C-terminal facet and this interaction also controls opening of the hexameric ring which has been shown to mediate RNA loading into the central cavity for processive translocation [4]. HDX can also be used to monitor conformational dynamics during the ATPase action of the packaging motor (Lisal et al., 2005). This and other developments together with basic principles of HDX are summarized in a review [6]. Further progress in molecular dynamics simulations allowed to interpret and quantitatively predict HDX from known high-resolution structures (Radou et al., 2014), providing a tool to validate structural models and assist high-throughput structure profiling.

Detection of assembly intermediates and assembly of complex viruses

Assembly of viruses was thought to be controlled by a unique and well-defined nucleation event early along the assembly pathway. This concept has been adopted from the well characterized assembly process of cellular filaments such as actin or microtubules (Prevelige et al., 1993). Given that formation of the nucleation complex is the rate limiting step (akin the transition state in catalysis) this species is only marginally populated and so are also other early assembly intermediates. This poses difficulty in delineating assembly pathways even for the simplest viruses. In few favorable cases the nature of the nucleation complex (e.g. number of coat protein subunits involved) can be inferred from the assembly kinetics (Prevelige et al., 1993). However, this approach involves numerous assumptions, for example that the nucleation complex is well defined and homogeneous. Also, the apparent lag phase in the assembly kinetics may not necessarily be linked to the nucleation event but is a direct result of the complex, multistep assembly reaction (Endres and Zlotnick, 2002). As shown by subsequent simulations of assembly reactions, many of these assumptions do not hold and heterologous nucleation complexes and multiple paths are certainly possible (Moisant et al., 2010). To test these valuable theoretical predictions, there is a need for sensitive detection of early intermediates.

Light scattering (often measured as increased turbidity), which monitors increase in weight average mass, has served as a method of choice for kinetic studies of virus assembly (Prevelige et al., 1993; Zlotnick et al., 1999). However, it lacks structural resolution. Other techniques that offer such resolution, such as rapid quench and staining electron microscopy, may be

used to explore the types of complexes present in the reaction but do not provide quantitative kinetic information. Advanced synchrotron X-ray sources allow to follow small angle X-ray scattering (SAXS) with millisecond time resolution. In relatively dilute solutions SAXS yields rotationally and ensemble averaged X-ray diffraction from individual particles, thus providing low resolution structural (shape) information about the dominant assemblies and enables to follow kinetics at the same time. We have applied this approach together with genetic engineering of viral proteins to follow assembly of bacteriophage P22 with 5 to 10 sec temporal resolution [7]. Co-assembly of coat protein with various oligomers and concentrations of the scaffolding helper protein revealed heterologous nucleation with radically different outcomes. The wild type weakly dimeric scaffolding protein favored a limited nucleation pathway followed by rapid growth with low population of intermediates. On the other hand, scaffolding covalent dimer (disulfide crosslinked mutant) caused over-nucleation and rapid population of partial viral shells, which failed to complete due to limited amount of the coat protein in the reaction (i.e. due to kinetic trapping). Thus, this confirmed that viral assembly pathways may be optimized in kinetic terms by fine tuning of weak interactions between subunits (Katen and Zlotnick, 2009). The idea of kinetic trapping is now explored for the design of novel antiviral drugs (Zlotnick and Mukhopadhyay, 2011).

Time-resolved SAXS requires relatively large quantities of reasonably well soluble coat protein and this is difficult to obtain for but a few model viruses. An alternative, albeit lower resolution technique, is dynamic light scattering (DLS) which, together with hydrodynamic modelling, enables determination of approximate shape (Sokolova et al., 2001). DLS can also analyze heterogenous mixtures and provide size distributions which is particularly useful when dealing with assembly of more complex viruses, e.g. those assembling from multiple distinct subunits. A model of such complex virus is dsRNA bacteriophage phi6 for which an *in vitro* assembly system has been developed [8] and exploited to characterize assembly intermediates by a combination of techniques, including light scattering kinetics and time resolved DLS. This study and subsequent characterization of another dsRNA bacteriophage phi8 [9] led to a proposal of a common assembly pathway for these complex viruses. Assembly is triggered by association of the major PC protein P1 with the core enzymes, the P2 polymerase and the P4 packaging hexamer. This interaction stabilizes an intermediate and assures that each PC incorporates the two essential enzymes. Since P2 is a monomer residing at two-fold edges of the PC and P4 is a hexamer residing at the five-fold vertices the early incorporation also solves the problem of the apparent symmetry mismatch in the final assembly. The lower (or even absence of) symmetry within the early intermediates is compatible with the incorporation of these enzymes at this stage.

Shedding light on RNA folding during virus assembly by single molecule fluorescence

The ensemble scattering techniques discussed above have one major shortcoming when dealing with more complex assemblies since they do not allow to selectively follow the fate of a particular subunit type. These techniques lack the necessary contrast. One way to introduce contrast is via isotope labelling and contrast variation together with small angle neutron scattering (SANS). But SANS requires high concentrations and large quantities of soluble

material. In addition, even the best neutron sources do not provide with sufficient flux to perform realistic time resolved studies. Therefore, only stationary observations are possible with much lower resolution than that offered by SAXS but still allowing e.g. spatial localization of subunits (Ikonen et al., 2003).

One way to circumvent these difficulties is to use fluorescent labels and correlation spectroscopy which is a type of fluctuation spectroscopy analogous to DLS. In this type of experiment the component of interest is labelled with a bright, synthetic fluorescent dye and introduced into the assembly reaction at low concentration (ideally sub-nanomolar, e.g. as a tracer) and then complemented with other unlabeled constituents. The reaction is then followed by a sensitive confocal microscope which is equipped by a single photon counting detectors and a correlator (either hardware or software). The correlator then yields an autocorrelation function which contains information about various processes which influence fluorescence intensity. These are related to the photo physics of the fluorescent dye (triplet state dynamics and quenching) as well as the duration of the passage through the confocal volume. The latter is related to diffusion (hydrodynamic shape) of the labeled molecules and thus can be used to follow assembly since larger species take longer to pass through the confocal volume and thus produce lasting correlations [10]. This technique also affords access to the low concentration regime which was not explored previously. The low concentration regime proved key in the study of selective packaging of ssRNA in simple RNA viruses such as those causing common cold (Human Rhinovirus) or foot and mouth disease (FMDV). Several of these simple RNA viruses were successfully reconstituted from purified components *in vitro* and they constitute one of the best characterized virus groups. But, contrary to the situation in infected cells it was found that the coat protein would package essentially any negatively charged polymer such as RNA, DNA or even polyglutamate or polysulfonate. These observations led to a widely accepted electrostatic mechanism of RNA packaging in which gradual neutralization of the RNA negative charge is responsible for encapsidation during co-assembly with the coat protein. However, this principle fails to explain the selective packaging of the viral RNA within the infected cells. While certain viruses solve this problem by physically linking replication and packaging, simple plant viruses and RNA bacteriophages seem to selectively acquire viral RNA in the cytoplasm that is otherwise full of ribosomal RNA and mRNA. The answer to this paradox came from exploration of the low concentration regime using FCS and labelled genomic RNA of the bacteriophage MS2 as a model system [11]. Somewhat counterintuitively, the viral RNA associates first with a sub-stochiometric amount of the coat protein and becomes condensed (collapsed), forming a core for further assembly. The initial collapse only happens with the cognate viral RNA and was also observed for a simple plant virus STNV (Satellite tobacco necrosis virus) [11]. From these studies it was clear that the initial collapse and specific packaging have to be mediated by multiple specific RNA-coat protein interactions which were subsequently characterized and mapped in detail for STNV [12] and other viruses including human pathogens (Patel et al., 2017; Stockley et al., 2013).

The single molecule approach was also useful in characterizing overall conformation of RNA molecules in solution. Single stranded homopolymers like poly(rA) have low persistence length and can thus be considered as gaussian coils of freely jointed or worm like chain. On the other hand, biologically relevant RNAs with high complexity sequences behave differently.

Due to stochastic base pairing such long RNAs exhibit a high degree of secondary structure with more than 60% of the chain in the duplex form. Thus, they are far from being linear and a randomly branched configuration has been proposed as the most plausible model for describing coarse grained properties, e.g. shape and hydrodynamic radius (Yoffe et al., 2008). A systematic sizing of various biologically relevant long RNAs in dilute solutions using FCS revealed a non-monotonic relation between their contour length (number of bases) and corresponding hydrodynamic size (hydrodynamic radius, R_h). This was explained by statistically accounting for the pattern of secondary structure which in turn determines the branching properties and thus the overall size [13]. This led to confirmation of the branching model and verified the new computational approach to estimate size of long RNAs from their sequence.

Another example of application of single molecule techniques to virus assembly is the investigation of an RNA chaperone sigmaNS from avian reovirus [14]. Using FCS we demonstrated that sigmaNS does not condense viral RNA but unfolds it. Unfolding is mediated by gradual binding and stabilization of single stranded RNA conformation as revealed by single molecule FRET (Forster resonance energy transfer) between labelled ends of RNA hairpin.

Conclusions

Here we have discussed and illustrated how technological advances foster paradigm changes in our understanding of virus assembly. Some techniques, such as SAXS and DLS are now well established and to some degree outdated. However, new technology is now revolutionizing electron microscopy so time-resolved structural studies using rapid cryopreservation of assembling reactions holds great promise. However, the challenges posed by the intrinsic sample heterogeneity and noise in the images may be surmounted only in conjunction with other emerging techniques, such as native mass spectrometry and laser-induced reversible cross-linking. Thus, in the future greater emphasis needs to be given to concerted application of several time-resolved experimental techniques which is underpinned by a quantitative theoretical approach and computation.

Cited references

Bancroft, J.B., Hills, G.J., and Markham, R. (1967). A study of the self-assembly process in a small spherical virus. Formation of organized structures from protein subunits *in vitro*. Virology *31*, 354-379.

Bancroft, J.B., Wagner, G.W., and Bracker, C.E. (1968). The self-assembly of a nucleic-acid free pseudo-top component for a small spherical virus. Virology *36*, 146-149.

Caspar, D.L., and Klug, A. (1962). Physical principles in the construction of regular viruses. Cold Spring Harbor symposia on quantitative biology *27*, 1-24.

Crick, F.H., and Watson, J.D. (1956). Structure of small viruses. Nature *177*, 473-475.

Endres, D., and Zlotnick, A. (2002). Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. Biophysical journal *83*, 1217-1230.

Ikonen, T., Kainov, D., Timmins, P., Serimaa, R., and Tuma, R. (2003). Locating the minor components of double-stranded RNA bacteriophage phi 6 by neutron scattering. Journal of Applied Crystallography *36*, 525-529.

Katen, S., and Zlotnick, A. (2009). The thermodynamics of virus capsid assembly. Methods in enzymology *455*, 395-417.

Klug, A., and Caspar, D.L. (1960). The structure of small viruses. Advances in virus research *7*, 225-325.

Knipe, D.M., and Howley, P.M. (2013). Fields Virology, Vol 1, 6 edn (Philadelphia: Lippincott Williams and Wilkins).

Lanman, J., Tuma, R., and Prevelige, P.E. (1999). Identification and characterization of the domain structure of bacteriophage P22 coat protein. Biochemistry *38*, 14614-14623.

Li, R., and Woodward, C. (1999). The hydrogen exchange core and protein folding. Protein Sci *8*, 1571-1590.

Lisal, J., Lam, T.T., Kainov, D.E., Emmett, M.R., Marshall, A.G., and Tuma, R. (2005). Functional visualization of viral molecular motor by hydrogen-deuterium exchange reveals transient states. Nature Structural & Molecular Biology *12*, 460-466.

Moisant, P., Neeman, H., and Zlotnick, A. (2010). Exploring the paths of (virus) assembly. Biophysical journal *99*, 1350-1357.

Patel, N., White, S.J., Thompson, R.F., Bingham, R., Weiss, E.U., Maskell, D.P., Zlotnick, A., Dykeman, E., Tuma, R., Twarock, R.*, et al.* (2017). HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. Nat Microbiol *2*, 17098.

Prevelige, P.E., Jr., and King, J. (1993). Assembly of bacteriophage P22: a model for ds-DNA virus assembly. Prog Med Virol *40*, 206-221.

Prevelige, P.E., Jr., Thomas, D., and King, J. (1993). Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. Biophys J *64*, 824-835.

Radou, G., Dreyer, F.N., Tuma, R., and Paci, E. (2014). Functional Dynamics of Hexameric Helicase Probed by Hydrogen Exchange and Simulation. Biophysical Journal *107*, 983-990.

Sokolova, A., Malfois, M., Caldentey, J., Svergun, D.I., Koch, M.H.J., Bamford, D.H., and Tuma, R. (2001). Solution structure of bacteriophage PRD1 vertex complex. Journal of Biological Chemistry *276*, 46187-46195.

Stockley, P.G., Twarock, R., Bakker, S.E., Barker, A.M., Borodavka, A., Dykeman, E., Ford, R.J., Pearson, A.R., Phillips, S.E.V., Ranson, N.A.*, et al.* (2013). Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. Journal of Biological Physics *39*, 277-287.

Thomas, G.J., Jr., and Agard, D.A. (1984). Quantitative analysis of nucleic acids, proteins, and viruses by Raman band deconvolution. Biophys J *46*, 763-768.

Tuma, R. (2005). Raman spectroscopy of proteins: from peptides to large assemblies. Journal of Raman Spectroscopy *36*, 307-319.

Tuma, R., Bamford, J.H.K., Bamford, D.H., Russell, M.P., and Thomas, G.J. (1996a). Structure, interactions and dynamics of PRD1 virus .1. Coupling of subunit folding and capsid assembly. Journal of Molecular Biology *257*, 87-101.

Tuma, R., Bamford, J.K.H., Bamford, D.H., and Thomas, G.J. (1999). Assembly dynamics of the nucleocapsid shell subunit (P8) of bacteriophage phi 6. Biochemistry *38*, 15025-15033.

Tuma, R., Parker, M.H., Weigele, P., Sampson, L., Sun, Y.H., Krishna, N.R., Casjens, S., Thomas, G.J., and Prevelige, P.E. (1998). A helical coat protein recognition domain of the bacteriophage P22 scaffolding protein. Journal of Molecular Biology *281*, 81-94.

Tuma, R., Prevelige, P.E., and Thomas, G.J. (1996b). Structural transitions in the scaffolding and coat proteins of P22 virus during assembly and disassembly. Biochemistry *35*, 4619-4627.

Tuma, R., and Thomas, G.J. (1997). Mechanisms of virus assembly probed by Raman spectroscopy: The icosahedral bacteriophage P22. Biophysical Chemistry *68*, 17-31.

Woodward, C., Carulla, N., and Barany, G. (2004). Native state hydrogen-exchange analysis of protein folding and protein motional domains. Methods Enzymol *380*, 379-400.

Yoffe, A.M., Prinsen, P., Gopal, A., Knobler, C.M., Gelbart, W.M., and Ben-Shaul, A. (2008). Predicting the sizes of large RNA molecules. Proc Natl Acad Sci U S A *105*, 16153-16158.

Zhang, Z., and Smith, D.L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. Protein Sci *2*, 522-531.

Zlotnick, A., Johnson, J.M., Wingfield, P.W., Stahl, S.J., and Endres, D. (1999). A theoretical model successfully identifies features of hepatitis B virus capsid assembly. Biochemistry *38*, 14644-14652.

Zlotnick, A., and Mukhopadhyay, S. (2011). Virus assembly, allostery and antivirals. Trends in microbiology *19*, 14-23.

List of selected papers included below and referred to in the text via numbers in square brackets []:

1. R Tuma, H Tsuruta, JM Benevides, PE Prevelige, GJ Thomas (2001) Characterization of subunit structural changes accompanying assembly of the bacteriophage P22 procapsid. Biochemistry 40, 665-674.
2. R Tuma, PE Prevelige, GJ Thomas, Jr. (1998) Mechanism of capsid maturation in a double-stranded DNA virus. Proc Natl Acad Sci USA 95, 9885-9890.
3. R Tuma, LU Coward, MC Kirk, S Barnes, PE Prevelige (2001) Hydrogen-deuterium exchange as a probe of folding and assembly in viral capsids. J Mol Biol 306, 389-396.
4. J Lísal, DE Kainov, TKT Lam, MR Emmett, H Wei, P Gottlieb, AG Marshall, (2006) Interaction of packaging motor with the polymerase complex of dsRNA bacteriophage. Virology 351, 73-79.
5. MM Poranen, R Tuma (2004) Self-assembly of double-stranded RNA bacteriophages. Virus Res 101, 93-100.
6. B Suchanova, R Tuma (2008) Folding and assembly of large macromolecular complexes monitored by hydrogen-deuterium exchange and mass spectrometry. Microbial cell factories 7, 12-32.
7. R Tuma, H Tsuruta, KH French, PE Prevelige (2008) Detection of intermediates and kinetic control during assembly of bacteriophage P22 procapsid. J Mol Biol 381, 1395-1406.
8. MM Poranen, AO Paatero, R Tuma, DH Bamford (2001) Self-assembly of a viral molecular machine from purified protein and RNA constituents. Mol Cell 7, 845-854.
9. DE Kainov, SJ Butcher, DH Bamford, R Tuma (2003) Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages. Journal of molecular biology 328, 791-804.
10. A Borodavka, R Tuma, PG Stockley (2013) A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence RNA Biology 10, 481-489.
11. A Borodavka, R Tuma, PG Stockley (2012) Evidence that viral RNAs have evolved for efficient, two-stage packaging. Proc Natl Acad Sci USA 109, 15769-15774.
12. N Patel, EC Dykeman, RHA Coutts, GP Lomonossoff, DJ Rowlands, SEV Phillips, N Ranson, R Twarock, R Tuma, PG Stockley (2015) Revealing the density of encoded functions in a viral RNA. Proc Natl Acad Sci USA 112, 2227-2232.
13. A Borodavka, SW Singaram, PG Stockley, WM Gelbart, A Ben-Shaul, R Tuma (2016) Sizes of long RNA molecules are determined by the branching patterns of their secondary structures. Biophys J 111, 2077-2085.

14. A Borodavka, J Ault, PG Stockley, R Tuma (2015) Evidence that avian reovirus sigma NS is an RNA chaperone: implications for genome segment assortment. Nucleic Acids Res 43, 7044-7057.

Characterization of Subunit Structural Changes Accompanying Assembly of the Bacteriophage P22 Procapsid†

Roman Tuma,‡,§ Hiro Tsuruta,‡,|| James M. Benevides,⊥ Peter E. Prevelige, Jr.,*,‡ and George J. Thomas, Jr.*,⊥

Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35205, and School of Biological Sciences, University of Missouri−Kansas City, Kansas City, Missouri 64110

Received August 17, 2000; Revised Manuscript Received November 16, 2000

ABSTRACT: P22 serves as a model for the assembly and maturation of icosahedral double-stranded DNA viruses. The viral capsid precursor, or *procapsid*, is assembled from 420 copies of a 47 kDa coat protein subunit (gp5) that is rich in β-strand secondary structure. Maturation to the *capsid*, which occurs in vivo upon DNA packaging, is accompanied by shell expansion and a large increase in the level of protection against deuterium exchange of amide NH groups. Accordingly, shell maturation resembles the final step in protein folding, wherein domain packing and an exchange-protected core become more fully developed [Tuma, R., Prevelige, P. E., Jr., and Thomas, G. J., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9885−9890]. Here, we exploit recent advances in Raman spectroscopy to investigate the P22 coat protein subunit under conditions which stabilize the *monomeric* state, viz., in solution at very low concentrations. Under these conditions, the monomer exhibits an elongated shape, as demonstrated by small-angle X-ray scattering. Raman spectra allow the identification of conformation-sensitive marker bands of the monomer, as well as the characterization of NH exchange dynamics for comparison with procapsid and capsid shell assemblies. We show that procapsid assembly involves significant ordering of the predominantly β-strand backbone. We propose that such ordering may mediate formation of the distinct subunit conformations required for assembly of a $T = 7$ icosahedral lattice. However, the monomer, like the subunit within the procapsid lattice, exhibits a moderate level of protection against low-temperature NH exchange, indicative of a nascent folding core. The environments and exchange characteristics of key side chains are also similar for the monomeric and procapsid subunits, and distinct from corresponding characteristics of the capsid subunit. The monomer thus represents a compact but metastable folding intermediate along the pathway to assembly of the procapsid and capsid.

Cellular functions are often performed by large macromolecular complexes, the assembly of which may involve a complex series of protein recognition events and conformational changes. Viruses serve as useful models for investigating the molecular mechanisms underlying such assembly processes. For example, the life cycle of a virus typically requires the intracellular assembly of a *capsid* from multiple copies of a few types of protein subunits. Constructing a topologically closed capsid from chemically identical or similar protein subunits requires switching between different conformations in a specific path-dependent manner (*1*). In the case of icosahedral double-stranded (ds)[1] DNA phages, the process involves the initial formation of a metastable precursor, or *procapsid*, which subsequently matures irreversibly to the stable capsid (*2*). Such a two-step mechanism, characterized by weaker intersubunit interactions in the procapsid state, may facilitate the correction of assembly errors prior to capsid maturation, stabilization, and genome packaging. Conversely, the greater stability of the mature capsid more effectively protects the viral genome from environmental insult.

Irreversible conformational changes in the capsid lattice are not limited to dsDNA bacteriophages. Recently, similar conformational changes have been observed for an insect RNA virus (*3*). Irreversible transitions also accompany proteolytic maturations in other RNA viruses (*4*). Picornaviruses employ such conformational changes to initiate disassembly and genome release upon infection (*5*−*7*). In many cases, conformational change is triggered by receptor binding or internalization (*8*). Understanding how this type of controlled plasticity can be programmed into the protein subunits of a capsid shell or precursor requires knowledge

† Supported by NIH Grant GM50776 (to G.J.T.) and NSF Grant DBI 9726698 (to P.E.P.). The SSRL facility is supported in part by the NIH NCRR (Grant RR-01209) and the DOE OBER.

* To whom correspondence may be addressed. P.E.P.: e-mail, prevelig@uab.edu; phone, (205) 975-5327; fax, (205) 975-5479. G.J.T.: e-mail, thomasgj@umkc.edu; phone, (816) 235-5247; fax, (816) 235-1503.

‡ University of Alabama at Birmingham.

§ Present address: Department of Biosciences, University of Helsinki, 00014 Helsinki, Finland.

|| Present address: Stanford Synchrotron Radiation Laboratory, MS 69, Palo Alto, CA 94309.

⊥ University of Missouri−Kansas City.

[1] Abbreviations: GuHCl, guanidinium hydrochloride; EDTA, ethylenediaminetetraacetic acid; SAXS, small-angle X-ray scattering; Tris, tris(hydroxymethyl)aminomethane; ds, double-stranded; H/D exchange, hydrogen/deuterium exchange; CD, circular dichroism; gp5, coat protein subunit of the P22 shell, encoded by viral gene 5; CCMV, cowpea chlorotic mottle virus; *T*, icosahedral triangulation number.

10.1021/bi001965y CCC: $20.00
Published on Web 12/22/2000

of the molecular structure and stability at various stages of the assembly and maturation pathway.

High-resolution structures of viruses reveal conformational switching of chemically identical subunits to accommodate different quasi-equivalent contacts within the icosahedral lattice (*9*). Although high-resolution structures unveil the different subunit conformations in great detail, in most cases they do not address the conformation of the unassembled subunit. Therefore, the mechanism by which subunits attain their final conformation remains elusive. In particular, there are two fundamentally different models of conformational switching. In one case, the discrete conformations observed in the viral shell also exist in solution, and their proportion determines the outcome of the assembly process. This mechanism has been proposed in the assembly of bacteriophage HK97 (*10*), wherein the subunits equilibrate between pentamers and hexamers, which then assemble into a $T = 7$ capsid. In the second case, unassembled subunits attain an average conformation that is distinct from any of the states observed within the viral shell assembly. Here, the final conformation is attained only upon assembly, in a manner analogous to the conventional pathway of protein folding (*11*, *12*). Such a mechanism most likely governs the morphogenesis of the RNA phage MS2 (*13*, *14*) and dsDNA phage PRD1 (*11*).

In this study, we employ a well-characterized icosahedral dsDNA virus, the *Salmonella* bacteriophage P22, to study structural changes in the coat protein subunit accompanying procapsid assembly. The assembly pathway of P22 is an established prototype for the morphogenesis of many dsDNA viruses, including the herpesviruses and adenoviruses. In each case, assembly proceeds through the construction of a procapsid, which subsequently undergoes lattice maturation and DNA packaging (*2*, *15*) (Figure 1). The P22 procapsid contains three essential components: 420 molecules of coat protein (gp5, 47 kDa) arranged in a $T = 7$ icosahedral lattice, 200−300 molecules of scaffolding protein (gp8, 34 kDa) contained in the interior, and a dodecameric ring of portal protein (gp1, 87 kDa) located at a singular vertex that forms the site of DNA entry and exit. Scaffolding protein is released at about the time of DNA packaging (*16*). DNA packaging triggers maturation, which results in a 10% increase in the shell diameter (expansion) and concurrent capsid stabilization (*17*, *18*). Maturation is mediated by domain interchange between coat protein subunits and leads to increased stability of subunits and to a small increase in the amount of β-strand secondary structure (*19*−*21*).

The symmetry of the P22 procapsid requires that the coat subunit attain multiple (up to seven) conformations. At least four distinct conformations can be detected by electron cryomicroscopy at intermediate resolution (*18*). In previous biophysical studies, including Raman spectroscopy, conformational changes of the shell subunit during the expansion−maturation step were characterized (*21*−*25*). However, due to technical limitations, it was not possible to obtain spectroscopic signatures of the coat protein monomer and identify the structural changes required for assembly of the procapsid shell. In the study presented here, we employ Raman spectroscopy together with hydrogen/deuterium (H/D) exchange and small-angle X-ray scattering (SAXS) to characterize changes in subunit secondary and tertiary structures with procapsid shell formation. On the basis of these results and the known domain structure of the P22 coat protein (*26*), we propose a mechanism for conformational switching during morphogenesis of the dsDNA viral capsid.

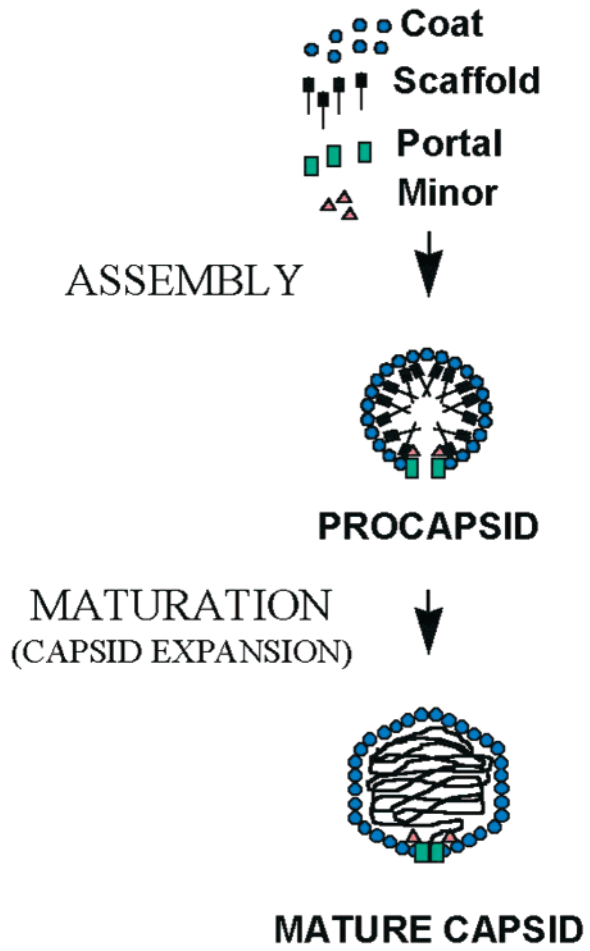


FIGURE 1: Assembly pathway of bacteriophage P22. The co-assembly (top) of 420 copies of coat, 200−300 copies of scaffold, 12 copies of portal, and a few copies of minor proteins yields a procapsid (middle), which matures through DNA packaging, scaffold ejection, and shell expansion to yield a mature capsid (bottom). Subsequent tail attachment is not shown.

MATERIALS AND METHODS

Protein and Procapsid Shell Purifications. Procapsid-like particles composed exclusively of coat and scaffolding proteins were expressed in *Escherichia coli* from a pET-based plasmid (pPM-5-Δ2) that encodes the coat and scaffolding genes (gift from S. Casjens, University of Utah, Salt Lake City, UT). Procapsid purification and scaffolding protein extraction were accomplished using previously described procedures (*27*).

Monomeric coat protein was obtained as follows. Purified procapsids were converted to denatured subunits by treatment with 6 M GuHCl. The denatured subunits were subsequently refolded at a protein concentration of 1 mg/mL by dialysis against buffer B (50 mM Tris, 25 mM NaCl, and 1 mM EDTA) at 4 °C. High-molecular mass aggregates were removed by centrifugation (40 000 rpm, Beckman 40.2 rotor, 2 °C, 2 h). The remaining protein solution was loaded onto a HiTrap-Q ion exchange column (Amersham-Pharmacia-Biotech, Inc., Piscataway, NJ) and eluted with a 0 to 0.3 M NaCl gradient. Aliquots of the peak fraction, which eluted

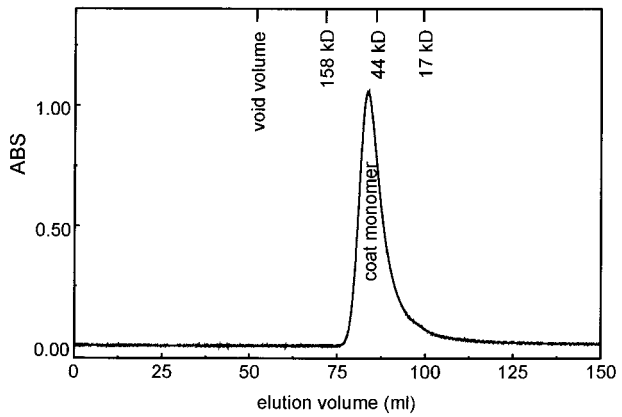


FIGURE 2: Size exclusion chromatography of the concentrated coat protein. Two milliliters of a coat protein solution (16 mg/mL) were loaded onto a TSK 3000 column. The position of the void volume (blue dextran) and elution of molecular mass standards are indicated on the top axis. Monomeric coat eluted in 50 mM Tris and 25 mM NaCl (pH 7.6) as a single peak at 84 mL, indicating an apparent molecular mass of 60 kDa.

at 0.2 M NaCl, were subsequently loaded onto a preparative TSK3000SW (22.5 mm × 300 mm) size exclusion column (TosoHass, Montgomeryville, PA) and eluted with buffer B at a rate of 4 mL/min and 4 °C. In a typical preparation, the Q column-concentrated coat protein was monomeric (Figure 2) and remained monodisperse and assembly active for about 24 h while stored on ice.

Raman Spectroscopy and Hydrogen/Deuterium Exchange. Samples for Raman spectroscopy were prepared from the peak fraction of the Q column. Identical Raman spectra were also obtained on a diluted sample of the protein (1.7 mg/mL) prepared from the monomer peak fraction of the TSK3000SW size exclusion column (Figure 2). Raman spectra were obtained on a Spex 500M single-monochromator spectrograph equipped with a notch filter and charge-coupled device detector (SpectrumOne, Instruments S. A., Edison, NJ). Spectra were excited with 120 mW of the 532 nm line from a solid-state, diode-pumped, frequency-doubled Nd:YVO_4 laser (model Verdi, Coherent, Santa Clara, CA). Samples were thermostated at 2 °C. Hydrogen/deuterium exchange was initiated by rapid buffer exchange [25 mM Tris and 0.2 M NaCl (pD 7.5)] using Biogel P-6 desalting columns (Bio-Rad, Hercules, CA). The exchange was monitored by time-resolved Raman spectroscopy at 2 °C using methods described previously in detail (*28*, *29*).

Small Angle X-ray Scattering. For small-angle X-ray scattering, the coat protein monomer was purified using size exclusion chromatography as described above (the protein concentration of the peak fraction ranging from 0.6 to 1.2 g/L). SAXS data were collected at beam line 4-2 of the Stanford Synchrotron Radiation Laboratory (*30*). The sample was placed in a thin (50 μL, 1.2 mm path) cell with mica windows. The cell was thermostated at 15 °C. Scattered X-rays were detected by use of a one-dimensional position sensitive proportional counter (EMBL/BioLogic model 210). The momentum transfer axis [$Q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength (1.38 Å)] of the detector was calibrated using the (100)-reflection of a cholesterol myristate powder sample. X-ray exposures 10 min in duration were subdivided into 10−20 exposures. Individual scattering curves were normalized to the intensity of the incident beam, background-subtracted, and averaged after confirmation of no potential radiation-induced aggregation.

RESULTS AND DISCUSSION

Conformation and Folding of the Coat Protein Subunit

Monomer Shape and Oligomerization. The refolded coat protein subunit eluted as a monomer from the anion exchange column and remained monomeric even at a high concentration (16 g/L) when stored in a buffer containing ≥0.2 M NaCl (Figure 2). The monomer assembled into procapsids when mixed with scaffolding protein in buffer B (data not shown). Compared with globular protein standards, the 47 kDa monomer eluted from a size exclusion column with a higher than expected apparent mass (60 kDa), indicative of an asymmetrically shaped molecule or an expanded (partially unfolded) globular conformation (Figure 2).

To further characterize the compactness and shape of the monomer, SAXS data were collected from dilute solutions (Figure 3). An apparent mass of 49 kDa was obtained by extrapolation of the data of Figure 3A to an angle of 0° using the Guinier approximation (*31*). The SAXS data confirm a monomeric state for the coat protein subunit. The monomer exhibited a radius of gyration of 36.7 ± 0.7 Å, significantly larger than the value of 18.7 Å predicted for a sphere of equivalent volume and mass. The SAXS results thus imply that the monomer is not spherically shaped (*32*).

To gain information about the distribution of mass within the molecule, the reciprocal space data of Figure 3A were transformed into a real space distance distribution function using the regularization method (*33*). The distance distribution function of Figure 3B provides information about the atom-weighted distribution of distances within the molecule. A maximum distance within the protein (D_{max}) of 120 ± 10 Å was obtained, a value considerably larger than expected for a spherical molecule (48 Å). The shape of the monomer was estimated using a simulated annealing procedure and a dummy atom model (*34*). The two best models, shown in Figure 3C, indicate a slightly bent and elongated shape about 100 Å in length.

Secondary Structure. The solution secondary structure of the monomer was probed by Raman spectroscopy using methods similar to those applied previously to the procapsid and capsid shells (*21*, *25*, *29*). The middle trace of Figure 4 shows the Raman spectrum in the region of 600−1800 cm^{-1}. The monomer exhibits a strong and broad amide I band centered near 1667 cm^{-1}, diagnostic of a relatively broad distribution of secondary structures in which β-strand nevertheless predominates (*35*, *36*). A quantitative estimate of the secondary structure distribution using the reference intensity profile method of Berjot et al. (*37*) indicates 22 ± 4% α-helix, 39 ± 2% β-strand, and 38 ± 5% undefined or irregular conformations (loops, turns, coils, and other aperiodic structures). The broad amide III band of the monomer is centered at 1251 cm^{-1}, which is consistent with the amide I profile (*35*, *36*). The Raman-based estimate of 22% α-helix in the monomer is also in agreement with the value of ∼20% α-helix estimated by circular dichroism (CD) (*38*). It should be noted that the significant α-helical content of the coat protein precludes a reliable estimate of β-strand content by

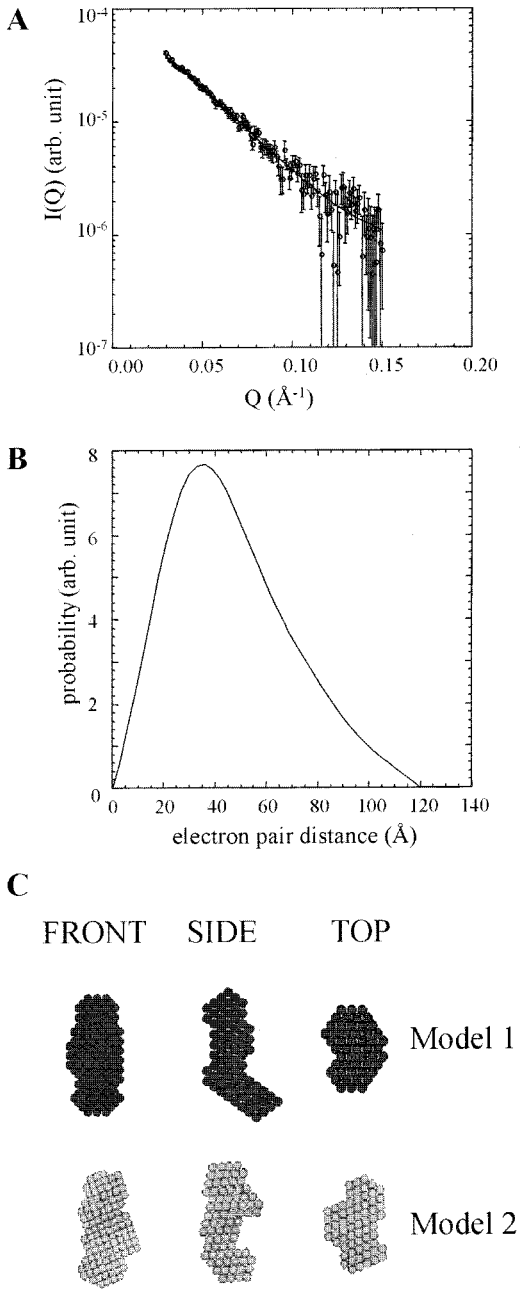


FIGURE 3: Small-angle X-ray scattering of the coat protein monomer. The monomer peak fraction (1.2 g/L) from the TSK 3000 column was used directly to obtain the data. (A) Scattering intensity vs momentum transfer *Q*. (B) The scattering intensity was reproduced by the distance distribution function using an indirect transform method (*32*, *33*). The apparent mass of 49 kDa was obtained by comparison of the intensity extrapolated to zero angle with that of a scaffolding protein standard [covalent dimer (68 kDa) containing the mutation R74C (*70*)]. (C) The two best models for the monomeric subunit are shown in orthogonal projections. The volume for each model is the same as that of the monomer. The models were generated by fitting the data in panel A using the DAMMIN procedure (*34*). A bead diameter of 5 Å was used for fitting and display.

CD spectroscopy. However, Raman spectroscopy clearly identifies β-strand as the prevalent conformation.

The finding of β-strand as the dominant secondary structure in the monomeric state of the coat protein is in accord with Raman results on subunits of the P22 procapsid and capsid shells (*22*, *24*, *25*), and suggests that the fundamental fold is preserved during assembly. Extensive

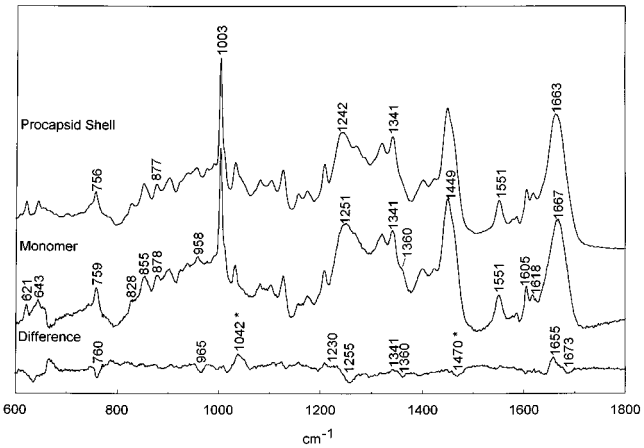


FIGURE 4: Raman spectra in the region of 600−1800 cm^{-1} (532 nm excitation) of the bacteriophage P22 procapsid shell at 18 mg/mL and 10 °C in buffer B (top) and unassembled coat protein monomer at 16 mg/mL and 2 °C in buffer B (middle). The difference spectrum (bottom) was computed by subtraction of the middle spectrum from the top spectrum. Spectra were corrected for Raman scattering of buffer B, and intensities were normalized using the 1003 cm^{-1} band of phenylalanine (*25*).

β-strand structure (in the form of a jellyroll fold) is also prevalent in the capsid subunits of many other icosahedral viruses (*39*−*44*). Recently, the same fold has been demonstrated by Raman spectroscopy for the bacteriophage PRD1 capsid subunit (protein P3, 43 kDa) (*11*). The PRD1 capsid subunit provides a useful comparison for that of P22 because it has been studied extensively by both Raman spectroscopy (*11*) and X-ray crystallography (*39*). On the basis of the Raman amide I profile (*11*), the estimated secondary structure (*37*) for the P3 subunit of the PRD1 capsid is 18% α-helix and 44% β-strand. This is in reasonably good agreement with the X-ray structure of P3, which gives 13% helix and 36% strand (*39*). The modest differences between Raman- and X-ray-determined values of helix and strand presumably reflect overlapping amide I band contributions from other types of structure (turns and irregular conformations) as well as the more strict classification of peptide backbone conformations by X-ray crystallography. Nevertheless, the ratio of β-strand to α-helix is about the same in both X-ray (2.75) and Raman (2.6) analyses. In contrast, a substantially lower β/α ratio (1.8) is obtained for the monomeric subunit of P22, as is apparent from inspection of the corresponding Raman amide I and amide III contours (compare Figure 4 with Figure 3 of ref *11*). The sharper amide I and amide III bands of the PRD1 shell resemble more closely those of the mature P22 shell (*23*). Interestingly, PRD1 is one of the few dsDNA viruses that assembles without an apparent shell maturation (expansion) step; the PRD1 coat protein assembles directly into the mature capsid architecture (*45*).

Finally, we note that the Raman signature of the mature capsid shell of the lambdoid phage HK97 (R. Hendrix, J. M. Benevides, and G. J. Thomas, Jr., unpublished data) exhibits clearly discernible α-helix markers, viz. an amide I shoulder at 1654 cm^{-1} and enhanced amide III intensity between 1250 and 1300 cm^{-1}, consistent with the recently determined X-ray crystal structure (*46*). Thus, despite the similar icosahedral capsid architectures of PRD1, P22, and HK97, significantly different subunit folds are evident from the respective Raman amide band signatures. The protein

folds of P22 and PRD1 subunits have more in common with one another than with that of HK97 (*46*).

Tertiary Structure. In addition to the assessment of secondary structure, Raman spectroscopy permits evaluation of tertiary structure through a variety of conformation- and environment-sensitive marker bands of selected side chains. Several of these marker bands are discussed next.

Raman indicators of the average hydrogen-bonding environment of the eight tyrosine phenolic groups of coat protein occur at 828 and 855 cm^{-1}. In the monomer spectrum (Figure 4, middle trace), these bands exhibit an intensity ratio (I_{855}/I_{828}) of ≈2. This is close to the maximum value of 2.5, which is diagnostic of a phenoxyl group that acts exclusively as the acceptor of a hydrogen bond (*47*). Accordingly, in the monomeric state, the majority of the eight tyrosines per coat protein subunit have the phenolic oxygen engaged as a hydrogen-bond acceptor. [The low intensity of the 828 cm^{-1} band can also arise from a tyrosine phenoxyl that is not hydrogen bonded, as occurs, for example, in the protein coat of Ff filamentous virus or in the vapor phase of *p*-cresol (*48*−*50*). However, such a condition is not expected for tyrosines of globular proteins or their assemblies.]

The relatively intense 760 cm^{-1} band of tryptophan (normal mode *W18*) suggests that the average environment of the six tryptophans per subunit is relatively polar (*51*). The *W17* mode of Trp, which is sensitive to hydrogen bonding of the indolyl N1H group, is observed at 880 cm^{-1}. This is consistent with relatively weak hydrogen bonding (*52*). The distinct shoulder at 1360 cm^{-1} is part of a tryptophan Fermi doublet [1340/1360 cm^{-1} (*53*)] that is sensitive to the hydrophobicity of the indolyl ring environment. Its intensity in the P22 shell monomer is similar to that observed for the small globular protein, *E. coli* thioredoxin (*54*), which contains two partially exposed Trp residues. Thus, the tryptophan residues of the coat protein are considered to be at least partially exposed to solvent in the monomeric state.

The tryptophan marker at 1551 cm^{-1} is assigned to the indolyl *W3* mode, the wavenumber value of which is diagnostic of the magnitude of the side chain torsion angle $\chi^{2,1}$ (*55*). The data presented here indicate that the average C3−Cβ bond has a $|\chi^{2,1}|$ of ≈95°. The position and sharpness of this marker suggest further that all indolyl rings adopt the same $|\chi^{2,1}|$ value. In previous work, a single-exponential decay of tryptophan fluorescence was reported for the coat protein monomer, implying that the six Trp residues per subunit share a similar amphipathic local environment (*56*). The results presented here are consistent with the fluorescence data and show further that all Trp side chain orientations are virtually identical with respect to $|\chi^{2,1}|$.

The Raman band corresponding to the sulfhydryl (S−H) stretching vibration of cysteine is diagnostic of the hydrogen-bonding environment of the S−H group (*57*). For the single cysteine (Cys 405) of the coat protein monomer, the Raman marker is observed at 2568 cm^{-1}, which indicates that the S−H group is the donor of a moderately strong hydrogen bond. The sharpness of the band further indicates that the S−H···X hydrogen bond adopts a single, well-defined configuration; e.g., there is no heterogeneity among different monomer molecules (*21*, *57*). Thus, the C-terminal domain attains the same well-defined tertiary structure in all subunits in solution.

Conformational Changes with Procapsid Assembly

To characterize structural changes accompanying subunit assembly, we compared Raman signatures of the monomeric subunit and the procapsid shell. The icosahedral procapsid shell ($T = 7$), which is composed solely of the coat protein subunit (gp5), is generated by selective removal of scaffolding and minor proteins using low concentrations of GuHCl. Despite subtle alterations in capsid morphology attendant with GuHCl treatment (*58*), previous studies have demonstrated that Raman spectra of GuHCl-treated and untreated shells are identical, provided that the spectral contribution from the scaffolding protein is compensated (*25*). These experiments demonstrate that GuHCl treatment does not appreciably alter the subunit conformation.

Secondary Structure. Although Raman spectra of the monomeric and procapsid shell states of coat protein share many similar features, it is clear from Figure 4 that procapsid assembly perturbs the structure of the coat protein monomer. The computed difference spectrum shown in the bottom trace of Figure 4 highlights the spectral changes accompanying assembly.

Importantly, the center of the Raman amide I marker is shifted from 1667 cm^{-1} in the monomer to 1663 cm^{-1} in the procapsid shell. This shift gives rise to Raman difference peaks at 1655 cm^{-1} (assigned to α-helix) and 1673 cm^{-1} (β-strand) and corresponding troughs at lower and higher wavenumber values. The difference band pattern of Figure 4 reflects a narrower amide I profile in the procapsid shell spectrum; i.e., the more ordered secondary structure of the shell subunit occurs at the expense of disordered chain segments (loops, coil, etc.) in the monomer. The more highly ordered secondary structure of the subunit in the assembled shell is further demonstrated in the difference spectrum of Figure 4 by the diminished intensity of the Raman amide III band at 1255 cm^{-1}, assigned to disordered structure (*25*, *36*, *59*) and enhanced intensity of the Raman amide III marker at 1230 cm^{-1} (β-strand). Despite overlapping contributions expected from various side chains near 900−1000 and 1300−1350 cm^{-1} (*53*, *60*), we may interpret the negative difference band at 965 cm^{-1} (*61*) and positive difference band at 1341 cm^{-1} (*62*) also as indications of a more highly α-helical secondary structure in subunits of the procapsid shell.

Overall, it is clear from Figure 4 that the polypeptide backbone of the coat protein monomer undergoes refolding during procapsid assembly. An estimate of the number of peptide linkages involved can be obtained from the normalized area of the amide I band differences. We obtain 4 ± 1% of the backbone, or 17 of 428 peptide bonds. In a recent study, Lanman and co-workers (*26*) demonstrated that a flexible hinge encompassing residues 180−205 of the subunit becomes protease-resistant upon procapsid assembly. Residues within this hinge are likely candidates for refolding. Thus, assembly of the P22 procapsid leads to local ordering of a small but significant percentage of the peptide residues. Similar subunit folding has been reported for capsid assembly of the P3 subunit of PRD1 (*11*). In the case of PRD1, however, a smaller subset (∼2%) of peptide linkages is involved. Also, the PRD1 capsomeres (P3 trimers) do not exhibit the assembly-induced increase in global stability (*11*) that characterizes subunits of the P22 shell lattice (*26*, *63*).

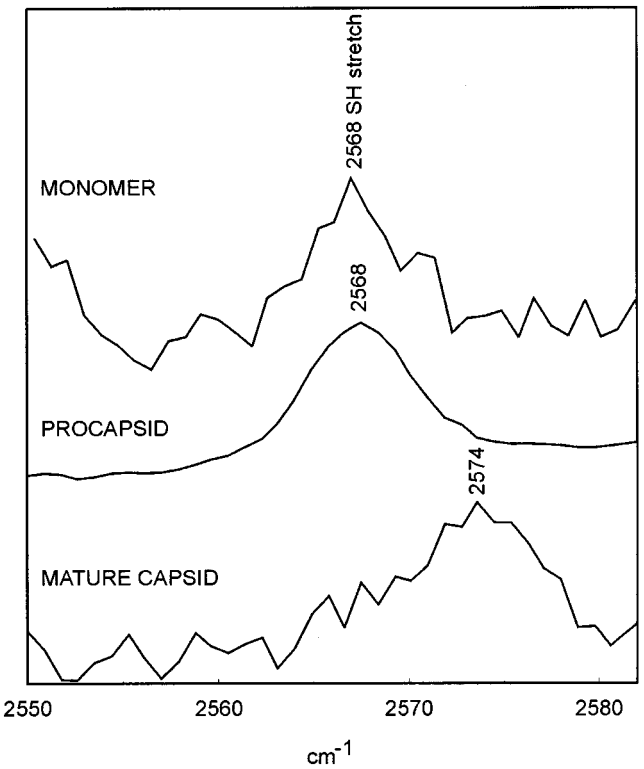


FIGURE 5: Raman spectra in the region of 2550−2580 cm^{-1} (532 nm excitation) of the P22 coat protein monomer (top), procapsid shell (middle), and mature capsid shell (bottom). Monomer and shell concentrations and other conditions are as given in the legend of Figure 4. Mature shells were prepared as previously described (*21*).

Tertiary Structure. In addition to the changes in subunit secondary structure noted above, P22 procapsid assembly leads to changes in tertiary structure that are also revealed by Raman spectroscopy. The shift of the tryptophan *W18* marker from 760 cm^{-1} in the monomer to 756 cm^{-1} in the procapsid shell with a concomitant large decrease in intensity (Figure 4) signifies a substantial change in the amphipathic environment of the average Trp side chain. Broadening of the 1360 cm^{-1} component of the Trp Fermi doublet indicates that this change is due to a more hydrophilic indolyl environment in the procapsid shell subunit.

The tryptophan *W3* marker (1551 cm^{-1}) exhibits no assembly-related shift in its band center, indicating a negligible change in the average $|\chi^{2,1}|$ torsion (*55*). Overall, Raman spectral features diagnostic of the Trp side chains exhibit only very small changes with the transition from the monomer to procapsid shell. This contrasts dramatically with the large Trp side chain reorientations that accompany shell expansion (*22*).

Invariant to the assembly process are Raman indicators of hydrogen bonding by tyrosine phenoxyls (markers at 830 and 850 cm^{-1}, Figure 4) and by the cysteine sulfhydryl of Cys 405 (marker at 2568 cm^{-1}, Figure 5). Additionally, other assigned Raman markers of the aliphatic and aromatic side chains of coat protein (*22*−*24*) are affected little by assembly of the procapsid (Figure 4). Thus, despite measurable folding of the coat protein backbone upon procapsid assembly, the tertiary structure of the subunit is not significantly affected. In this regard, the P22 shell assembly is similar to that of PRD1 (*11*).

It is interesting to note, however, that in contrast to the formation of the P22 procapsid shell from monomers, the process of shell maturation (procapsid-to-capsid transformation) leads to large and extensive changes in both the environments and hydrogen-bonding interactions of numerous side chains (*21*, *22*). A striking example of this effect, which is illustrated in Figure 5, is the large wavenumber shift (2568 → 2574 cm^{-1}) that occurs in the Raman S−H marker of Cys 405 with shell maturation. Clearly, the moderate S−H···X hydrogen bond of the Cys 405 sulfhydryl in the monomer is replaced by a much weaker hydrogen bond in the subunit of the mature capsid (*21*, *57*). [See also the discussion of SH → SD exchange dynamics of the monomer and shells (below).] Because of the location of Cys 405 near the C-terminus of the subunit sequence, it is concluded that the C-terminal domain, and particularly the local environment of the Cys 405 sulfhydryl group, plays a key role in monomer folding and capsid maturation processes, but not in the procapsid assembly step.

Hydrogen/Deuterium Exchange Dynamics

Amide Exchange. In previous work, we compared the Raman signatures and the rates of deuterium exchange of amide NH groups (H/D exchange) in subunits of the P22 procapsid and capsid (*21*). The data revealed that although the procapsid-to-capsid maturation occurs with conservation of subunit secondary structure (*22*), it is accompanied by a very large (2-fold) increase in the level of exchange protection of shell subunits, as well as the above-noted extensive changes in side chain environments (*22*). These results indicate that the shell transformation involves subunit domain movement accompanied by a large increase in the level of effective exchange protection in the core of the subunit, without appreciable refolding of subunit domains. Here, we extend the previous approaches (*21*, *28*) and employ Raman amide bands to compare H/D exchange dynamics of the monomer and procapsid shell subunit. H/D exchange was previously probed both at low temperatures (exchange via the native conformation) and at elevated temperatures (exchange via a locally unfolded conformation). However, because the monomer is intrinsically less thermostable than the shell assembly and subject to off-pathway aggregation at elevated temperatures, this study has been limited to exchange under strictly native conditions.

Figure 6A compares Raman markers in the amide I region of the monomer and procapsid shell subunit, following exposure of each sample to D_2O (pD 7.8) under native conditions (2 °C for the monomer and 10 °C for the procapsid shell). As expected, the band centers are shifted by deuteration to wavenumber values lower (amide I′) than those that occur in the corresponding nondeuterated samples (cf. amide I markers of Figure 4). Deuterium exchange was allowed to reach a plateau, about 30 min for the monomer and 120 min for the procapsid shell, after which further exchange was negligible. However, not all amide NH sites become deuterated under these native conditions (see also amide III′ markers in Figure 6C). The dominant Raman band in each spectrum of Figure 6A actually represents an overlap of contributions from amide I′ (due to exchanged ND peptides) and amide I (due to residual nonexchanged NH peptides). The composite amide I/amide I′ band of the monomer is centered at 1658 cm^{-1}, while that of the

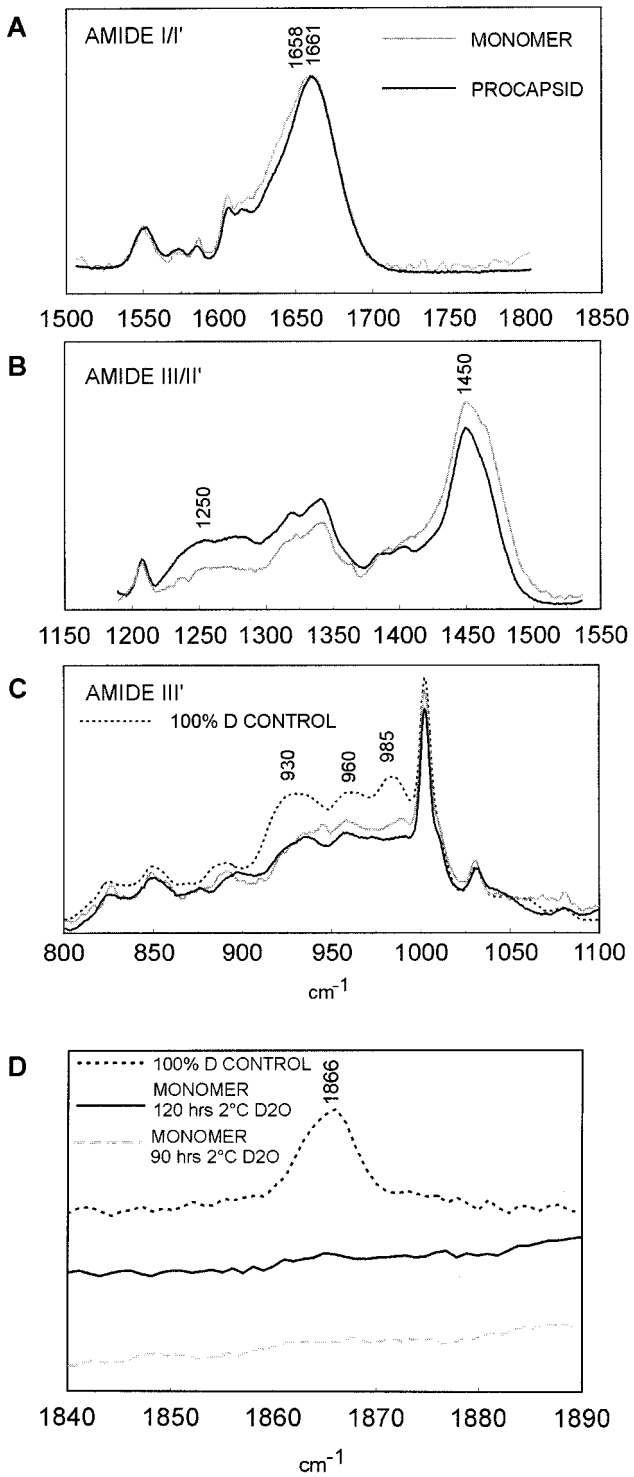


FIGURE 6: Normalized Raman signatures of native-state H/D exchanges of the P22 coat protein monomer (2 °C) and procapsid shell (10 °C), following exposure to D_2O buffer for 12 h (see the text): (A) amide I/amide I′ profiles, (B) amide III/amide II′ profiles, (C) amide III′ profiles, and (D) S−D stretching band profiles, following prolonged exposure to D_2O. Included in panel D as a control is the S−D band profile of the fully exchanged procapsid subunit incubated at 40 °C (see the text).

procapsid shell is centered at 1661 cm^{-1}. The lower wavenumber value for the composite band of the monomer reflects both a slightly different secondary structure and a greater degree of amide H/D exchange vis-à-vis the subunit of the procapsid shell. (By comparison, the nonexchanged monomer and procapsid shell subunit yield amide I bands centered at 1664 and 1667 cm^{-1}, respectively, as seen in Figure 4.)

The corresponding Raman amide III profile (due to nonexchanged NH groups) and amide II′ profile (due to exchanged ND groups), shown in Figure 6B, are entirely consistent with the amide I/amide I′ data depicted in Figure 6A. Thus, the monomer exhibits a lower amide III intensity (∼1250 cm^{-1}) and a greater amide II′ intensity (∼1450 cm^{-1}) than the procapsid shell subunits, owing to more extensive native-state H/D exchange in the monomer.

Figure 6C compares Raman profiles in the amide III′ region (900−1000 cm^{-1}, due to exchanged ND groups) of the monomer and procapsid shell subunit after native-state exchange. Also shown in Figure 6C is the amide III′ profile of a coat protein control in which all NH groups have been exchanged with deuterium under denaturing conditions prior to refolding of the protein in D_2O. Again, these data confirm that H/D exchange is more extensive in the monomer than in the procapsid shell subunits. In particular, greater amide III′ intensity is observed at 930−945 cm^{-1} which can be assigned to deuterated α-helix (*62*, *64*), at 960 cm^{-1} which can be assigned to deuterated disordered peptides (*21*, *35*, *36*), and at 980−990 cm^{-1} which can be assigned to deuterated *β*-strand (*21*, *35*, *36*, *64*). The amide III′ exchange profile of the monomer indicates that the protein fold of the monomer comprises unprotected and protected residues in regions of α-helical, *β*-stranded, and disordered secondary structure. Panels B and C of Figure 6 also show that neither the procapsid shell subunit nor the monomer undergoes complete exchange of all peptide residues located in α-helical, *β*-stranded, or disordered regions of secondary structure. The data suggest that the most protected residues of the monomer are those located in regions of *β*-strand structure.

The substantial exchange protection of the monomer, estimated from Figure 6C as 30% of all NH sites, indicates that the protein fold is sufficiently well defined to produce a relatively compact tertiary structure. Accordingly, the less compact and "molten-globule-like" segments of the protein implied by binding of the hydrophobic probe bis-ANS (*38*) may be limited to small regions localized at the protein surface and potentially involved in shell assembly. One locus of the monomer that is likely to undergo a folding-like conformational change upon assembly is the interdomain hinge region (residues 180−205), which becomes more protease-resistant and exchange-protected in the procapsid shell (R. Tuma and P. E. Prevelige, manuscript submitted for publication) (*26*).

Side Chain Exchange. H/D exchange of selected side chains can be used to localize and monitor the folding core of a protein (*11*, *21*, *25*, *28*, *46*). For example, the single sulfhydryl of the coat protein (Cys 405) provides a selective probe of exchange of the C-terminal region. Figure 6D shows Raman spectra in the region of 1840−1890 cm^{-1}, where the stretching vibration of the deuterated sulfhydryl group (S−D) is expected to occur. The absence of an S−D band in the spectrum of the monomer shows unequivocally that the Cys 405 sulfhydryl is fully protected against H/D exchange. Protection persists for samples incubated under native conditions (pD 7.8 and 2 °C) for up to 6 days. The SH group can be exchanged only under conditions that favor global unfolding (40 °C). Thus, the Cys 405 sulfhydryl is buried within the hydrophobic folding core of the monomer. A

similar degree of sulfhydryl exchange protection is observed for the procapsid shell subunit (*21*).

The hydrogen-bonding state of the Cys 405 sulfhydryl is also similar in the monomer and procapsid shell subunit [cf. Figure 5 and the data of Tuma at al. (*21*)]. However, both the hydrogen-bonding environment and exchange protection of the sulfhydryl are dramatically altered in the mature capsid, where S−H···X hydrogen bonding is considerably weaker (*21*). Shell expansion and SH → SD exchange of the procapsid subunit have been shown to be governed by similar activation energies (*21*). This finding strongly suggests that shell maturation is mediated by a structural rearrangement at the locus of Cys 405 (exchange-protected folding core), consistent with a mechanism involving an interchange of subunit domains (*21*). Therefore, it is reasonable to conclude that S−H···X hydrogen bonding of the Cys 405 sulfhydryl provides an important source of stabilization of the monomer in a coupled folding−assembly process.

CONCLUSIONS

A biological requirement of the unassembled coat protein monomer of P22 is a well-defined tertiary structure, which can support the specific biomolecular recognition required for shell assembly without compromising the flexibility that is essential for quasi-equivalent contacts in a $T = 7$ icosahedron. This study demonstrates that this requirement is met by an elongated molecule containing much of the β-strand secondary structure also present in the subunit of the mature capsid assembly. Nevertheless, subtle structural alterations are shown to take place between monomeric and assembled states of the subunit.

In particular, approximately 17 of the 430 amino acid residues undergo a refolding event in which they become assimilated as regions of α-helix and β-strand secondary structure. These residues may correspond to the flexible linker region recently identified in studies of proteolytic susceptibility (*26*). A similar linker region has been identified in the FG loop of bacteriophage MS2 coat protein, which also adopts multiple conformations along the assembly pathway (*13*, *14*). Another example of such refolding is the selective formation of a β-strand annulus at the pseudo-3-fold axis in the capsid of the single-stranded RNA plant virus CCMV (*40*).

Previous H/D exchange studies revealed that the procapsid shell subunit retains a considerable degree of conformational flexibility, and that this degree of freedom is lost during the subsequent expansion−maturation step (*21*). On the basis of this finding, it was suggested that the subunit within the procapsid resembles a late folding intermediate. Present results show that assembly of the monomer into the procapsid shell has surprisingly little effect on the environments of subunit side chains. There is essentially no change in tyrosine and cysteine hydrogen-bonding environments and, at most, a very small change in the average indolyl hydrophobicity of tryptophan. In addition, both H/D exchange and SAXS measurements indicate that the unassembled subunit adopts a relatively compact tertiary structure. Therefore, the coat protein monomer represents a compact, late-folding intermediate that is only slightly more locally disordered than its procapsid shell counterpart.

The interrelationship between folding and assembly can be depicted in the context of an energy landscape, as shown in Figure 7. At high dilutions, the coat protein folds into a soluble monomer and, apart from a flexible loop (∼20 residues) of the polypeptide chain, adopts secondary and tertiary structures similar to those of the subunit incorporated within the procapsid shell. At higher concentrations (>1 mg/mL; R. Tuma, H. Tsuruta, and P. E. Prevelige, Jr., unpublished data), the monomer gradually overcomes the high entropic barrier, generated in part by ordering of the loop region, and assembles (*65*, *66*). The activation barrier is presumed to be lowered by the presence of the scaffolding protein, which acts as an entropy sink by utilizing the energy of scaffolding protein dimerization to promote coat subunit association (*67*, *70*). Thus, the disorder−order transition within the loop region is thermodynamically coupled to the binding of the scaffolding protein and in effect exerts control over the assembly process.

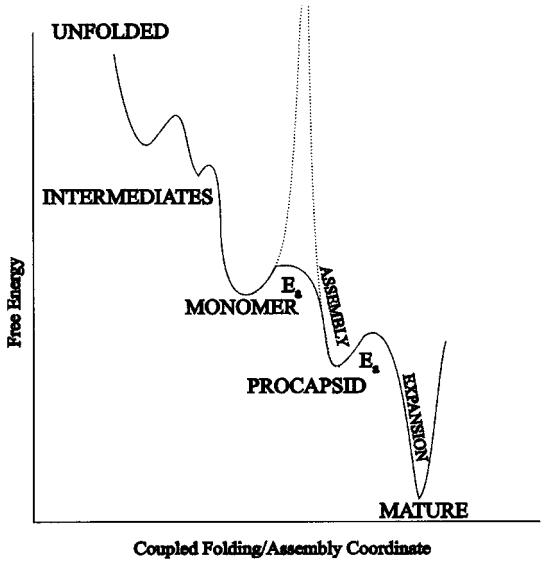


FIGURE 7: Schematic energy landscape for folding and assembly of P22 coat protein. The broader minima for the monomer and procapsid subunit indicate greater conformational flexibility of these states. The height of the free energy barrier separating the monomer and procapsid states is lowered in the presence of the scaffolding protein. The sharp and deep minimum for the mature subunit reflects its greater dynamic stability as revealed by hydrogen isotope exchange. The energy landscape that the monomer experiences at infinite dilution is shown by a dotted line. In this case, the procapsid state is separated from the monomeric state by an infinitely high free energy barrier.

Both the monomeric coat protein and the coat protein subunits within the procapsid retain a considerable degree of conformational freedom, a phenomenon typical of (late) folding intermediates (*68*). It has been pointed out that viral capsids are faced with the conflicting demands of achieving fidelity during assembly and stability of the final structure, and that these demands can be met through the use of many weak interactions (*69*). Although providing sufficient energy to form a stable precursor capsid, such weak interactions might be insufficient to form extremely stable viral capsids. A general scheme for circumventing this problem is a two-step assembly process in which a metastable precursor form is subsequently stabilized. In the case of phage HK97, such stabilization is achieved through extensive covalent cross-linking, whereas in certain other cases (e.g., adenovirus and herpesvirus), it is achieved through the use of "glue" proteins. However, in P22 and many other dsDNA viruses, a procapsid

with weak subunit−subunit contacts is assembled initially, and then is converted into a capsid with much stronger intersubunit interactions. This maturation process, which is encoded within the primary sequence of the coat protein, is accomplished via a shift in the energy landscape of the subunit within the procapsid lattice. On this landscape, a new low-energy state, corresponding to the mature capsid, becomes accessible. This state is populated after expansion. During expansion, the subunit secondary structure is perturbed only modestly (2−3% increase in β-strand content), but the tertiary structure is radically altered (domain rearrangements), and the exchange-protected folding core that is common to both the procapsid-assembled subunit and the monomer undergoes a major (2-fold) increase in its level of protection. These findings, which for the first time include studies of the coat protein monomer, extend and confirm the previously proposed two-step folding model for the P22 coat protein (*21*).

REFERENCES

1. Rossmann, M. G. (1984) *Virology 134*, 1−11.
2. King, J., and Chiu, W. (1997) in *Structural Biology of Viruses* (Chiu, W., Burnett, R. M., and Garcea, R., Eds.) pp 288−311, Oxford University Press, New York.
3. Canady, M. A., Tihova, M., Hanzlik, T. N., Johnson, J. E., and Yeager, M. (2000) *J. Mol. Biol. 299*, 573−584.
4. Zlotnick, A., Reddy, V. S., Dasgupta, R., Schneemann, A., Ray, W. J., Jr., Rueckert, R. R., and Johnson, J. E. (1994) *J. Biol. Chem. 269*, 13680−13684.
5. Chow, M., Basavappa, R., and Hogle, J. M. (1997) in *Structural Biology of Viruses* (Chiu, W., Burnett, R. M., and Garcea, R., Eds.) pp 157−186, Oxford University Press, New York.
6. Wien, M. W., Chow, M., and Hogle, J. M. (1996) *Structure 4*, 763−767.
7. Belnap, D. M., Filman, D. J., Trus, B. L., Cheng, N., Booy, F. P., Conway, J. F., Curry, S., Hiremath, C. N., Tsang, S. K., Steven, A. C., and Hogle, J. M. (2000) *J. Virol. 74*, 1342−1354.
8. Tsang, S. K., Danthi, P., Chow, M., and Hogle, J. M. (2000) *J. Mol. Biol. 296*, 335−340.
9. Rossmann, M. G., and Erickson, J. W. (1985) in *Virus Structure and Assembly* (Casjens, S., Ed.) pp 30−73, Jones and Bartlett, Boston.
10. Duda, R. L., Martincic, K., and Hendrix, R. W. (1995) *J. Mol. Biol. 247*, 636−647.
11. Tuma, R., Bamford, J. H. K., Bamford, D. H., Russell, M. P., and Thomas, G. J., Jr. (1996) *J. Mol. Biol. 257*, 87−101.
12. Kumar, S., Ma, B., Tsai, C. J., Sinha, N., and Nussinov, R. (2000) *Protein Sci. 9*, 10−19.
13. Golmohammadi, R., Valegard, K., Fridborg, K., and Liljas, L. (1993) *J. Mol. Biol. 234*, 620−639.
14. Ni, C. Z., Syed, R., Kodandapani, R., Wickersham, J., Peabody, D. S., and Ely, K. R. (1995) *Structure 3*, 255−263.
15. Casjens, S., and Hendrix, R. (1988) in *The Bacteriophages* (Calendar, R., Ed.) pp 15−91, Plenum Press, New York.
16. King, J., and Casjens, S. (1974) *Nature 251*, 112−119.
17. Earnshaw, W., Casjens, S., and Harrison, S. C. (1976) *J. Mol. Biol. 104*, 387−410.
18. Prasad, B. V., Prevelige, P. E., Marietta, E., Chen, R. O., Thomas, D., King, J., and Chiu, W. (1993) *J. Mol. Biol. 231*, 65−74.
19. Galisteo, M. L., and King, J. (1993) *Biophys. J. 65*, 227−235.
20. Prevelige, P. E., Jr., and King, J. (1993) *Prog. Med. Virol. 40*, 206−221.
21. Tuma, R., Prevelige, P. E., Jr., and Thomas, G. J., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9885−9890.
22. Prevelige, P. E., Jr., Thomas, D., Aubrey, K. L., Towse, S. A., and Thomas, G. J., Jr. (1993) *Biochemistry 32*, 537−543.
23. Prevelige, P. E., Jr., Thomas, D., King, J., Towse, S. A., and Thomas, G. J., Jr. (1990) *Biochemistry 29*, 5626−5633.
24. Thomas, G. J., Jr., Li, Y., Fuller, M. T., and King, J. (1982) *Biochemistry 21*, 3866−3878.
25. Tuma, R., Prevelige, P. E., Jr., and Thomas, G. J., Jr. (1996) *Biochemistry 35*, 4619−4627.
26. Lanman, J. K., Tuma, R., and Prevelige, P. E., Jr. (1999) *Biochemistry 38*, 14614−14623.
27. Prevelige, P. E., Jr., Thomas, D., and King, J. (1993) *Biophys. J. 64*, 824−835.
28. Tuma, R., and Thomas, G. J., Jr. (1996) *Biophys. J. 71*, 3454−3466.
29. Tuma, R., and Thomas, G. J., Jr. (1997) *Biophys. Chem. 68*, 17−31.
30. Tsuruta, H., Brennan, S., Rek, Z. U., Irving, T. C., Tompkins, W. H., and Hodgson, K. O. (1998) *J. Appl. Crystallogr. 31*, 672−682.
31. Serdyuk, I. N., Tsalkova, T. N., Svergun, D. I., and Izotova, T. D. (1987) *J. Mol. Biol. 194*, 126−128.
32. Feigin, L. A., and Svergun, D. I. (1987) *Structure Analysis by Small-Angle X-ray and Neutron Scattering*, Plenum Press, New York.
33. Svergun, D. I. (1991) *J. Appl. Crystallogr. 24*, 485−492.
34. Svergun, D. I. (1999) *Biophys. J. 76*, 2879−2886.
35. Chen, M. C., and Lord, R. C. (1974) *J. Am. Chem. Soc. 96*, 4750−4752.
36. Chen, M. C., Lord, R. C., and Mendelsohn, R. (1974) *J. Am. Chem. Soc. 96*, 3038−3042.
37. Berjot, M., Marx, J., and Alix, A. J. P. (1987) *J. Raman Spectrosc. 18*, 289−300.
38. Teschke, C. M., King, J., and Prevelige, P. E., Jr. (1993) *Biochemistry 32*, 10658−10665.
39. Benson, S. D., Bamford, J. K., Bamford, D. H., and Burnett, R. M. (1999) *Cell 98*, 825−833.
40. Speir, J. A., Munshi, S., Wang, G., Baker, T. S., and Johnson, J. E. (1995) *Structure 3*, 63−78.
41. Rossmann, M. G., and Johnson, J. E. (1989) *Annu. Rev. Biochem. 58*, 533−573.
42. Hartman, K. A., McDonald-Ordzie, P. E., Kaper, J. M., Prescott, B., and Thomas, G. J., Jr. (1978) *Biochemistry 17*, 2118−2123.
43. Verduin, B. J., Prescott, B., and Thomas, G. J., Jr. (1984) *Biochemistry 23*, 4301−4308.
44. Li, T., Chen, Z., Johnson, J. E., and Thomas, G. J., Jr. (1992) *Biochemistry 31*, 6673−6682.
45. Butcher, S. J., Bamford, D. H., and Fuller, S. D. (1996) *EMBO J. 14*, 6078−6086.
46. Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W., and Johnson, J. E. (2000) *Science 289*, 2129−2133.
47. Siamwiza, M. N., Lord, R. C., Chen, M. C., Takamatsu, T., Harada, I., Matsuura, H., and Shimanouchi, T. (1975) *Biochemistry 14*, 4870−4876.
48. Overman, S., and Thomas, G. J., Jr. (1995) *Biochemistry 34*, 5440−5451.
49. Overman, S., and Thomas, G. J., Jr. (1998) *J. Raman Spectrosc. 29*, 23−29.
50. Arp, Z., Laane, J., Overman, S. A., and Thomas, G. J., Jr. (2001) *Biochemistry 40* (in press).
51. Miura, T., Takeuchi, H., and Harada, I. (1991) *Biochemistry 30*, 6074−6080.
52. Takeuchi, H., and Harada, I. (1986) *Spectrochim. Acta 42A*, 1069−1078.
53. Harada, I., Miura, T., and Takeuchi, H. (1986) *Spectrochim. Acta 42A*, 307−312.
54. Li, H., Hanson, C., Fuchs, J. A., Woodward, C., and Thomas, G. J., Jr. (1993) *Biochemistry 32*, 5800−5808.
55. Miura, T., Takeuchi, H., and Harada, I. (1989) *J. Raman Spectrosc. 20*, 667−671.
56. Prevelige, P. E., Jr., King, J., and Silva, J. L. (1994) *Biophys. J. 66*, 1631−1641.
57. Li, H., and Thomas, G. J., Jr. (1991) *J. Am. Chem. Soc. 113*, 456−462.

58. Thuman-Commike, P. A., Greene, B., Jakana, J., McGough, A., Prevelige, P. E., and Chiu, W. (2000) *J. Virol. 74*, 3871−3873.
59. Tuma, R., Parker, M. H., Weigele, P., Sampson, L., Sun, Y., Krishna, N. R., Casjens, S., Thomas, G. J., Jr., and Prevelige, P. E., Jr. (1998) *J. Mol. Biol. 281*, 81−94.
60. Overman, S. A., and Thomas, G. J., Jr. (1999) *Biochemistry 38*, 4018−4027.
61. Krimm, S. (1987) in *Biological applications of Raman spectroscopy* (Spiro, T. G., Ed.) pp 1−45, Wiley-Interscience, New York.
62. Overman, S. A., and Thomas, G. J., Jr. (1998) *Biochemistry 37*, 5654−5665.
63. Galisteo, M. L., Gordon, C. L., and King, J. (1995) *J. Biol. Chem. 270*, 16595−165601.
64. Yu, T.-J., Lippert, J. L., and Peticolas, W. L. (1973) *Biopolymers 12*, 2161−2176.
65. Earnshaw, W., and King, J. (1978) *J. Mol. Biol. 126*, 721−747.
66. Casjens, S., and King, J. (1974) *J. Supramol. Struct. 2*, 202−224.
67. Thuman-Commike, P. A., Greene, B., Jakana, J., Prasad, B. V., King, J., Prevelige, P. E., Jr., and Chiu, W. (1996) *J. Mol. Biol. 260*, 85−98.
68. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) *Science 262*, 1680−1685.
69. Burnett, R. M. (1985) *J. Mol. Biol. 185*, 125−143.
70. Parker, M. H., Stafford, W. F., III, and Prevelige, P. E., Jr. (1997) *J. Mol. Biol. 268*, 655−665.

BI001965Y

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 9885–9890, August 1998
Biophysics

Mechanism of capsid maturation in a double-stranded DNA virus

(virus assembly/protein folding/H/D exchange/P22 phage/Raman spectroscopy)

ROMAN TUMA*†, PETER E. PREVELIGE, JR.‡, AND GEORGE J. THOMAS, JR.*§

*Division of Cell Biology and Biophysics, School of Biological Sciences, University of Missouri, Kansas City, MO 64110-2499 and ‡Department of Microbiology, University of Alabama, Birmingham, AL 35294-0005

Communicated by Mostafa A. El-Sayed, Georgia Institute of Technology, Atlanta, GA, June 4, 1998 (received for review April 3, 1998)

ABSTRACT **Folding mechanisms of proteins incorporated within supramolecular assemblies, including viruses, are little understood and may differ fundamentally from folding mechanisms of small globular proteins. We describe a novel Raman dynamic probe of hydrogen-isotope exchange to investigate directly these protein folding/assembly pathways. The method is applied to subunit folding in assembly intermediates of the double-stranded DNA bacteriophage P22. The icosahedral procapsid-to-capsid maturation (shell expansion) of P22 is shown to be accompanied by a large increase in exchange protection of peptide β-strands. The molecular mechanism of shell expansion involves unfolding of metastable tertiary structure to form more stable quaternary contacts and is governed by a surprisingly high activation energy. The results demonstrate that coat subunit folding and capsid expansion are strongly coupled processes. Subunit structure in the procapsid represents a late intermediate along the folding/assembly pathway to the mature capsid. Coupling of folding and assembly is proposed as a general pathway for the construction of supramolecular complexes.**

The assembly of a macromolecular complex from multiple copies of protein subunits requires intersubunit interactions that are highly specific. The strength of protein interactions, which determines the stability of the assembly, is roughly proportional to the sequestered surface area (1). Accordingly, the driving force for assembly is the sequestration of hydrophobic residues at subunit interfaces (2). Conversely, exposure of hydrophobic surfaces may compete with productive folding of subunits and lead to off-pathway aggregation rather than the native supramolecular architecture. A plausible mechanism for avoiding unproductive aggregation is assembly in several discrete steps with significant changes in subunit structure at each stage of the process and attainment of the native fold only within the mature assembly. In effect, some assembly intermediates may consist of subunits in incompletely folded states. These states are likely to possess compact shapes with well-defined secondary and tertiary structures, resembling the late-folding intermediates observed for small globular proteins (2). Such folding intermediates can be distinguished by the extent of protection of their peptide NH groups against deuterium exchange (3).

Until recently, the measurement of protium/deuterium (H/D) exchanges in proteins has been limited to relatively small globular structures that are amenable to study by high-field NMR spectroscopy. Current developments in time-resolved Raman spectroscopy offer a new approach for monitoring H/D exchanges in larger protein molecules and their complexes (4–8), thus facilitating the detection and characterization of protein-folding intermediates in supramolecular assemblies. Changes in secondary structure involving as few as 2% of residues can be measured accurately and interpreted structurally by use of Raman amide I and amide III bands, as exemplified recently for purified P22 coat (9) and scaffolding (10) molecules. In the present study, we have used the Raman dynamic probe to reveal differences in the fractions of deuterated peptide groups in assembly intermediates of the *Salmonella* bacteriophage P22, an icosahedral double-stranded DNA virus (11). Corresponding changes in subunit secondary structure along the pathway of P22 assembly are also characterized.

The assembly of phage P22 constitutes a model, multistep pathway for icosahedral viral morphogenesis (11, 12). In the initial step, 420 copies of the major coat protein (gp5, 429 residues, 47 kDa) assemble into a precursor procapsid in the presence of scaffolding protein subunits (gp8, 303 residues, 33 kDa). The metastable procapsid (diameter 58 nm) is subsequently transformed into the mature capsid (diameter 63 nm) concurrent with scaffolding release and DNA packaging, a process common to many other double-stranded DNA viruses, including the herpes viruses (13, 14). The P22 capsid transformation involves extensive rearrangements in the surface lattice and a 10% increase in radius (shell expansion) (15). However, this shell maturation is accompanied by only a small change in subunit secondary structure, despite large changes in side chain environments (9). These observations suggest that domain movement involving hinge bending mediates shell expansion (9, 15). Mild denaturants may also be used to extract scaffolding protein from procapsids without altering the coat protein lattice (15–17). Subsequent expansion of the empty procapsid shell can be induced *in vitro* either by heating to 65°C or by mild detergent treatment (17, 18).

The experiments reported here extend previous spectroscopic, calorimetric, and electron cryomicroscopic investigations of the P22 shell transformation and provide new insights into the molecular mechanisms of coat subunit folding and capsid maturation.

MATERIALS AND METHODS

Procapsids were prepared using *Salmonella typhimurium* strain DB7136 infected with P22 strain 2^{-}*am H200*/13^{-}*am H101* as described (16) and further purified by Sephacryl S-1000 chromatography. Empty procapsid shells were prepared by repeated scaffolding (gp8) extraction with 0.5 M GuHCl at 4°C and sucrose gradient centrifugation.

Expanded shells were prepared from procapsid shells by heat treatment at 65°C (17). The progress of expansion was monitored by migration of particles on 1.2% agarose gels. The same protocol was used for shells assembled in D_2O. The composition and integrity of particles were examined by SDS/PAGE and 1.2% Sea-Kem agarose gels.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†Present address: Department of Microbiology, University of Alabama, Birmingham, AL 35294-0005.
§To whom reprint requests should be addressed. e-mail: thomasgj@cctr.umkc.edu.

Exchange experiments at 35°C were performed on samples twice pelleted and resuspended in isotopically pure D_2O buffer. Samples were sealed in glass capillaries and allowed to complete fast exchange at 10°C. During subsequent incubation at 35°C, samples were periodically examined by Raman spectroscopy. At the end of each experiment (average duration, 3 wk), samples were examined for structural integrity (agarose gels) and composition (SDS/PAGE). Reversibility of local unfolding was confirmed by Raman spectroscopy on control samples that were incubated in H_2O.

Subunits were refolded in D_2O buffer as described by Teschke and King (19). Refolded deuterated subunits were assembled in the presence of scaffolding protein and purified by sucrose gradient centrifugation (16). Empty procapsid shells were prepared as described above.

Spectral measurements were performed on a Spex 1877 triple spectrograph equipped with a charge-coupled device array detector. The spectra were excited at 514.5 nm. Time-resolved spectra were collected at 1-min intervals. Time-resolved Raman spectra were normalized using the 1003 cm^{-1} band of phenylalanine as internal standard. Amide III′ bands were extracted by subtraction of the spectra of protiated particles. The resulting amide III′ difference bands were normalized with respect to the fully deuterated particles. The percentages of exchanged sites were calculated from these normalized amide III′ band areas. The protocol is fully described in previous papers (4–7).

The initial rates of procapsid shell expansion were obtained from separation of heat-treated products on 1.2% agarose gel (separation buffer: 40 mM Tris, 10 mM CH_3COONa, 2 mM EDTA, pH 8.0) followed by Coomassie staining and densitometry.

RESULTS

Three Regimes for Exchange of Subunit Peptide (NH) Groups. The effects of H/D exchange of peptide NH groups on Raman spectra of procapsid shells and expanded shells are illustrated in the *Upper* and *Lower* panels, respectively, of Fig. 1. For each type of shell, the Raman amide I band is centered near 1660–1665 cm^{-1} and the companion amide III band is centered near 1243 cm^{-1} (trace A in each panel of Fig. 1). The corresponding Raman amide bands for deuterated shells, designated amide I′ and amide III′, occur at approximately 1652 and 984 cm^{-1}, respectively (trace B in each panel of Fig. 1). The positions of the amide I, amide I′, amide III, and amide III′ bands all indicate that β-strand is the dominant secondary structure of the coat protein subunit in both the procapsid shell and expanded shell assemblies, in accord with previous studies (9, 10).

We have identified three classes of exchangeable (NH → ND) peptide groups in shell subunits by varying the temperature or solvent environment at which Raman measurements are made. First, unprotected sites of shells maintained in native buffer at 10°C exchange relatively rapidly (19). These generate the Raman difference profile indicated as trace C in each panel of Fig. 1. A second class of sites, which is resistant to exchange at 10°C, can be exchanged in native buffer at 35°C via local unfolding (3). Heating to 35°C promotes partial unfolding of shell subunits without compromising the integrity of the shell assembly itself (17). This type of peptide exchange generates the Raman difference profile indicated as trace D in each panel of Fig. 1. Finally, a third class of peptide sites is resistant to exchange, even in shells containing partially unfolded (35°C) subunits. These are peptides of the exchange-protected core of the shell subunit. Amide protons of these peptides can be deuterated only by inducing shell disassembly and coat subunit unfolding in 6 M GuDCl buffer. Subsequent subunit refolding and shell reassembly in native D_2O buffer permit structural characterization of the exchange-protected core by Raman difference spectroscopy, as shown in trace E of each panel of Fig. 1. We next discuss these three classes of peptide NH → ND exchange in turn.

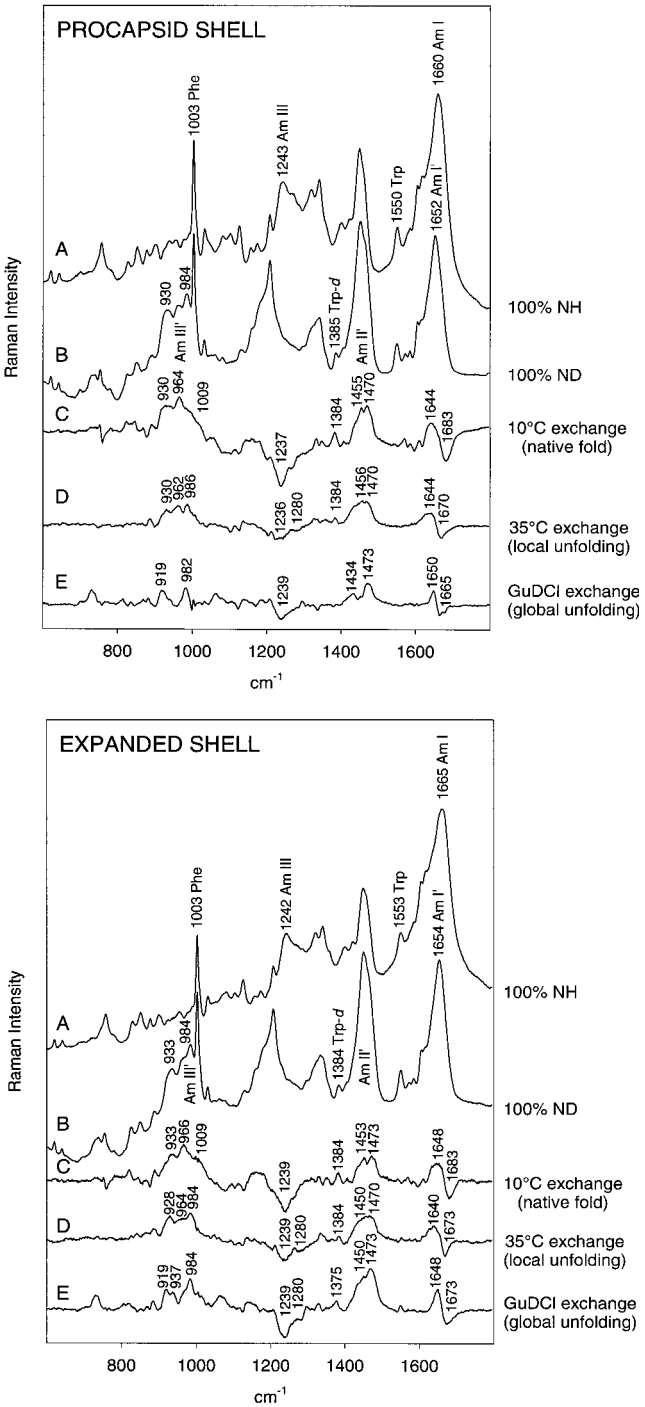


FIG. 1. Raman monitoring of deuteration of peptide NH groups of procapsid (*Upper*) and expanded shells (*Lower*). In each panel: (*A*) H_2O solution spectrum. (*B*) D_2O solution spectrum of fully deuterated particle, obtained by refolding and reassembly in native D_2O buffer of a sample previously treated with 6 M GuDCl. (*C*) Difference spectrum between *A* and particles exposed to D_2O solution for 3 h at 10°C (signature of peptides exchanging from the native state). (*D*) Difference spectrum between particles exposed to D_2O for 600 h at 35°C and particles exposed to D_2O for 3 h at 10°C (signature of peptides exchanging via local unfolding). (*E*) Difference spectrum between *B* and particles exposed to D_2O for 600 h at 35°C (signature of peptides in the exchange-protected core). Spectra were obtained from samples at 70 mg/ml concentration in 10 mM Tris buffer (pH 7.4) and 10°C.

Facile Exchange of Native Structure. Relatively rapid exchange of peptide NH groups at 10°C is limited to domains of the shell subunits that are readily accessible to solvent and probably located on the surface of the assembly. With exchange of these residues, broad Raman bands appear at 930, 964, and 1009 cm^{-1}, which are assignable to amide III′ markers of exposed peptide groups in loops, turns, and other irregular conformations (21). This conclusion is supported by the decrease of amide I intensity at 1683 cm^{-1} and concomitant increase of amide I′ intensity at 1640–1650 cm^{-1}. Facile NH → ND exchange also occurs in α-helices and is manifested by Raman amide II′ intensity appearing near 1455 cm^{-1} in difference trace C. The amide II′ intensity adds to the prominent bands of the 1440–1460 cm^{-1} interval that are due to C-H bending modes of aliphatic side chains. All exchanges from the native state are complete within 3 h of exposure to D_2O, both for procapsid shells and expanded shells. Facile NH → ND exchanges, including solvent-accessible peptides in loop, turn, and irregular conformations as well as rapidly exchanging peptides in native α-helical conformations, encompass 47 ± 3% of peptide sites in the expanded shell and 54 ± 3% of peptide sites of the procapsid shell (Fig. 2). The lower percentage of facile exchange in the expanded shell represents a small but significant increase in exchange protection upon shell expansion.

Exchange of Labile Structure by Local Unfolding. The Raman amide III′ signature of peptides exchanging via local unfolding at 35°C consists of two bands which are centered at 930 and 986 cm^{-1} in the procapsid shell (trace D, Fig. 1 *Upper*) and at 928 and 984 cm^{-1} in the expanded shell (trace D, Fig. 1 *Lower*). Intensities of these markers correlate with the intensity decreases of amide III in the interval 1230–1240 cm^{-1} and amide I at 1665–1675 cm^{-1}, both of which are characteristic of β-strand (9, 10, 22). Similar correlation has been observed for the all β-strand coat protein P3 of bacteriophage PRD1. Thus, we assign amide III′ at 929/985 ± 1 cm^{-1} to peptides in the β-strand conformation. An additional amide III′ feature at 962 cm^{-1} in the procapsid shell spectrum indicates exchange from peptides with irregular secondary structure. The 962 cm^{-1} contribution is diminished in the expanded shell spectrum, indicating that the irregular structure is either protected or already exchanged at 10°C in this assembly. The latter interpretation is consistent with the exchange-protected core comprising exclusively β-strand (see next section). In both procapsid shells and expanded shells, the exchange of peptides via local unfolding is complete within 250 h of exposure to D_2O. In the procapsid shell, 32 ± 3% of peptides exchange via local unfolding, whereas in the expanded shell only 27 ± 2% of sites are exchanged by this mechanism (Fig. 2).

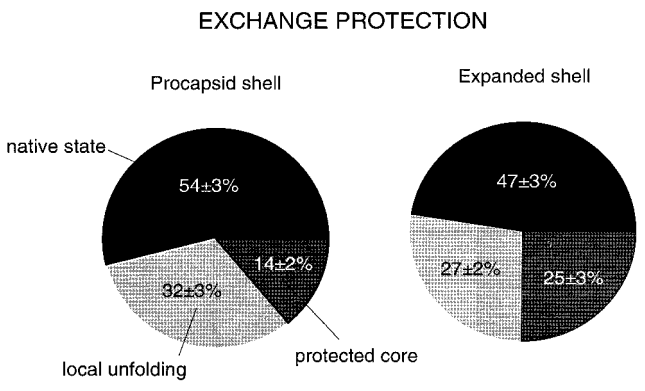


FIG. 2. Percentages of peptide NH groups exchanging in native, locally unfolded, and globally unfolded states of procapsid shells (54%, 32%, and 14%, respectively) and expanded shells (47%, 27%, and 25%, respectively). Exchange fractions were calculated from amide III′ band areas in the spectral interval 900-1020 cm^{-1}. Exchange by local unfolding is defined as exchange that occurs at 35°C but not at 10°C. Exchange by global unfolding, which requires shell disassembly and subunit denaturation in 6 M GuDCl, defines peptides of the exchange-protected core.

Exchange of the Protected Core by Global Unfolding. The Raman profile of the exchange-protected core is indicated for subunits of the procapsid shell and expanded shell in trace E of the appropriate panel of Fig. 2. In both types of shell, the Raman signature of the protected core consists of amide III′ bands near 919 and 982–984 cm^{-1}, demonstrating that the exchange-protected core consists exclusively of β-strand. However, the amide III′ intensities are markedly different in the two types of shell, revealing a striking increase in the percentage of protected β-strand NH groups accompanying expansion. This interpretation is supported by the intensities observed for other deuteration-shifted amide bands. Thus, the shifted amide I → amide I′ and amide III → amide III′ bands are considerably more intense for the expanded shell than for the procapsid shell. On the basis of the measured amide intensities in traces E of Fig. 1, we have determined that the effective exchange-protected core of the procapsid shell subunit comprises 14 ± 2% of peptide groups, whereas that of the expanded shell subunit comprises 25 ± 3% of peptide groups (Fig. 2). Thus, the apparent exchange-protected core of the shell subunit nearly doubles with expansion.

Exchanges of Subunit Side Chains. *Tryptophans.* Expansion of the P22 shell is accompanied by significant changes in environments and orientations of subunit side chains (9). The Raman spectrum is particularly sensitive to changes involving tryptophan, including differences in lability of the indole N1-H bond in procapsid shells and expanded shells. Indole N1H → N1D exchange, which serves as a gauge of tryptophan accessibility to solvent and local unfolding dynamics, is conveniently monitored by the intensity increase in the 1375–1385 cm^{-1} region of the Raman spectrum (23, 24). As seen in Fig. 1, the six tryptophans of the procapsid shell are more susceptible to native state exchange of N1H than are those of the expanded shell. We estimate that four to five Trp residues per subunit undergo facile (10°C) exchange, and another one to two exchange via local unfolding (35°C) in the procapsid shell. None of the six tryptophans of the procapsid shell is completely protected from exchange, i.e., there is no difference band near 1375–1385 cm^{-1} in trace E of the *Upper* panel of Fig. 1. Conversely, at least one and possibly two Trp residues resist exchange in the expanded shell, evidenced by the 1375 cm^{-1} difference band in trace E of the *Lower* panel of Fig. 1. Thus, one or two Trp side chains become part of the exchange-protected core upon shell expansion. This is consistent with previous data indicating a change from hydrophilic to hydrophobic local environment for the average Trp side chain following shell expansion (9).

Cysteine-405. The single cysteine residue (Cys 405) of the P22 capsid subunit resists chemical modification of its sulfhydryl group in all assemblies investigated, including the coat protein monomer, procapsid shell, and expanded shell (P.E.P., unpublished results). Further information about the Cys 405 sulfhydryl environment can be obtained from the well-resolved and relatively sharp Raman band near 2570 cm^{-1}, which is due to the S-H bond stretching vibration and is diagnostic of SH hydrogen bonding (25). The SH Raman band serves as a unique probe of local structure and dynamics (26). Fig. 3 shows that the band center differs significantly between procapsid and expanded shell forms, undergoing a shift from 2569 to 2573 cm^{-1} with expansion. The higher frequency and broader band shape associated with the expanded shell are indicative of weaker hydrogen bonding and a less hydrophobic sulfhydryl environment in the expanded shell than in the procapsid shell (25). In accord with this finding, we also observe more rapid

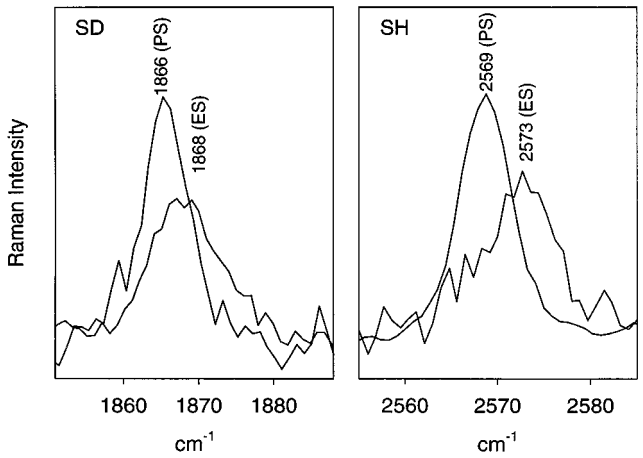


FIG. 3. Raman markers of the sulfhydryl group of Cys 405 in subunits of the procapsid shell (PS) and expanded shell (ES). (*Right*) The Raman S-H marker bands observed for H_2O solutions of shells. (*Left*) The corresponding S-D markers observed for D_2O solutions following complete SH → SD exchange.

SH → SD exchange of Cys 405 in the expanded shell than in the procapsid shell.

Although SH groups are well protected against exchange in the native procapsid ($t < 20$°C), SH → SD exchange is observed at a slightly elevated temperature, via local unfolding. For example, in the procapsid at 30°C, the sulfhydryl exchange rate constant (k_{SH}) is measured as 0.035 h^{-1} (Fig. 4 *Upper*). The activation energy of local unfolding (E_a^{local}) calculated from the temperature dependence of procapsid SH exchange over the range 20–40°C is 44 kcal·mol^{-1} (1 kcal·mol^{-1} = 4.184 kJ·mol^{-1}) (Fig. 4 *Lower*).

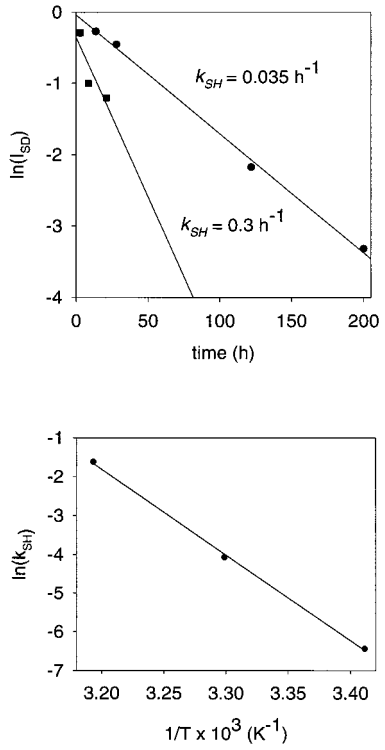


FIG. 4. (*Upper*) Rates (k_{SH}) of Cys 405 sulfhydryl exchange in subunits of procapsid shells at 30°C (●) and expanded shells at 2°C (■). Data were obtained by measuring the intensity decay of the Raman S-H band of the shell as a function of time of exposure to D_2O. (*Lower*) Arrhenius plot showing temperature dependence of the sulfhydryl exchange rate (k_{SH}) in subunits of the procapsid shell. The slope of the plot corresponds to an activation energy for local unfolding (E_a^{local}) of 44 kcal·mol^{-1}.

Conversely, for the expanded shell, SH → SD exchange is relatively facile at all temperatures. Thus, even at 2°C, we find for the expanded shell that $k_{SH} \geq 0.3$ h^{-1} (Fig. 4 *Upper*) and E_a^{local} is negligible in comparison to that of the procapsid. Interestingly, the Cys 405 sulfhydryl is less protected against exchange than is the average Trp indolyl group of the expanded shell.

Activation Energy of Shell Expansion. To estimate the energy barrier opposing heat-induced shell expansion (E_a^{expan}), we measured the temperature dependence of the expansion rate constant, k_{expan} (Fig. 5 *Upper*). The Arrhenius plot of the measured rate constants (Fig. 5 *Lower*) exhibits considerable curvature, indicating significant temperature dependence of the apparent E_a^{expan}. We find $E_a^{expan} = 40 \pm 10$ kcal·mol^{-1} in the range 40–55°C and $E_a^{expan} = 90 \pm 20$ kcal·mol^{-1} in the range 55–70°C. Thus, the apparent activation energy of shell expansion at physiological temperature (37°C) is estimated as roughly equal to that of local unfolding, the latter determined from SH exchange dynamics (above). It should be noted that the expansion activation energy, E_a^{expan}, is expressed per mole of shells (420 subunits), while the local unfolding activation energy, E_a^{local}, is expressed per mole of subunits.

DISCUSSION

The striking increase in exchange protection that is observed with shell expansion (14% vs. 25%, Fig. 2) implies a corre-

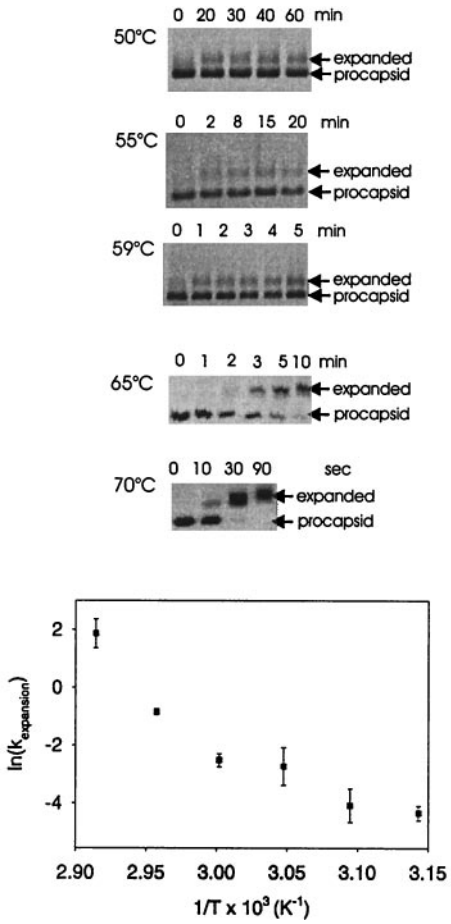


FIG. 5. (*Upper*) Kinetics of heat-induced procapsid expansion. Procapsid shells were heated at the indicated temperatures and aliquots of the reaction mixtures were withdrawn and separated on a 1.2% (wt/vol) agarose gel. Initial expansion rates (k_{expan}) were estimated from gel band intensities. (*Lower*) Arrhenius plot showing temperature dependence of the expansion rate.

sponding increase in compactness of the icosahedrally arranged coat protein subunits. Similarly large increases in exchange protection are conferred by close packing of hydrophobic side chains during folding of the cores of typical globular proteins (27). Subunits of the mature (expanded) P22 shell may be viewed as representing the final native state of the coat protein, while those of the precursor shell are analogous to a late-folding intermediate. The native coat protein conformation, i.e., the state of lowest free energy, is thus attained only within the context of the expanded shell. An energy level diagram depicting coupling of coat protein folding with shell maturation is given in Fig. 6.

The present results could be explained by a simple protection mechanism in which hydrophobic residues that are exposed in the procapsid shell become sequestered in the expanded shell. However, such sequestration seems unlikely because the procapsid shell lacks exposed hydrophobic surfaces, as judged by its resistance to the reagents ANS and bisANS (28). A scheme that involves movement of coat protein domains, as illustrated in the upper panel of Fig. 6, is deemed more likely. Here, hydrophobic residues that form an intrasubunit interface between two coat protein domains in the procapsid shell become immobilized in the expanded shell. Evidence for domain structure of the coat protein has been obtained from unfolding studies (29). Immobilization is a consequence of domain interchange between neighboring

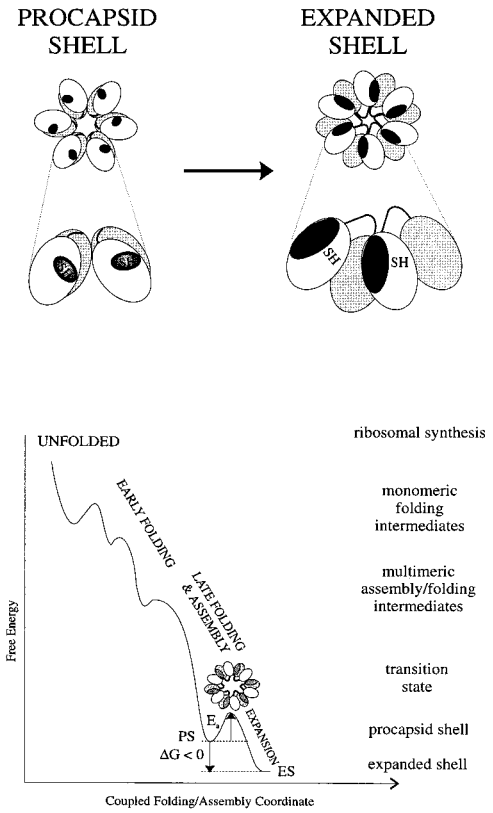


FIG. 6. (*Upper*) A mechanism accounting for increased exchange protection in the expanded shell (right) vis-à-vis the procapsid shell (left). The effective increase in the subunit exchange-protected core with expansion is due to domain interchange between neighboring subunits. The shell lattice (upper diagram) is represented as a cluster of six subunits (hexon) in which the protected core is indicated by dark shading and the two coat protein domains are shown as unshaded and lightly shaded. The enlargement (lower diagram) depicts rearrangement of contacts and domains, including Cys 405 sulfhydryls, between two neighboring subunits. (*Lower*) Energy landscape representation of coupling between subunit folding and capsid assembly in P22. A transition state containing partially exposed hydrophobic surfaces (dark gray) is proposed for the heat-induced expansion.

subunits, wherein metastable tertiary contacts of procapsid subunits are replaced by more stable quaternary contacts of the expanded shell subunits. Such a model is supported by the Raman difference spectrum between procapsid and expanded shells, which reveals many changes in side chain environments despite a small change in subunit secondary structure (9). On the basis of the Raman signature observed for the protected core in the expanded shell subunit and in view of the relatively small change in secondary structure accompanying shell expansion, we conclude that the interchanged domain is mainly β-stranded. The domain interchange model proposed here is also supported by results of electron cryomicroscopy (15). The proposed mechanism of domain interchange during capsid expansion avoids excessive exposure of hydrophobic surfaces on isolated subunits, thus averting unproductive aggregation of subunits. At the same time, the stability of the mature assembly is not compromised.

The distinctive environments and different exchange dynamics observed for the Cys 405 sulfhydryl in procapsid and expanded shells suggest that Cys 405 is located at or near the interface between interchanging domains. If the activation energy of local unfolding (E_a^{local} = 44 kcal·mol^{-1}, as measured by SH exchange) represents the free energy of domain opening, then simultaneous opening of the equivalent of one such interface would be required to account for the activation energy that triggers shell expansion at physiological temperature ($E_a^{exp} \approx$ 35–40 kcal·mol^{-1} at 37°C). This explanation is based on the assumption that shell expansion is initiated at a single (but not necessarily unique) site in each shell and is propagated rapidly through the lattice. Such a mechanism is supported by the absence of partially expanded shells or other intermediates in both *in vivo* and *in vitro* preparations.

The apparent temperature dependence of the expansion activation energy (Fig. 5 *Lower*) may stem from temperature-dependent disruption of hydrophobic interactions. This would require transient exposure of hydrophobic residues to initiate expansion (Fig. 6), which is consistent with the previously reported acceleration of shell expansion by SDS (18). Alternatively, the expansion mechanism may involve a different transition state at higher temperatures.

The present results show conclusively that subunits in the procapsid shell of P22 are less protected against amide exchange than are those of the expanded shell. The subunit main chain in the procapsid shell presumably represents a more flexible conformation, which is appropriate to achieving the transition state. The driving force required to surmount the high activation energy barrier is evidently provided *in vivo* by ATP-driven DNA packaging. The *in vitro* effects of chemical agents such as SDS and urea may facilitate particular steps in the expansion, although at the present time the exact mechanism of their action is not known.

Finally, we note that formation of a metastable intermediate and subsequent transition to a stable final state has been observed for influenza virus hemagglutinin (30), in which the most stable state is attained only upon interaction with membrane at low pH. Interestingly, an irreversible, heat-induced, capsid expansion has been detected for poliovirus and other picornaviruses (31). This expansion is thought to correspond to capsid transformation during virus entry into the host cell (31). Since the most stable fold of subunits is attained only within the context of the mature assembly or during the membrane fusion, we propose that manipulation of the subunit folding pathway could provide a means for programming morphological transformations in such supramolecular ensembles.

Support from National Institutes of Health Grants GM50776 (to G.J.T.) and GM47980 (to P.E.P.) is gratefully acknowledged.

1. Janin, J. (1995) *Biochimie* **77,** 497–505.
2. Price, N. C. (1994) in *Mechanisms of Protein Folding*, ed. Pain, R. H. (Oxford Univ. Press, New York), pp. 160–193.
3. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995) *Science* **269,** 192–197.
4. Li, T., Johnson, J. E. & Thomas, G. J., Jr. (1993) *Biophys. J.* **65,** 1963–1972.
5. Reilly, K. E. & Thomas, G. J., Jr. (1994) *J. Mol. Biol.* **241,** 68–82.
6. Tuma, R. & Thomas, G. J., Jr., (1996) *Biophys. J.* **71,** 3454–3466.
7. Tuma, R., Bamford, J. H. K., Bamford, D. H., Russell, M. P. & Thomas, G. J., Jr. (1996) *J. Mol. Biol.* **257,** 87–101.
8. Tuma, R. & Thomas, G. J., Jr., (1997) *Biophys. Chem.* **68,** 17–31.
9. Prevelige, P. E., Jr., Thomas, D., Aubrey, K. L., Towse, S. A. & Thomas, G. J., Jr. (1993) *Biochemistry* **32,** 537–543.
10. Tuma, R., Prevelige, P. E., Jr. & Thomas, G. J., Jr. (1996) *Biochemistry* **35,** 4619–4627.
11. Prevelige, P. E., Jr. & King, J. (1993) *Prog. Med. Virol.* **40,** 206–221.
12. King, J. & Chiu, W. (1997) in *Structural Biology of Viruses*, eds. Chiu, W., Burnett, R. M. & Garcea, R. (Oxford Univ. Press, New York), pp. 1–30.
13. Newcomb, W. W., Homa, F. L., Thomsen, D. R., Booy, F. P., Trus, B. L., Steven, A. C., Spencer, J. V. & Brown, J. C. (1996) *J. Mol. Biol.* **263,** 432–446.
14. Trus, B. L., Booy, F. P., Newcomb, W. W., Brown, J. C., Homa, F. L., Thomsen, D. R. & Steven, A. C. (1996) *J. Mol. Biol.* **263,** 447–462.
15. Prasad, B. V. V., Prevelige, P. E., Marietta E., Chen, R. O., Thomas, D., King, J. & Chiu, W. (1993) *J. Mol. Biol.* **231,** 65–74.
16. Prevelige, P. E., Jr., Thomas, D. & King, J. (1988) *J. Mol. Biol.* **202,** 743–757.
17. Galisteo, M. L. & King, J. (1993) *Biophys. J.* **65**, 227–235.
18. Earnshaw, W., Casjens, S. & Harrison, S. C. (1976) *J. Mol. Biol.* **104,** 387–410.
19. Teschke, C. M. & King, J. (1993) *Biochemistry* **32,** 10839–10847.
20. Finucane, M. D. & Jardetzky, O. (1996) *Protein Sci.* **5,** 653–662.
21. Sugawara, J., Harada, I., Matsuura H. & Shimanouchi, T. (1978) *Biopolymers* **17,** 1405–1421.
22. Chen, M. C. & Lord, R. C. (1974) *J. Am. Chem. Soc.* **96,** 4750–4752.
23. Miura, T., Takeuchi, H. & Harada, I. (1988) *Biochemistry* **27,** 88–94.
24. Miura, T., Takeuchi, H. & Harada, I. (1989) *J. Raman Spectrosc.* **20,** 667–671.
25. Li, H. & Thomas, G. J., Jr. (1991) *J. Am. Chem. Soc.* **113,** 456–462.
26. Li, H., Hanson, C., Fuchs, J. A., Woodward, C. & Thomas, G. J., Jr. (1993) *Biochemistry* **32,** 5800–5808.
27. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. (1993) *Science* **262,** 1680–1685.
28. Teschke, C. M., King, J. & Prevelige, P. E., Jr. (1993) *Biochemistry* **32,** 10658–10665.
29. Prevelige, P. E., Jr., King, J. & Silva, J. L. (1994) *Biophys. J.* **66,** 1631–1641.
30. Carr, C. M., Chaudhry, C. & Kim, P. S. (1997) *Proc. Natl. Acad. Sci USA* **94,** 14306–14313.
31. Chow, M., Basavappa, R. & Hogle, J. M. (1997) in *Structural Biology of Viruses*, eds. Chiu, W., Burnett, R. M. & Garcea, R. (Oxford Univ. Press, New York), pp. 1–30.

doi:10.1006/jmbi.2000.4383 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **306**, 389–396

JMB

COMMUNICATION

Hydrogen-deuterium Exchange as a Probe of Folding and Assembly in Viral Capsids

Roman Tuma[1], Lori U. Coward[2], Marion C. Kirk[2], Stephen Barnes[2,3] and Peter E. Prevelige Jr[1]*

[1]*Department of Microbiology, The University of Alabama at Birmingham, Birmingham AL 35294-2170, USA*

[2]*Mass Spectrometry Shared Facility and Comprehensive Cancer Center, The University of Alabama at Birmingham Birmingham AL 35294-0019, USA*

[3]*Department of Pharmacology & Toxicology, The University of Alabama at Birmingham Birmingham AL 35294-0019, USA*

**Corresponding author*

The dynamics of proteins within large cellular assemblies are important in the molecular transformations that are required for macromolecular synthesis, transport, and metabolism. The capsid expansion (maturation) accompanying DNA packaging in the dsDNA bacteriophage P22 represents an experimentally accessible case of such a transformation. A novel method, based on hydrogen-deuterium exchange was devised to investigate the dynamics of capsid expansion. Mass spectrometric detection of deuterium incorporation allows for a sensitive and quantitative determination of hydrogen-deuterium exchange dynamics irrespective of the size of the assembly. Partial digestion of the exchanged protein with pepsin allows for region-specific assignment of the exchange. Procapsids and mature capsids were probed under native and slightly denaturing conditions. These experiments revealed regions that exhibit different degrees of flexibility in the procapsid and in the mature capsid. In addition, exchange and deuterium trapping during the process of expansion itself was observed and allowed for the identification of segments of the protein subunit that become buried or stabilized as a result of expansion. This approach may help to identify residues participating in macromolecular transformations and uncover novel patterns and hierarchies of interactions that determine functional movements within molecular machines.

Keywords: hydrogen/deuterium exchange; virus; capsid; protein folding; dynamics

Most of the biochemical reactions within a cell do not occur through the random collision of protein molecules but rather within well defined, large (ten or more protein subunits), macromolecular complexes. These complexes, which form by self-assembly from constituent macromolecules, are efficient molecular machines that undergo many duty cycles per second. During a typical duty cycle, the subunits of the assembly often rearrange and undergo a series of conformational changes.[1] For high processivity, the rearrangements must take place within the intact assembly and therefore require considerable mobility of domains and subunits. Such motions occur *via* the formation of high energy transition state ensembles of intact assemblies, and the formation of these transition states is in turn determined by the strength of local interactions within and between subunits. To understand the function of such an assembly therefore requires an identification of the regions undergoing transformation, elucidation of the pathway of transformation, and characterization of the energetic barriers to transformation. As originally pioneered by Englander and co-workers,[2,3] functional labeling by hydrogen/deuterium exchange (HDX) studies provide a means to obtain the required information. Advances in the use of mass spectrometry to analyze exchange patterns provides the opportunity to utilize this technique to obtain region-specific rather than global information.

Present address: R. Tuma, Department of Biosciences and Institute of Biotechnology, Biocenter 2, P.O. Box 56, University of Helsinki, FIN-00014, Finland.

Abbreviations used: HDX, hydrogen deuterium exchange; MALDI, matrix assisted laser desorption ionization; TOF, time of flight.

E-mail address of the corresponding author: prevelig@uab.edu

0022-2836/01/030389–8 $35.00/0

Maturation of bacteriophage P22 is one of the best-characterized transitions within large macromolecular complexes.[4–7] P22 is a representative member of the class of dsDNA viruses that assemble in two steps. In the first step, P22 coat protein (gp5, 46,700 Da) assembles with the assistance of approximately 300 molecules of scaffolding protein into a $T = 7$ procapsid (420 subunits, 19.6 MDa) that serves as a precursor for DNA packaging (Figure 1). In the second step, triggered *in vivo* by DNA packaging, the scaffolding protein is released, the procapsid lattice expands, and as a consequence of expansion, the capsid is stabilized.[4] Expansion results in a significant morphological change.[8] Raman spectroscopy, which can provide a global (but not region-specific) measurement of HDX, has revealed that expansion results in a large increase in exchange protection with only a small change in secondary structure.[6] Capsid expansion can be triggered *in vitro* at elevated temperature (55-65 °C) and is governed by large, temperature-dependent, activation energy[4,6] (Figure 1).

It has been proposed that expansion is mediated by domain interchange between neighboring subunits (Figure 1).[6,7,9] Such a large-scale rearrangement requires a transient disruption of interdomain interactions followed by a swiveling motion of whole domains around a flexible hinge. Although the hinge region has been mapped by limited proteolysis to the middle of the protein (residues 180-205)[10] it is not known which residues constitute the invariant interfaces, nor is it known which regions dissociate to allow for the domain movement. The large difference in HDX protection between the procapsid and the mature capsid suggests that the regions involved in the transformation are less protected in the procapsid, and thus that HDX could be used to identify these regions. Additionally, the regions which become transiently exposed to solvent during expansion may be identified by labeling during expansion.

NMR spectroscopy has traditionally been used for high resolution mapping of HDX.[11] However, due to the slow tumbling rates and spectral complexity of large complexes, this technique is limited to relatively small proteins. Recently, mass spectrometry (MS)-based techniques for monitoring site-specific HDX have been developed.[12–15] This

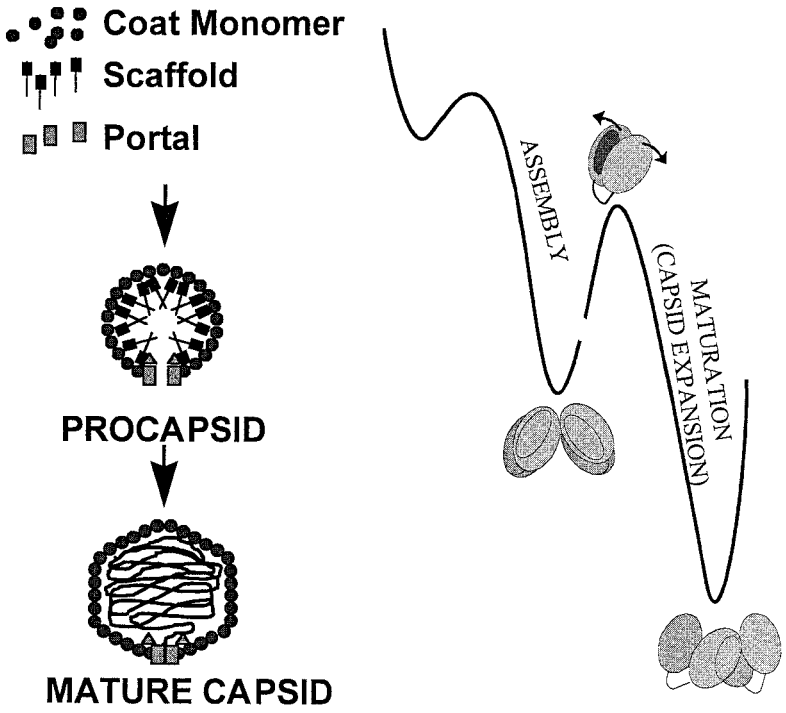


Figure 1. Left: The assembly pathway of bacteriophage P22. Monomeric coat protein assembles with the assistance of scaffolding and portal proteins to form a metastable $T = 7$ procapsid in which the scaffolding protein is contained internally and a portal protein complex is located at one of the icosahedral vertices. Upon DNA packaging through the portal vertex, the scaffolding protein exits, and the lattice expands to become the mature $T = 7$ capsid. The samples used in this study correspond to procapsids from which the scaffolding has been extracted, and expanded shells, analogs of the mature form obtained by heating procapsid like particles.[4] Right: Assembly and maturation of P22 coat protein is superimposed on a hypothetical energy landscape. Domain interchange is proposed as the mechanism of capsid expansion and is illustrated schematically. Domains of the subunit which form internal contacts in the procapsid (buried shaded areas) open up in the transition state (exposed shaded areas) and then form new contacts with their neighbors in the expanded form (buried, swapped, shaded areas). The transition may be mediated by an ensemble of transition states in which the interface between domains has been disrupted allowing for mobility within the assembly.

approach, which is not in principle limited by the size of the assembly, has proven useful for mapping residues constituting subunit interfaces in oligomeric proteins[16–19] as well as to study chaperone-assisted protein folding.[13]

In this study, we demonstrate the utility of HDX and MS for the study of protein dynamics in large assemblies. Deuterium labeling during expansion was employed to identify regions within domain interfaces involved in the transformation from procapsid to capsid. Our results demonstrate that HDX combined with MS can identify regions involved in transformation of large biological complexes.

Measurement of deuterium exchange in a viral capsid

A method, similar to that described,[20] which relies on exchange of the sample under a defined set of conditions, chemical quenching of the back exchange at low pH, rapid digestion of sample under acidic conditions into a reproducible set of short (10-15 amino acid residues) peptides (Figure 2(a)), and determination of the mass distribution by MALDI-TOF (Figure 2(c)), was employed. HDX of amide protons results in a shift of the fragment mass distribution towards higher masses; because side-chain protons back-exchange rapidly during digestion in H_2O they do not contribute to the mass shift. In a single experiment it is possible to obtain mass distribution information for many peptide fragments.

P22 capsids proved to be stable even at pH 2.5, and it was necessary to disrupt the viral capsids prior to digestion by inclusion of a guadinidine hydrochloride (GuHCl) dissociation step. This step was only necessary for efficient digestion of the stable mature shells. For the less stable procapsid shells, identical results were obtained with and without the dissociation step, indicating that the GuHCl treatment did not alter the observed exchange patterns (data not shown). The back-exchange during dissociation and digestion was estimated to be between 20-30 % by comparison of the observed and expected centroids for the fully deuterated control. The results in Figure 2 show that both good signal-to-noise ratio and sufficient resolution can be achieved even in the presence of residual salts (~50 mM total) in the blotting matrix. This makes the MALDI interface preferable over the electrospray interface for the direct analysis of samples. Nine strong peaks in the MALDI were selected for assignment by electrospray MS:MS. These nine fragments cover approximately 28 % of the primary sequence (Figure 2(b)).

The three panels of Figure 2(c) show that region-specific monitoring of HDX is possible in a large macromolecular assembly irrespective of its size, and compare the HDX profiles for folding and assembly intermediates. Although the viral capsid employed in this study is composed primarily of the coat protein (with minor contribution of the portal protein) the same protocol may be used for assemblies containing several protein species. For more complex assemblies, such as ribosomal subunits or the nuclear pore, overlapping fragments may complicate the analysis, and a rapid separation of the peptide digest and the more elaborate electrospray methodology would be preferred over the simpler MALDI protocol.[21]

Detection of subunit-subunit interface rearrangements due to capsid transformation

Changes in the extent and stability of intersubunit interactions during assembly and maturation will result in changes in the degree of HDX protection within the regions involved, and these changes will be evidenced as shifts in the mass distribution of the peptide fragments. Structural elements, which remain relatively constant through assembly, should display unchanged exchange protection. For example, the regions comprising the folding core of the subunit are expected to exhibit similar degrees of protection in the monomer, procapsid, and mature virion forms. A good candidate for a component of the exchange-protected core are the residues 415-430 (Figure 2(c), middle). In comparison with unexchanged and fully exchanged controls, this region shows a high degree of exchange protection in all three forms (monomer, procapsid, and expanded shell). These data are consistent with the observation that the single cysteine residue 415 is buried in both the monomer and procapsid structure.[6,22]

In contrast, regions that comprise interfaces between subunits generated during assembly are expected to show little protection in the monomer and increased protection in the procapsid. Peptides that have their origin in a previously identified flexible hinge region display this behavior (Figure 3, peptides 182-188, 189-195). Recent spectroscopic results have shown that about 20 residues undergo folding during assembly.[22] The present results indicate that the hinge is the likely candidate for the folding region.

The regions that are directly involved in expansion are likely to become more protected in the expanded state. The degree of protection of the assigned peptides was determined for the procapsid and expanded shells under conditions that promote exchange *via* local unfolding (35 °C) (Figures 2 and 3). These conditions are best suited to highlight the differences in protection between the two states.[6] Figure 2(c), left and right, provide examples of peptides (residues 196-208, 156-168) that become protected upon expansion. For the peptide 196-208, the mass distribution for the procapsid is similar to that of the fully exchanged control, suggesting little protection for these species. Peptide156-168 is protected relative to the fully exchanged form, but the protection in the procapsid is similar to that in the monomer. Presumably, this is protection due to folding. In both cases, the mass envelope obtained for the peptide from the expanded shell is signifi-

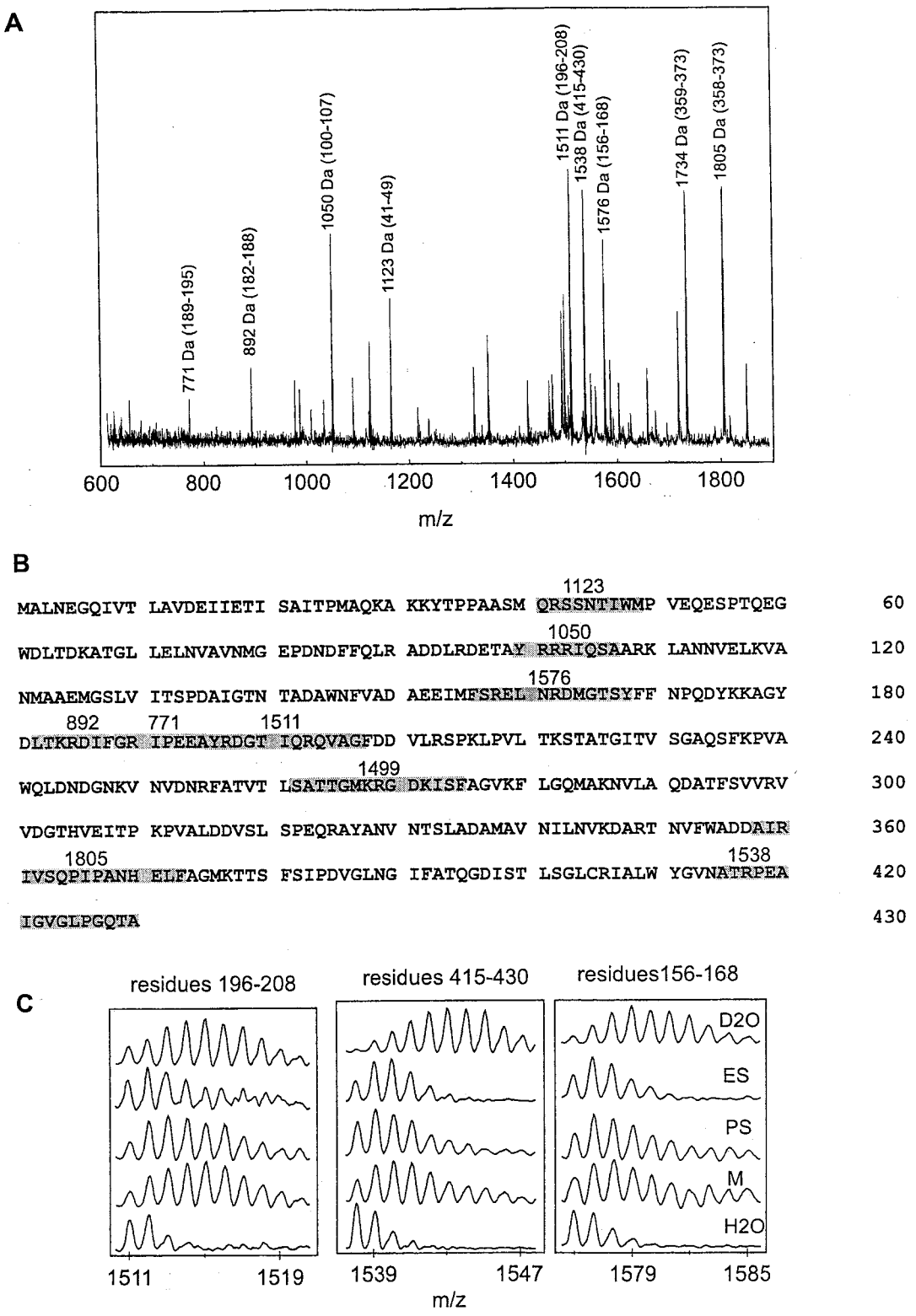


Figure 2 (*legend shown opposite*)

cantly narrower and on average shifted towards lower mass than that in the procapsid. This indicates increased protection as a result of capsid expansion. The observed protection is in accordance with the location of peptide 196-208 within the flexible hinge region that becomes protease-resistant in the expanded shell (residues 180-205).[10] Similarly, other regions within and near the hinge

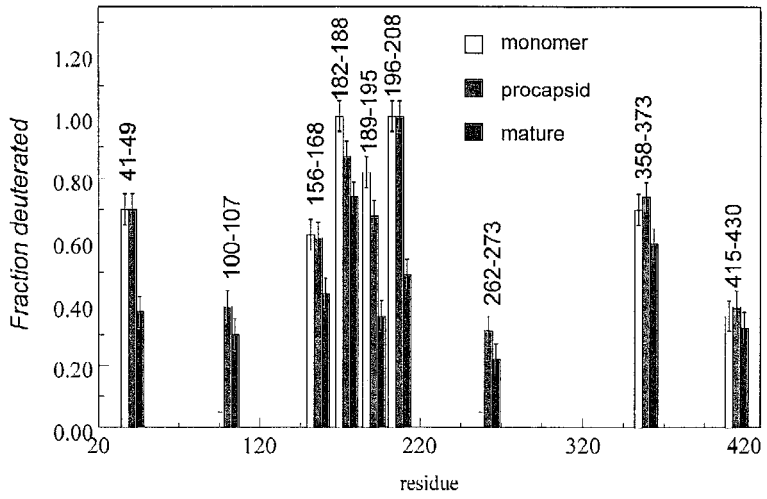


Figure 3. Deuterium exchange of coat protein subunit in different assembly states determined as described in Figure 2. Bar graph above each sequence region indicates fraction of deuterated peptides in the monomer (after 24 hours, 4 °C, yellow), procapsid (after 100 hours, 35 °C, red), expanded shell (after 100 hours, 35 °C, blue). The data represent the average and standard deviations obtained from four independent experiments.

(182-188,189-195) exhibit less protection in the procapsid than in the mature shell (Figure 3). Several regions (41-49, 156-168) become significantly more protected upon expansion (summarized in Figure 3). A slightly increased protection was further observed for residues 100-107, 262-273 and 358-373. The low magnitude of these changes, and the fact that they are distributed throughout the protein, suggests that they arise from the global stabilization which is known to accompany expansion.[4]

Interestingly, we have as yet been unable to identify a single contiguous region that would remain completely unprotected in all three assembly states and thus would represent a surface epitope. This may in part be due to incomplete "coverage" of the coat sequence in the peptide fragments. On the other hand, none of the peptides for which exchange was reported shows 100 % protection. Thus, it is conceivable that a stable viral capsid assumes a polypeptide fold in which short exposed epitopes are interspersed within regions that are stabilized by cooperative interactions. For

Figure 2. (a) MALDI-TOF of peptic digest of P22 coat protein. The amino acid residue numbers defining the fragment are indicated in parenthesis. Identification of the peptides was carried out in separate MS/MS experiments using a PE Sciex (Concorde, Ontario, Canada) API III triple quadrupole mass spectrometer with an electrospray ionization interface. The molecular ions of peptides identified in the MALDI-TOF analysis were subject to collision with Ar/N_2 (90 %:10 %) gas to generate daughter fragment ion spectra. Proteins and assemblies were prepared and purified as described previously.[6,10] $H_2O/^2H_2O$ buffer changes were achieved using BioGel P-6 desalting spin columns which were pre-equilibrated three times with the desired H_2O or ^2H_2O buffer (50 mM sodium phosphate, 25 mM NaCl, pH 7.5). Exchange was initiated by a transfer of the sample into ^2H_2O and incubation at the desired temperature (4 and 35 °C). Because mature shells were stable even at pH 2.5, all assemblies were subject to dissociation in a buffer (10 mM ammonium phosphate, pH 2.5) containing 3 M guanidine hydrochloride. After dilution 1:10 into 10 mM ammonium phosphate (pH 2.5) the dissociated subunits were digested with agarose-immobilized pepsin (0 °C, two minutes, 200 units of Sigma insoluble pepsin). The resulting solution of peptides was diluted 1:10 into matrix solution (0.1 % trifluoroacetic acid in 50 % v/v acetonitrile/water, saturated with α-hydroxycinnamic acid) and spotted onto a chilled gold-coated MALDI plate. Masses of the peptides were determined using a Perseptive Biosystems (Framingham, MA) Voyager Elite MALDI-TOF mass spectrometer in the reflector mode (mass resolution approximately 1:2000) using a pulsed nitrogen laser operating at 337 nm. Each spectrum represented data accumulated from up to 100 laser shots. (b) The amino acid sequence of bacteriophage P22 coat protein. The shaded regions represent peptides resulting from pepsin digestion and assigned by MS:MS. The numbers above the shaded regions correspond to the mass of the peptide as determined in (a). (c) Mass spectra of three peptides corresponding to different coat protein regions. The mass distribution for exchanged forms is shifted towards higher mass due to the replacement of the amide protons by deuterons. In each panel the top and bottom traces represent mass spectrum of fully exchanged and non-exchanged controls, respectively. The two controls were used in quantitative analysis for determination of the fraction of deuterated peptide groups according to: $d = (c - n)/(f - n)$, where c is the centroid of the isotopic distribution of the peptide peak, n and f are the centroids of the fully protonated and the fully deuterated control, respectively. The centroid was calculated as a weighted average of positions of five strongest peaks in the isotopic distribution. The centroid reproducibility was ±0.5 Da between different experiments. The second and third traces from top (ES and PS, respectively) represent endpoint of exchange *via* local unfolding in expanded shells and procapsid shells, respectively (100 hours in ^2H_2O at 35 °C[6]). The second trace from the bottom (M) represents the endpoint of exchange for the native state of coat protein monomer (24 hours, 4 °C).

P22 and other dsDNA bacteriophages this feature is consistent with great stability of the mature capsid shell and resistance to protease cleavage.[4,10,23] Therefore, it may not be possible to use the difference in the protection of the monomeric subunit and the assembled subunit to identify subunit-subunit interfaces directly, as demonstrated for smaller, less stable and reversible, protein oligomers such as cAMP-dependent protein kinase/inhibitor and thrombin/thrombomodulin binding.[19] In order to determine subunit association within large stable assemblies, for which subunit folding and assembly are often coupled, characterization of the energetics of local unfolding may be required.[24]

The use of H/^{2}H exchange to follow dynamic transformations

While structural studies can characterize stable states, often in great detail, and thereby suggest pathways for structural transformations, characterizing the actual pathways of these transformations remains a significant challenge. In principle, HDX techniques should be applicable to this problem in much the same way they have been used to characterize protein folding pathways. As a proof of principle, we have pulse-labeled procapsid shells during the process of expansion and examined the pattern of deuterium retention after back-exchange. We use exchange of the peptide corresponding to the hinge region (residues 196-208) to illustrate dynamic behavior during expansion (Figure 4).

Procapsids were incubated in ^2H_2O at 68 °C for 30 seconds, conditions under which approximately 40 % of the shells expand (rate of expansion at 68 °C is one min^{-1}),[6] and the distribution of deuterium in the peptide corresponding to amino acid residues 196-208 was determined (Figure 4(a), trace II). Pre-expanded shells were incubated in parallel as a control (Figure 4(b), trace II). Within the pro-

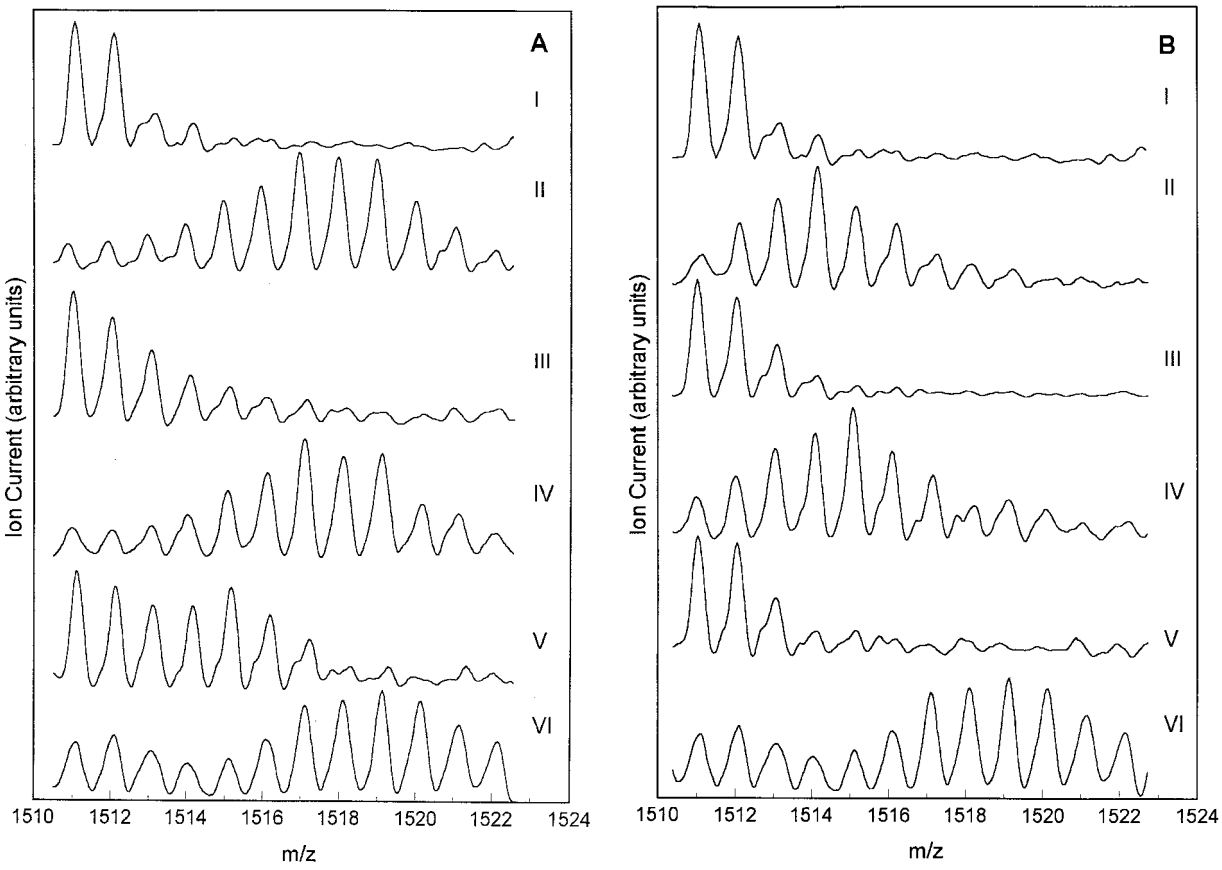


Figure 4. Deuterium labeling during procapsid shell expansion at 68 °C. (a) (I) Mass spectra of protonated control; (II) procapsids labeled at 68 °C for 30 seconds during which 40 % shells expanded (the extent of expansion was assessed by agarose gel electrophoresis);[6] (III) procapsids labeled at 68 °C for 30 seconds and back-exchanged in H_2O for 2.5 hours at 35 °C; (IV) procapsid shells were heated at 68 °C in ^2H_2O for ten minutes during which 70 % shells expanded; (V) procapsid shells were heated at 68 °C in ^2H_2O for ten minutes and back-exchanged in H_2O for ten minutes at 68 °C; (VI) fully deuterated control. (b) The same as (a) but expanded shells were used instead of procapsids. The fully deuterated control spectra for both procapsid and pre-expanded shells were superimposable and were averaged to improve the signal-to-noise ratio, Note that the appearance of bimodality in the fully deuterated sample ((a) and (b), trace VI) is due to the overlap of the deuterated tail of the m/z 1499 peak, This does not affect the conclusions.

capsid, the peptide became almost fully deuterated whereas in the expanded shell only approximately half of the NH sites were exchangeable. The fact that a similar degree of labeling can be achieved in the procapsids at 35 °C (a temperature where there is no expansion[6]) (Figure 3) suggests that the deuterium is likely incorporated before the expansion. To determine if this exchange was reversible, both samples were then back-exchanged at 35 °C for 2.5 hours (at 35 °C no detectable expansion is observed,[6] and data not shown). For the pre-expanded shells, the forward exchange was completely reversible (compare Figure 4(b) traces I and III). For the procapsid-derived samples, the reaction was not fully reversible, some deuterons were retained as evidenced by the peaks observed between m/z 1515-1519 (Figure 4(a), III). In contrast, when the forward exchange was performed at 35 °C (Figure 2(c)) it was fully reversible (not shown). Thus, the retained deuterons were stabilized by the expansion and face a higher activation barrier for back-exchange (Figure 1).

More extensive incubation at 68 °C for ten minutes results in the continued expansion of procapsids until they comprise 70 % of the population. Once again, it is evident that the pre-expanded shells are more resistant to forward exchange than the procapsid-derived samples and upon back-exchange at 68 °C the pre-expanded shells display complete reversibility (compare Figure 4(b) traces I and V). However, for the procapsid-derived samples approximately half the initial amount of deuterium was retained as a second peak centered at around m/z 1515 in the overall envelope (Figure 4(a), trace V). We interpret these data as follows: the 196-208 region is exposed prior to expansion. Upon expansion, approximately one half of the amino acid residues in this fragment become fully protected against exchange and retain deuterium, while the other half continue to exchange freely. This confirms the existence of a large increase in the activation barrier for local exchange upon expansion affecting approximately half of the residues within the hinge. The high activation barrier is a result of higher stability of this region in the mature shell as depicted on the energy landscape in Figure 1. Flexibility within the hinge region is required for expansion.[10] Furthermore when the mobility of the loop was increased by proteolytic cleavage expansion was greatly accelerated.[10] The use of $H/^2H$ exchange permits better mapping of the hinge region dynamics.

Several other regions (residues 41-49 and 156-168) also became labeled during expansion and retained deuterium (data not shown). The retention of deuterium reflects the increased energy barrier for exchange that results from stabilization of the expanded state. On the other hand, as would be expected if the C-terminal region was a component of the constant folding core, none of the C-terminal regions trapped deuterium during expansion.[6]

Implications for dynamics of viruses and other assemblies

A viral capsid undergoes a series of controlled structural transformations during the viral life cycle and as such represents a dynamic protein assembly. This dynamic character is critical to their function, and mutations which alter the stability of the capsid frequently have a deleterious effect on viral replication.[25] Stabilization of the capsid structure, thereby preventing uncoating, is in fact the mechanism of action of one class of antiviral agents.[26] However, investigating the progressive alterations in backbone dynamics as the proteins comprising viral capsids fold, assemble, and function has proven to be a challenge. Proteolytic cleavage and mass spectrometry has been used to detect flexible regions,[10,25] and HDX studies by Raman spectroscopy have been used to gain an overall estimate of exchange kinetics and protection.[6,27,28] The use of HDX and mass spectrometry extends these studies and allows for determination of flexibility throughout the polypeptide backbone. The present method exploited the irreversibility of the procapsid-to-capsid transition to probe the expansion dynamics. Although molecular machines undergo cyclic changes to return to the identical initial state, individual steps in the cycle are usually coupled to an irreversible chemical reaction (e.g. ATP hydrolysis) and thus the complex moves along an energy landscape similar to that of the P22 capsid (Figure 1). Thus, the methodology presented here could be applied to more complex molecular machines.

Acknowledgments

These studies were supported by NSF grant DBI-9726698 (P.E.P). The mass spectrometer was purchased by funds from a NIH Instrumentation Grant (S10RR11329) and from this institution. Operation of the UAB Mass Spectrometry Shared Facility has been supported in part by a NCI Core Research Support Grant (P30 CA13148-27) to the University of Alabama at Birmingham Comprehensive Cancer Center.

References

1. Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, **92**, 291-294.
2. Englander, S. W. & Rolfe, A. (1973). Hydrogen exchange studies of respiratory proteins. 3. Structural and free energy changes in hemoglobin by use of a difference method. *J. Biol. Chem.* **248**, 52-61.
3. Englander, S. W., Englander, J. J., McKinnie, R. E., Ackers, G. K., Turner, G. J., Westrick, J. A. & Gill, S. J. (1992). Hydrogen exchange measurement of the free energy of structural and allosteric change in hemoglobin. *Science*, **256**, 1684-1687.
4. Galisteo, M. L. & King, J. (1993). Conformational transformations in the protein lattice of phage P22 procapsids. *Biophys. J.* **65**, 227-235.

5. Prevelige, P. E., Jr & King, J. (1993). Assembly of bacteriophage P22: a model for ds-DNA virus assembly. *Prog. Med. Virol.* **40**, 206-221.
6. Tuma, R., Prevelige, P. E., Jr & Thomas, G. J., Jr (1998). Mechanism of capsid maturation in a double-stranded DNA virus. *Proc. Natl Acad. Sci. USA,* **95**, 9885-9890.
7. Zhang, Z., Greene, B., Thuman-Commike, P. A., Jakana, J., Prevelige, P. E., Jr, King, J. & Chiu, W. (2000). Visualization of the maturation transition in bacteriophage P22 by electron cryomicroscopy. *J. Mol. Biol.* **297**, 615-626.
8. Prasad, B. V., Prevelige, P. E., Marietta, E., Chen, R. O., Thomas, D., King, J. & Chiu, W. (1993). Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. *J. Mol. Biol.* **231**, 65-74.
9. Prevelige, P. E., Jr, Thomas, D., Aubrey, K. L., Towse, S. A. & Thomas, G. J., Jr (1993). Subunit conformational changes accompanying bacteriophage P22 capsid maturation. *Biochemistry,* **32**, 537-543.
10. Lanman, J. K., Tuma, R. & Prevelige, P. E., Jr (1999). Domain structure of the P22 coat protein. *Biochemistry,* **38**, 14614-14623.
11. Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. (1996). Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **6**, 18-23.
12. Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T. & Dobson, C. M. (1993). Detection of transient protein folding populations by mass spectrometry. *Science,* **262**, 896-900.
13. Miranker, A., Robinson, C. V., Radford, S. E. & Dobson, C. M. (1996). Investigation of protein folding by mass spectrometry. *FASEB J.* **10**, 93-101.
14. Wagner, D. S. & Anderegg, R. J. (1994). Conformation of cytochrome c studied by deuterium exchange-electrospray ionization mass spectrometry. *Anal. Chem.* **66**, 706-711.
15. Zhang, Z. & Smith, D. L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **2**, 522-531.
16. Deng, Y. & Smith, D. L. (1998). Identification of unfolding domains in large proteins by their unfolding rates. *Biochemistry,* **37**, 6256-6262.
17. Deng, Y. & Smith, D. L. (1999). Hydrogen exchange demonstrates three domains in aldolase unfold sequentially. *J. Mol. Biol.* **294**, 247-258.
18. Deng, Y., Zhang, Z. & Smith, D. L. (1999). Comparison of continuous and pulsed labeling amide hydrogen exchange/mass spectrometry for studies of protein dynamics. *J. Am. Soc. Mass Spectrom.* **10**, 675-684.
19. Mandell, J. G., Falick, A. M. & Komives, E. A. (1998). Identification of protein-protein interfaces by decreased amide proton solvent accessibility. *Proc. Natl Acad. Sci. USA,* **95**, 14705-14710.
20. Mandell, J. G., Falick, A. M. & Komives, E. A. (1998). Measurement of amide hydrogen exchange by MALDI-TOF mass spectrometry. *Anal. Chem.* **70**, 3987-3995.
21. Smith, D. L., Deng, Y. & Zhang, Z. (1997). Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J. Mass Spectrom.* **32**, 135-146.
22. Tuma, R., Tsuruta, H., Benevides, J. M., Prevelige, P. E., Jr & Thomas, G. J., Jr (2001). Characterization of subunit structural changes accompanying assembly of the bacteriophage P22 procapsid. *Biochemistry,* **40**, 665-674.
23. Jardine, P. J., McCormick, M. C., Lutze-Wallace, C. & Coombs, D. H. (1998). The bacteriophage T4 DNA packaging apparatus targets the unexpanded prohead. *J. Mol. Biol.* **284**, 647-659.
24. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science,* **269**, 192-197.
25. Bothner, B., Dong, X. F., Bibbs, L., Johnson, J. E. & Siuzdak, G. (1998). Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J. Biol. Chem.* **273**, 673-676.
26. Diana, G. D., McKinlay, M. A. & Treasurywala, A. (1997). The use of structural information in the design of picornavirus capsid binding agents. In *Structural Biology of Viruses* (Chiu, W., Burnett, R. M. & Garcea, R., eds), pp. 157-186, Oxford University Press, New York.
27. Tuma, R., Bamford, J. H. K., Bamford, D. H., Russell, M. P. & Thomas, G. J., Jr (1996). Structure, interactions and dynamics of PRD1 virus. I. Coupling of subunit folding and capsid assembly. *J. Mol. Biol.* **257**, 87-101.
28. Tuma, R. & Thomas, G. J., Jr (1997). Mechanisms of virus assembly probed by Raman spectroscopy: the icosahedral bacteriophage P22. *Biophys. Chem.* **68**, 17-31.

Edited by C. R. Matthews

(Received 29 August 2000; received in revised form 4 October 2000; accepted 21 November 2000)

Available online at www.sciencedirect.com
SCIENCE DIRECT
Virology 351 (2006) 73–79
VIROLOGY
www.elsevier.com/locate/yviro

Interaction of packaging motor with the polymerase complex of dsRNA bacteriophage

Jiří Lísal [a], Denis E. Kainov [a,1], TuKiet T. Lam [b,2], Mark R. Emmett [b], Hui Wei [c], Paul Gottlieb [c], Alan G. Marshall [b], Roman Tuma [a,*]

[a] *Institute of Biotechnology and Department of Biological and Environmental Sciences, University of Helsinki, Viikinkaari 1 PL 65, 00014 Helsinki, Finland*
[b] *National High Magnetic Field Laboratory and Department of Chemistry and Biochemistry, 1800 E. Paul Dirac Dr., Florida State University, Tallahassee, FL 32310, USA*
[c] *Department of Microbiology and Immunology, The Sophie Davis School of Biomedical Education, The City College of New York, New York, NY 10031, USA*

Received 9 January 2006; returned to author for revision 15 February 2006; accepted 12 March 2006
Available online 27 April 2006

Abstract

Many viruses employ molecular motors to package their genomes into preformed empty capsids (procapsids). In dsRNA bacteriophages the packaging motor is a hexameric ATPase P4, which is an integral part of the multisubunit procapsid. Structural and biochemical studies revealed a plausible RNA-translocation mechanism for the isolated hexamer. However, little is known about the structure and regulation of the hexamer within the procapsid. Here we use hydrogen–deuterium exchange and mass spectrometry to delineate the interactions of the P4 hexamer with the bacteriophage phi12 procapsid. P4 associates with the procapsid via its C-terminal face. The interactions also stabilize subunit interfaces within the hexamer. The conformation of the virus-bound hexamer is more stable than the hexamer in solution, which is prone to spontaneous ring openings. We propose that the stabilization within the viral capsid increases the packaging processivity and confers selectivity during RNA loading.

Keywords: Structure; Assembly; Interactions; Hydrogen–deuterium exchange; Dynamics; RNA; Packaging

Introduction

Genome encapsidation is an essential step in viral morphogenesis. Viruses utilize different strategies that can be divided into two basic classes: (1) nucleic acid–coat protein co-assembly, (2) genome packaging into a capsid precursor (procapsid, PC). While many ssRNA viruses adopt the first strategy, the latter is common among dsDNA viruses and dsRNA bacteriophages. The second type of genome packaging requires a molecular motor that converts chemical energy into mechanical work in the form of translocation and condensation of the nucleic acid. Genome packaging can be performed by a polymerase that replicates the genome while translocating it into the capsid precursor. In many cases, a specialized packaging ATPase performs the translocation. In dsDNA viruses, the packaging motor consists of a connector, which is embedded into one vertex of the capsid, and a multisubunit terminase with ATPase activity. Although the structure of bacteriophage ϕ29 connector was elucidated (Simpson et al., 2000), only limited structural information is available for terminases and their interactions with the connector (Feiss and Catalano, 2005).

In comparison with dsDNA viruses, the packaging motor (protein P4) from dsRNA bacteriophages (ϕ6–ϕ14, *Cystoviridae* family, (Mindich and Bamford, 1988; Mindich et al., 1999)) is much simpler (Kainov et al., 2003), and consists of a

Abbreviations: ESI Electrospray ionization; FT–ICR Fourier transform ion cyclotron resonance; HDX hydrogen_deuterium exchange; MS mass spectrometry; PC procapsid.

* Corresponding author. Fax: +358919159920.
E-mail address: roman.tuma@helsinki.fi (R. Tuma).
[1] Present address: IGBMC, CNRS, 1 rue Laurent Fries, BP 10142, 67404 Illkirch CEDEX, France.
[2] Present address: WM Keck Foundation Biotechnology Resource Laboratory, 300 George St., Yale University, New Haven, CT 06511, USA.

0042-6822/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2006.03.025

hexameric ring of identical ATPase subunits that encircle a central channel. Twelve hexamers are attached to the five-fold vertices of the procapsid (empty polymerase complex) (de Haas et al., 1999) (Fig. 1). The polymerase complex is a large icosahedral molecular machine that constitutes the core of the bacteriophage (Butcher et al., 1997). The icosahedral framework of PC is built from 120 copies of major structural protein P1. The P1 framework also exposes specific RNA binding sites in a segment-specific manner on its surface during packaging. This assures precise packaging of all three single-stranded genomic precursors (Mindich, 1999) because RNA binding to isolated P4 hexamer is not specific (Juuti et al., 1998; Kainov et al., 2003). During packaging initiation RNA carrying a specific *pac* site is bound to P1, presumably in the vicinity of P4 hexamer. The close proximity of RNA may trigger hexamer opening and RNA loading into the central channel as observed for isolated hexamers from bacteriophage ϕ8 (Lisal et al., 2005). After loading ATP hydrolysis by P4 powers unidirectional translocation of the enclosed RNA into PC.

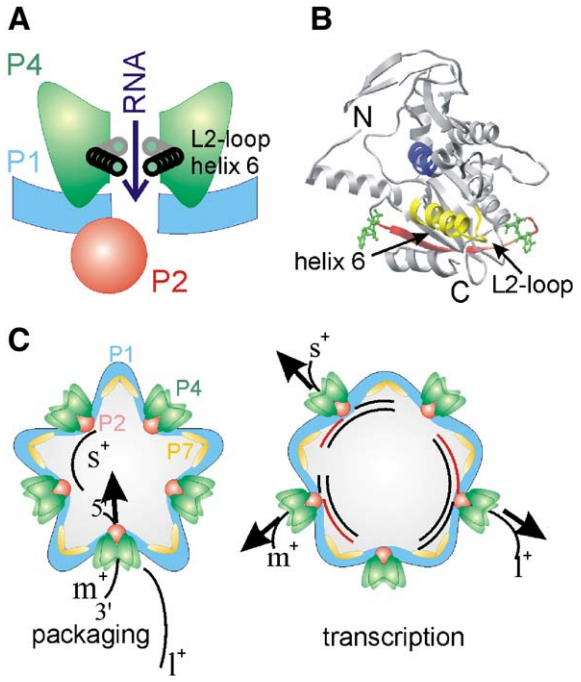


Fig. 1. Hexameric molecular motor P4. (A) Schematic model of RNA translocation through the central channel. The power stroke direction is delineated by blue arrow. The two limiting positions of the swiveling helix α6 are indicated. (B) Functional epitopes of P4 subunit (Lisal et al., 2005): The protected core (blue), cooperative unit encompassing the helix α6 and loop L2 (yellow) together with β-sheet 15 (red). ADP molecules are shown (green). (C) Dual role of P4 in the viral life cycle: active motor during packaging (left) and passive conduit for RNA, which is transcribed by the polymerase P2 (right). Schematic diagram shows bacteriophage ϕ12 polymerase complex (blue) and the packaging motor (green). The empty PC (procapsid) is assembled from 120 copies of proteins P1 (major structural protein). 12 P2 monomers (RNA-dependent RNA polymerase), 12 P4 hexamers (packaging ATPase) and 60 P7 subunits (minor structural protein) (Poranen et al., 2001). The procapsid specifically recognizes the three viral single-strand plus-sense genomic precursors (s^+, m^+, l^+) by their unique packaging signals situated at the 5′-ends and P4 motor packages a single copy of each (Mindich, 1999, 2004). During packaging P4 translocates ssRNA in the 5′ to 3′ direction (Kainov et al., 2003).

In addition to its roles in packaging, P4 hexamer is also required for nucleation of PC assembly (Poranen et al., 2001; Poranen and Tuma, 2004) and for transcription (Pirttimaa et al., 2002) (Fig. 1C). Interestingly, ATPase activity of P4 is shut-off during transcription and consequently the hexamer may act as a passive pore for RNA exit (Kainov et al., 2004) (Fig. 1). This suggests that P4 activity is regulated within the polymerase complex via its interactions with the PC framework (Benevides et al., 2002).

Recently, atomic structures of P4 hexamer from bacteriophage ϕ12 in several nucleotide binding states were solved and a translocation mechanism has been proposed (Lisal and Tuma, 2005; Mancini et al., 2004a). In this mechanism RNA is propelled by a swiveling motion of helix α6 and loop L2, which protrudes into the central channel (Figs. 1A–B). The motion was also visualized by solution hydrogen–deuterium exchange for P4 protein from bacteriophage ϕ8 (Lisal et al., 2005). The method further unraveled concerted changes in the dynamics of a cooperative unit that encompasses part of the catalytic Rec-A domain (Fig. 1B).

The three-dimensional structure of P4 and the EM maps suggest that P4 hexamer may interact with the P1 framework via either N-terminal dome or C-terminal facet (Fig. 1B). Due to the symmetry mismatch between the five-fold vertex and P4 hexamer EM has not provided any further information about these interactions (de Haas et al., 1999). It was shown genetically that the last 13 amino acids of ϕ6 P4 are essential for assembly and presumably this region interacts with P1 (Paatero et al., 1998). Here we take advantage of the high-resolution structure of P4 from bacteriophage ϕ12 and characterize the procapsid bound structure of ϕ12 P4 by hydrogen–deuterium exchange (HDX) (Englander et al., 1996) and high-resolution mass spectrometry (MS) (Lanman et al., 2004; Lanman and Prevelige, 2004). We compare HDX kinetics of the free hexamer with that of the PC-associated P4. HDX revealed protection of the C-terminus upon assembly and stabilization of the hexameric interfaces. Implications of the interactions for assembly and regulation are discussed. This contribution also demonstrates that HDX can be successfully applied to fairly complex virus capsids and complement the traditional techniques of structural virology.

Results and discussion

Deuterium exchange reveals stabilization of many regions

In a HDX experiment proteins or their complexes are first incubated in a D_2O buffer under native conditions to allow exchange of labile hydrogens, among others amide N–H groups of the polypeptide backbone. The pattern of deuterium incorporation reflects both solvent accessibility and local stability (Englander et al., 1996). HDX is first quenched by acidification and the denatured proteins are then fragmented by a non-specific protease to obtain fragments that are then analyzed by mass spectrometry (Zhang and Smith, 1993). Under the digestion conditions, only deuterium incorporated at the backbone amide sites is preserved while side chain

groups rapidly back exchange. This allows selective determination of the backbone amide exchange. In the case of protein assemblies (e.g. virus polymerase complex) the peptide mixture is complex and often generates overlapping isotopic envelopes. In such a case a high-resolution mass spectrometric technique like FT-ICR is required to resolve the isotopic envelopes of the selected protein. We selected the hexameric packaging NTPase P4 for which the three-dimensional (3-D) structure is known (Mancini et al., 2004a). As discussed below, knowledge of the 3-D structure plays pivotal role in interpretation of HDX data.

Deuterium incorporation into a peptide fragment (i.e., amide hydrogen exchange) caused progressive mass increase. The mass increments were converted to exchange kinetics (Fig. 2, 2nd column). Because each peptide fragment contains multiple amide sites that often exchange with different rates the kinetics are rather complex. Additional information was obtained by transforming the complex kinetics into exchange rate distributions (Fig. 2, middle column) (Zhang et al., 1997). From these distributions the number of protected sites ($k_{ex} < 0.1\ h^{-1}$), sites with intermediate exchange rate ($0.1 < k_{ex} < 100\ h^{-1}$) and fast exchanging sites ($k_{ex} > 100\ h^{-1}$) were computed (Fig. 2 right column) (Lisal et al., 2005). As seen from all three representations, some regions (e.g. Fig. 2 panels B–F) were more protected in the PC-associated state than in the isolated P4, showing that HDX reveals changes associated with interactions between P4 and PC. Altogether HDX kinetics were obtained for 24 peptic fragments which cover 68% of P4 structure.

In order to visualize the changes associated with PC binding we employed apparent HDX rates to color-code the ribbon representation of the P4 structure (Fig. 3). The apparent rates were obtained as weighted averages from the rate distributions (Lisal et al., 2005). Many regions in the isolated P4 hexamer exhibited fast and intermediate exchange rates (Fig. 3A), consistent with relatively unhindered solvent accessibility of the surface and relative instability of subunit interfaces. In contrast, exchange of many regions in PC-bound P4 hexamer was slower (Fig. 3B). In the following sections we distinguish between a global stabilization of the PC-bound hexamer and stabilization due to direct P4–PC interactions, which most likely involve residues on the surface of the hexamer.

Based on the protection pattern and the location within the three-dimensional structure the peptic fragments can be grouped into two classes. Fragments that are located on the surface of the hexamer exhibited fast or intermediate exchange rates in the isolated P4 (Figs. 2A, E, F) and belong to the first class. Some of them became protected in the PC-bound P4 (Figs. 2E–F). Such protection is most likely due to decreased solvent accessibility in the P4–PC complex and thus constitutes a “footprint” of the interactions on the surface of P4.

The second class encompasses fragments that are located within the hexameric structure as judged by the 3-D model. Some of them were partially protected in the isolated hexamer (Figs. 2B–C) and many became further stabilized within PC (Figs. 2B–D). This protection is due to global stabilization because it is unlikely that these partially buried regions would directly interact with PC.

P4 hexamer interacts with the procapsid via its C-terminal facet

On the bases of EM reconstruction for PC from the related phage ϕ6 (de Haas et al., 1999) and the doughnut structure of P4 hexamer (Mancini et al., 2004a) the interactions between P4 and PC should be limited to either the N-terminal dome or the C-terminal facet of P4. Based on the compact structure of P4 we assume that only regions exposed on the surface of P4 hexamer will interact with PC. The surface representation of exchange rates is shown in Fig. 4. Conspicuously, the fragments lining the bottom facet of the hexamer (CT and fragments within the adjacent β-sheet) exhibited significantly decreased HDX rates. In the isolated hexamer these C-terminal fragments exhibit negligible number of protected sites (Figs. 2E–F, right column) while upon binding to PC both fragments gained significant number of protected sites. This is compatible with a decreased solvent accessibility of the C-terminus in the P4–PC complex. On the other hand, surface-exposed regions within the N-terminal apical domain were not significantly protected upon association with PC (Fig. 2A). Only two fragments within the N-terminal facet became more protected upon PC association. These fragments are situated in close vicinity to and are part of the subunit interfaces that are protected via global stabilization of the hexamer (see below). Thus, it is unlikely that these regions interact directly with PC.

Taken together, the C-terminal facet, in particular the C-terminus, C-terminal helix and the adjacent β-strand, constitute the likely sites of direct P4–PC interaction. Notably, this orientation is also consistent with the direction of RNA translocation proposed on the basis of the structure-based mechanism in which the RNA translocation direction is from the apical towards the C-terminal domain (Mancini et al., 2004a) (Fig. 1A). Furthermore, the C-terminal 13 residues of P4 from the related phage ϕ6 were essential for assembly (Paatero et al., 1998). We conclude that the C-terminal facet of P4 hexamer is proximal to and interacts with the procapsid.

Procapsid controls hexamer opening

As noted above PC association results in global stabilization of the hexamer. Most of these regions are located at subunit interfaces. For example, the two helices at subunit interfaces (IH1 and IH2 in Fig. 3B, see also Fig. 2C) become significantly protected only upon incorporation of the hexamer into PC. These helices in the related ϕ8 virus P4 were shown to be exposed only upon opening of the hexameric ring during RNA loading into the central channel (Lisal et al., 2005). Thus, ϕ12 P4 hexamer encounters frequent spontaneous ring openings in solution, a feature manifested by its low RNA translocation processivity in comparison with that of the ϕ8 protein (Lisal and Tuma, 2005). However, this spontaneous opening ceases within the structural context of the PC. In the context of the virus PC may trigger ring openings in a controlled fashion in order to selectively load the viral RNA during packaging initiation. In this way loading of heterologous RNA is prevented.

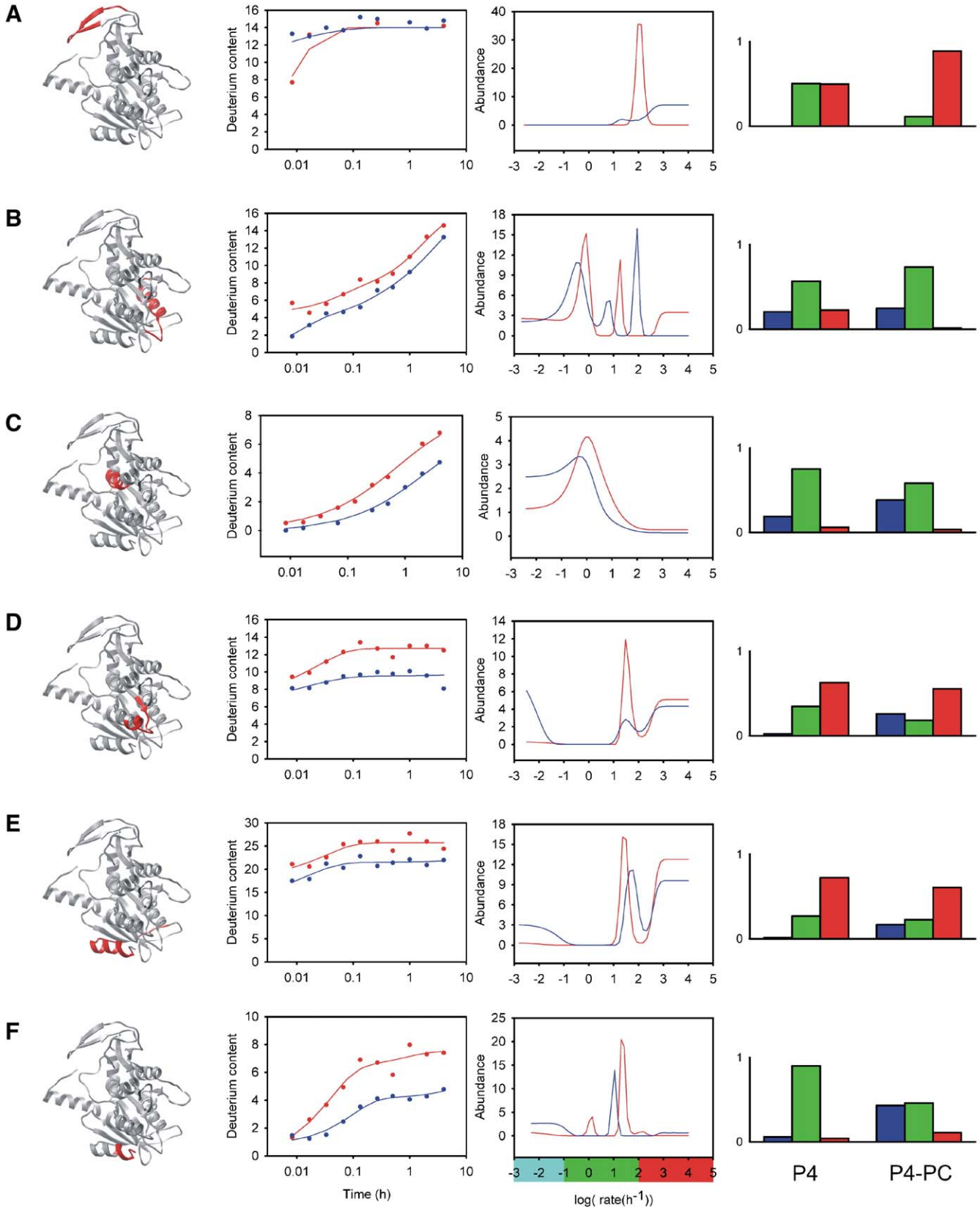


Fig. 2. Hydrogen–deuterium exchange measured by mass spectrometry. HDX kinetics of selected regions (P4 hexamer in solution in red, P4–PC complex in blue): residues 69–83 (A), 131–152 (B), 215–224 (C), 230–244 (D), 294–320 (E), 316–324 (F). Panels from left to right: Localization of the region within the structure of the P4 subunit, HDX kinetics, rate distribution, number of sites exchanging with slow (<0.1 h^{-1}, blue), intermediate (0.1–100 h^{-1}, green) and fast (>100 h^{-1}, red) rates under the two conditions (as shown below the panel).

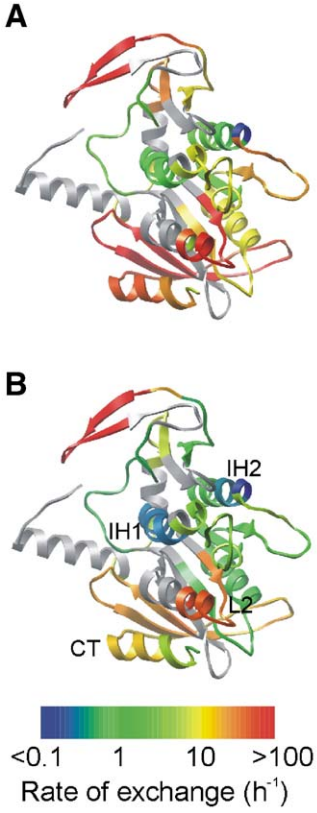


Fig. 3. Ribbon diagrams for one ϕ12 P4 subunit colored according to the HDX rates for isolated P4 hexamer (A) computed from rate distributions similar to those in Fig. 2 (for details see (Lisal et al., 2005)). The hexamer within the procapsid (B). The central channel of P4 is in the front, the C-terminal base is at the bottom; the N-terminal apical domain points up. The scale bar for HDX rates (h^{-1}) is shown at the bottom. IH1 and IH2 are the two helices located at the subunit interfaces. L2 designates the L2 loop and CT is the C-terminal helix.

The ring opening mechanism is not limited to P4. For example, the DnaC-mediated loading of the hexameric replicative helicase DnaB onto DNA also proceeds via a ring opening (Delagoutte and von Hippel, 2003). Similarly, association with other component of the replication machinery controls the specificity of DNA loading. DnaB hexamer alone can load onto ssDNA using ring opening. However, DnaC prevents this non-specific process and promotes loading at the origin of replication (*OriC*) sites. As in P4 DnaB interacts with DnaC via one facet of the hexameric ring (Barcena et al., 2001). Therefore, it is likely that global stabilization may control ring opening also in this case. Association between DnaB and DnaC also inhibits ATPase activity of DnaB (Galletto et al., 2003). This resembles the effect seen for P4 within the polymerase complex that accompanies switching from the packaging to RNA transcription mode (Kainov et al., 2004). In the next section we speculate about the structural basis for such regulation.

Mechanism of activity regulation

All regions close to the C-terminal helix were affected by P4 binding to the procapsid (cf. Figs. 3A and B and Supplementary material). Dynamic stabilization propagates from the C-terminal helix via β15 sheet to α6 helix and L2 loop (Fig. 2D). This region also constitutes the cooperative unit of the ATPase as proposed for P4 from ϕ8 phage (Fig. 1B). Association of the C-terminus with PC appears to affect dynamics of the cooperative unit and thus may also affect the ATPase activity. The influence of the C-terminus on the ATPase activity was examined by use of a C-terminal deletion mutant P4ΔC (ϕ12 P4 missing amino acids 312–331) (Mancini et al., 2004b). Although the basal ATPase activity (k_{cat}) of P4ΔC in the absence of RNA (0.66 s^{-1}) was slightly below that of the full-length P4 (0.84 s^{-1}), the truncated protein was stimulated by RNA to a greater degree (k_{cat} = 3.95 s^{-1} versus 2.52 s^{-1} for wt). Thus, the C-terminus acts as an inhibitor of ATPase activity in the isolated hexamer. We propose that in the virus core this helix may be either withdrawn from the cooperative unit to activate the ATPase (procapsid in packaging state) or "jammed" onto the catalytic RecA-like domain to shut off the activity (polymerase complex in transcription mode), respectively.

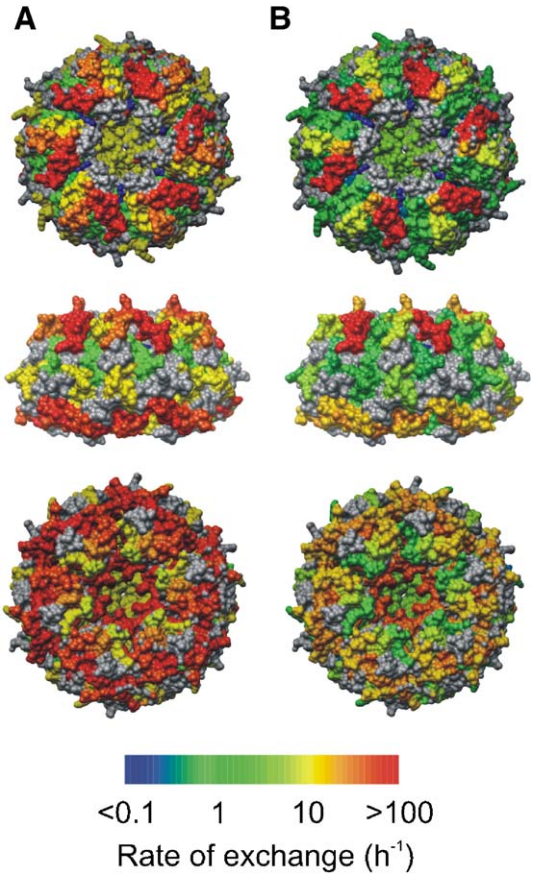


Fig. 4. Surface representations of ϕ12 P4 hexamer colored according to the weighted average HDX rates. Panel A shows the HDX rate for isolated hexamer whereas panel B shows HDX rates for P4 in the procapsid. The color bar in bottom is the color scale for rates (h^{-1}); segments for which no peptic fragments were available are shown in gray. The top panel shows a top view of the hexamer facing the N-terminal apical domain, the middle panel shows a side view and the bottom panel shows the C-terminal base of the hexamer.

Conclusions

In this study, we have demonstrated that hydrogen–deuterium exchange can be selectively determined for a selected subunit type within a complex, multisubunit, virus assembly. The exchange kinetics allowed comparing the subunit solution and crystal structures with that adopted within a large multi-subunit assembly. This approach revealed the interactions of the packaging motor with the viral core and provided insight into roles of different regions in assembly, genome packaging and transcription regulation. The approach holds promise for studying other packaging motors, especially those that are only transiently associated with the virus capsid (e.g., terminases or various non-structural proteins of dsRNA viruses) for which structural analysis by EM or X-ray crystallography is difficult or impossible. The technique shall be applicable to other molecular machines like ribosomes or replisomes and their transient states.

Methods

Protein and virus preparations

The ϕ12 PC was assembled in *Escherichia coli* strain BL21 (DE3) transformed with plasmid pPL1 carrying a cDNA copy of the entire l segment after the T7 promoter. The expression was induced with 0.5 mM IPTG and particles were purified from cell lysates by zonal centrifugation in 5% to 20% sucrose gradients. The gradients were fractionated and aliquots from these fractions were analyzed by SDS-PAGE. The particles were composed of P1, P2 and P4 proteins, although a fraction of P4 was lower than in the virus. For further details, see (Gottlieb et al., 1988, 2002).

The ϕ12 P4 wild type (wt) protein was expressed in *Escherichia coli* and purified as previously described (Mancini et al., 2004b). To produce the C-terminal deletion mutant P4ΔC (residues 1–311), the 5′-terminal part of P4 gene was PCR amplified from pPG27 (using primers 5′-GGAGTGACATATGATCCATCTGTACGA-3′, 5′-CTACGGATCC**TCA**AGCTTTGCCGGAGGCTT-3′) and inserted into pT7-7 plasmid at *Nde*I–*Hin*dIII sites to yield an expression plasmid pDK50. The expression, purification and activity measurements for the truncated protein were essentially the same as for the wild type protein.

Hydrogen–deuterium exchange

HDX analysis was performed as described in (Lam et al., 2002). P4 or procapsids were diluted into D_2O containing buffer and exchanged for variable period of time ranging from 30 s to 4 h. Exchange was quenched by acidification and rapid freezing in liquid nitrogen. Fragmentation was achieved by pepsin digestion for 2 min on ice. Peptides were identified and assigned by either exact isotopic mass matching or by MS/MS analysis using ESI-ion trap Esquire 3000+ mass spectrometer from Bruker Daltonics. Deuterium contents were measured with a home built 9.4 T Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (Lam et al., 2002; Marshall et al., 1998). Exchange was corrected for back exchange during sample processing using a fully deuterated control as described in (Lisal et al., 2005). 24 peptide fragments covering 68% of P4 sequence provided sufficient signal in both PC and P4 hexamer spectra for reliable analysis.

Acknowledgments

We thank Ayman Abu Ramadan for the software development. Further we thank Rabah Soliymani and Dr. Marc Baumann for help with the MS/MS experiments. J.L. is supported by Viikki Graduate School in Biosciences, D.E.K. was supported by the National Graduate School in Informational and Structural Biology. This work was supported by Academy of Finland grant 206926 (RT), the Finnish Centre of Excellence Program 2000–2005, the NSF National High Field FT-ICR Mass Spectrometry Facility (CHE-99-09502), Florida State University, and the National High Magnetic Field Laboratory in Tallahassee, FL. P.G. was supported by The National Science Foundation Career Award MCB9984310, a grant from The City University of New York from the PSC-CUNY Research Award Program and the NIH-Research Center for Minority Institutions (RCMI) grant G12RR-A103060. We wish to thank Kenneth M. Kruta for critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2006.03.025.

References

Barcena, M., Ruiz, T., Donate, L.E., Brown, S.E., Dixon, N.E., Radermacher, M., Carazo, J.M., 2001. The DnaB.DnaC complex: a structure based on dimers assembled around an occluded channel. EMBO J. 20, 1462–1468.

Benevides, J.M., Juuti, J.T., Tuma, R., Bamford, D.H., Thomas Jr., G.J., 2002. Characterization of subunit-specific interactions in a double-stranded RNA virus: Raman difference spectroscopy of the phi6 procapsid. Biochemistry 41, 11946–11953.

Butcher, S.J., Dokland, T., Ojala, P.M., Bamford, D.H., Fuller, S.D., 1997. Intermediates in the assembly pathway of the double-stranded RNA virus phi6. EMBO J. 16, 4477–4487.

de Haas, F., Paatero, A.O., Mindich, L., Bamford, D.H., Fuller, S.D., 1999. A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage phi6. J. Mol. Biol. 294, 357–372.

Delagoutte, E., von Hippel, P.H., 2003. Helicase mechanisms and the coupling of helicases within macromolecular machines. Part II: integration of helicases into cellular processes. Quart. Rev. Biophys. 36, 1–69.

Englander, S.W., Sosnick, T.R., Englander, J.J., Mayne, L., 1996. Mechanisms and uses of hydrogen exchange. Curr. Opin. Struct. Biol. 6, 18–23.

Feiss, M., Catalano, C.E., 2005. Bacteriophage lambda terminase and the mechanisms of viral DNA packaging. In: Catalano, C.E. (Ed.), Viral Genome Packaging Machines: Genetics, Structure, and Mechanism. Plenum Publishers, New York, pp. 5–39.

Galletto, R., Jezewska, M.J., Bujalowski, W., 2003. Interactions of the *Escherichia coli* DnaB helicase hexamer with the replication factor the DnaC protein. Effect of nucleotide cofactors and the ssDNA on protein–protein interactions and the topology of the complex. J. Mol. Biol. 329, 441–465.

Gottlieb, P., Strassman, J., Bamford, D.H., Mindich, L., 1988. Production of a polyhedral particle in *Escherichia coli* from a cDNA copy of the large genomic segment of bacteriophage phi 6. J. Virol. 62, 181–187.

Gottlieb, P., Potgieter, C., Wei, H., Toporovsky, I., 2002. Characterization of phi12, a bacteriophage related to phi6: nucleotide sequence of the large double-stranded RNA. Virology 295, 266–271.

Juuti, J.T., Bamford, D.H., Tuma, R., Thomas Jr., G.J., 1998. Structure and NTPase activity of the RNA-translocating protein (P4) of bacteriophage phi 6. J. Mol. Biol. 279, 347–359.

Kainov, D.E., Pirttimaa, M., Tuma, R., Butcher, S.J., Thomas Jr., G.J., Bamford, D.H., Makeyev, E.V., 2003. RNA packaging device of double-stranded RNA bacteriophages, possibly as simple as hexamer of P4 protein. J. Biol. Chem. 278, 48084–48091.

Kainov, D.E., Lisal, J., Bamford, D.H., Tuma, R., 2004. Packaging motor from double-stranded RNA bacteriophage phi12 acts as an obligatory passive conduit during transcription. Nucleic Acids Res. 32, 3515–3521.

Lam, T.T., Lanman, J.K., Emmett, M.R., Hendrickson, C.L., Marshall, A.G., Prevelige, P.E., 2002. Mapping of protein:protein contact surfaces by hydrogen/deuterium exchange, followed by on-line high-performance liquid chromatography-electrospray ionization Fourier-transform ion-cyclotron-resonance mass analysis. J. Chromatogr., A 982, 85–95.

Lanman, J., Prevelige Jr., P.E., 2004. High-sensitivity mass spectrometry for imaging subunit interactions: hydrogen/deuterium exchange. Curr. Opin. Struct. Biol. 14, 181–188.

Lanman, J., Lam, T.T., Emmett, M.R., Marshall, A.G., Sakalian, M., Prevelige Jr., P.E., 2004. Key interactions in HIV-1 maturation identified by hydrogen–deuterium exchange. Nat. Struct. Mol. Biol. 11, 676–677.

Lisal, J., Tuma, R., 2005. Cooperative mechanism of hexameric packaging motor. J. Biol. Chem. 280, 23157–23164.

Lisal, J., Lam, T., Kainov, D.E., Emmett, M.R., Marshall, A.G., Tuma, R., 2005. Functional visualization of viral molecular motor by hydrogen–deuterium exchange reveals transient states. Nat. Struct. Mol. Biol. 12, 460–466.

Mancini, E.J., Kainov, D.E., Grimes, J.M., Tuma, R., Bamford, D.H., Stuart, D.I., 2004a. Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. Cell 118, 744–755.

Mancini, E.J., Kainov, D.E., Wei, H., Gottlieb, P., Tuma, R., Bamford, D.H., Stuart, D.I., Grimes, J.M., 2004b. Production, crystallization and preliminary X-ray crystallographic studies of the bacteriophage phi 12 packaging motor. Acta Crystallogr., D 60, 588–590.

Marshall, A.G., Hendrickson, C.L., Jackson, G.S., 1998. Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom. Rev. 17, 1–35.

Mindich, L., 1999. Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage phi6. Microbiol. Mol. Biol. Rev. 63, 149–160.

Mindich, L., 2004. Packaging, replication and recombination of the segmented genome of bacteriophage Phi6 and its relatives. Virus Res. 101, 83–92.

Mindich, L., Bamford, D.H., 1988. Lipid-containing bacteriophages. In: Calendar, R. (Ed.), The Bacteriophages, vol. 2. Plenum Publishing Corporation, New York, pp. 475–519.

Mindich, L., Qiao, X., Qiao, J., Onodera, S., Romantschuk, M., Hoogstraten, D., 1999. Isolation of additional bacteriophages with genomes of segmented double-stranded RNA. J. Bacteriol. 181, 4505–4508.

Paatero, A.O., Mindich, L., Bamford, D.H., 1998. Mutational analysis of the role of nucleoside triphosphatase P4 in the assembly of the RNA polymerase complex of bacteriophage phi6. J. Virol. 72, 10058–10065.

Pirttimaa, M.J., Paatero, A.O., Frilander, M.J., Bamford, D.H., 2002. Nonspecific nucleoside triphosphatase P4 of double-stranded RNA bacteriophage phi6 is required for single-stranded RNA packaging and transcription. J. Virol. 76, 10122–10127.

Poranen, M.M., Tuma, R., 2004. Self-assembly of double-stranded RNA bacteriophages. Virus Res. 101, 93–100.

Poranen, M.M., Paatero, A.O., Tuma, R., Bamford, D.H., 2001. Self-assembly of a viral molecular machine from purified protein and RNA constituents. Mol. Cell 7, 845–854.

Simpson, A.A., Tao, Y., Leiman, P.G., Badasso, M.O., He, Y., Jardine, P.J., Olson, N.H., Morais, M.C., Grimes, S., Anderson, D.L., Baker, T.S., Rossmann, M.G., 2000. Structure of the bacteriophage phi29 DNA packaging motor. Nature 408, 745–750.

Zhang, Z., Smith, D.L., 1993. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. Protein Sci. 2, 522–531.

Zhang, Z., Li, W., Logan, T.M., Li, M., Marshall, A.G., 1997. Human recombinant [C22A] FK506-binding protein amide hydrogen exchange rates from mass spectrometry match and extend those from NMR. Protein Sci. 6, 2203–2217.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Virus Research 101 (2004) 93–100
Virus Research
www.elsevier.com/locate/virusres

Self-assembly of double-stranded RNA bacteriophages

Minna M. Poranen*, Roman Tuma

Department of Biological and Environmental Sciences, University of Helsinki, Viikinkaari 5, PL 56, FIN-00014, Finland

Abstract

Double-stranded RNA viruses infecting bacterial hosts belong to the Cystoviridae family. Bacteriophage ϕ6 is one of the best characterized dsRNA viruses and shares structural as well as functional similarities with other well-studied eukaryotic dsRNA viruses (e.g. L-A, rotavirus, bluetongue virus, and reovirus). The assembly pathway of the enveloped, triple-layered ϕ6 virion has been well documented and can be divided into four distinct steps which are (1) procapsid formation, (2) genome encapsidation and replication, (3) nucleocapsid surface shell assembly, and (4) envelope formation. In this review, we focus primarily on the procapsid and nucleocapsid assembly for which in vitro systems have been established. The in vitro assembly systems have been instrumental in revealing assembly intermediates and conformational changes that are common to ϕ6 and ϕ8, two cystoviruses with negligible sequence homology. Two viral enzymes, the packaging NTPase (P4) and the RNA-dependent RNA polymerase (P2), were found essential for the nucleation step. The nucleation complex contains one or more tetramers of the major procapsid protein (P1) and is further stabilized by protein P4. Interaction of P1 and P4 during assembly is accompanied by an additional folding of their respective polypeptide chains. The in vitro assembled procapsids were shown to selectively package and replicate the genomic ssRNA. Furthermore, in vitro assembly of infectious nucleocapsids has been achieved in the case of ϕ6. The in vitro studies indicate that the nucleocapsid coat protein (P8) assembles around the polymerase complex in a template-assisted manner. Implications for the assembly of other dsRNA viruses are also presented.

Keywords: Virus assembly; In vitro; dsRNA virus; Nucleation complex; Molecular machine

1. Introduction

Bacteriophage ϕ6 is the type organism of the family Cystoviridae (http://www.ncbi.nlm.nih.gov/ICTV/). For a long time ϕ6 represented the only member of this family and was the only known double-stranded RNA virus (dsRNA virus) infecting bacteria. Recently eight similar viruses have been isolated (Mindich et al., 1999). Some of the new isolates (ϕ7, ϕ9, ϕ10, and ϕ11) have clear sequence similarity with ϕ6, while others (ϕ8, ϕ12, and ϕ13) are more distantly related. However, the virion architecture and the genomic organization are conserved. In this review we focus on the bacteriophage ϕ6 because most of the current knowledge on the structure, assembly, and molecular biology of dsRNA bacteriophages is derived from this model system. The other dsRNA bacteriophages, especially ϕ8, are invoked when related information is available.

The ϕ6 system has been of interest because its internal polymerase complex is a cytoplasmic molecular machine, translocating, replicating and transcribing RNA in a highly specific manner. This multi-functional complex can now be reconstituted from its component proteins, and the assembled complexes are fully functional both in vitro and in vivo (Poranen et al., 2001). Such in vitro assembly system has opened up the possibility to extract novel information on the sequence of molecular interactions operating during ϕ6 virion assembly. A similar in vitro system has been recently described for bacteriophage ϕ8 (Kainov et al., 2003b) and allows for assessment of assembly pathway evolution and conservation among members of the Cystoviridae. The compilation of the in vitro assembly results and description of the common features of polymerase complex assembly constitute the major part of this review.

The similarities among dsRNA viruses are not restricted to viruses infecting bacterial hosts, instead, common features have been observed between cystoviruses and eukaryotic dsRNA viruses of the Reoviridae family (Bamford et al., 2002). The similarity is evident in the innermost virion layer, the polymerase complex, which carries out

* Corresponding author. Tel.: +358-9-19159106; fax: +358-9-19159098.
E-mail addresses: minna.poranen@helsinki.fi (M.M. Poranen), roman.tuma@helsinki.fi (R. Tuma).

0168-1702/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2003.12.009

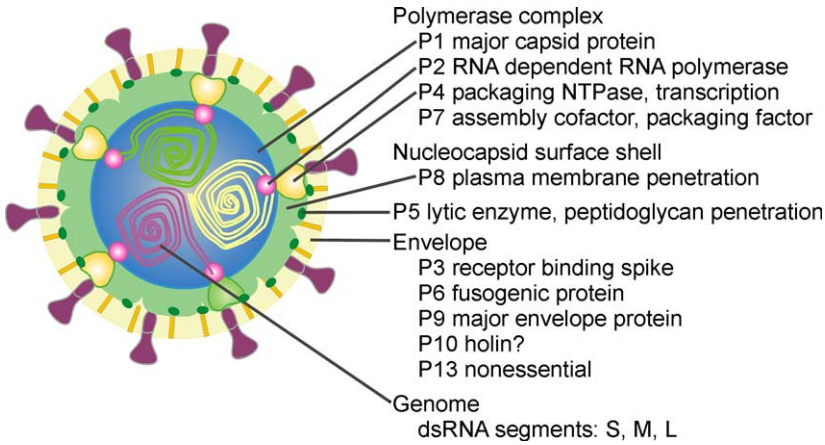


Fig. 1. Architecture of ϕ6 virion and protein functions.

genome packaging, replication, and transcription. Given such high degree of conservation, conclusions reached for the bacteriophage have clear implications for the assembly pathways of related eukaryotic viruses.

Because the life-cycle, genomes and replication strategies of cystoviruses have been described in detail in the Mindich paper in this special issue we limit our introduction to structural features that are relevant to the virion assembly. The structure of ϕ6 virion is schematically depicted in Fig. 1. The innermost layer, the viral core (or in more general terms, the polymerase complex) is composed of four protein species, P1, P2, P4, and P7 (Bamford and Mindich, 1980; Mindich and Davidoff-Abelson, 1980). P1 is the major component of the polymerase complex forming the polyhedral skeleton of the particle (Olkkonen and Bamford, 1987; Ktistakis and Lang, 1987). There are 120 copies of protein P1 arranged as 60 asymmetric dimers on a $T = 1$ icosahedral lattice (Butcher et al., 1997). Such a capsid organization exhibiting two chemically identical subunits in different conformations within the icosahedral asymmetric unit has been observed only in the cores of dsRNA viruses and is sometimes referred to as a "$T = 2$" structure (Grimes et al., 1998). A hexameric protein, P4, forms turret-like protrusions on the P1 lattice creating a symmetry mismatch at the five-fold vertices (Butcher et al., 1997; de Haas et al., 1999). The polymerase P2 is located under the five-fold vertex (Ikonen et al., 2003). P7 has been proposed to reside at two-fold symmetry axes of the polymerase complex (Juuti and Bamford, 1997; Poranen et al., 2001). In addition to the above described symmetry mismatch recent studies indicated that one of the 12 P4 hexamers is physically and functionally different (Fig. 2) (Pirttimaa et al., 2002).

The core is surrounded by a second proteinaceous layer, called the nucleocapsid surface shell (Fig. 1), composed mainly of a single viral protein, P8 (van Etten et al., 1976; Bamford and Mindich, 1980). P8 is arranged on a $T = 13$ lattice with the exception of the regions close to the five-fold symmetry axes that are occupied by the hexameric P4 turrets extending from the polymerase complex (Butcher et al., 1997). This double-shell assembly intermediate is designated as a nucleocapsid (NC).

The outermost layer of the virion is the lipid bilayer (Fig. 1) (van Etten et al., 1976). The envelope of ϕ6 contains phospholipids originating from the host plasma membrane (Sands, 1973), and four virally encoded integral membrane proteins, P6, P9, P10, P13 (Gottlieb et al., 1988a; Sinclair et al., 1975; van Etten et al., 1976), of which protein P6 anchors the receptor-binding spike protein P3 (Stitt and Mindich, 1983b). The envelope lacks icosahedral symmetry (personal communication, F. de Haas, A.O. Paatero, S.J. Butcher, D.H. Bamford and S.D. Fuller).

2. Procapsid assembly

The early studies of ϕ6 infected cells revealed that the assembly pathway involves formation of empty polymerase complexes with dodecahedral morphology (Fig. 2) (Bamford and Mindich, 1980). These precursors are called procapsids (PC). PC has the same protein composition as the virion core but is devoid of RNA. In early studies no assembly intermediates for the PC were detected. However, since then substantial information on the structural and functional organization of the PC has been obtained from analyses of assemblies formed during expression of different combinations of viral proteins (Gottlieb et al., 1988b, 1990; Paatero et al., 1998). Additional details of the ϕ6 and ϕ8 polymerase complex morphogenesis are based on recent in vitro assembly assays (Poranen et al., 2001; Kainov et al., 2003b), which have facilitated elucidation of the sequence of protein-protein interactions and structural characterization of assembly intermediates.

2.1. Assembly of recombinant procapsids

When expressed within a host cell or in a heterologous organism, the polymerase complex proteins assemble into PCs similar to those found early during ϕ6 infection (Gottlieb

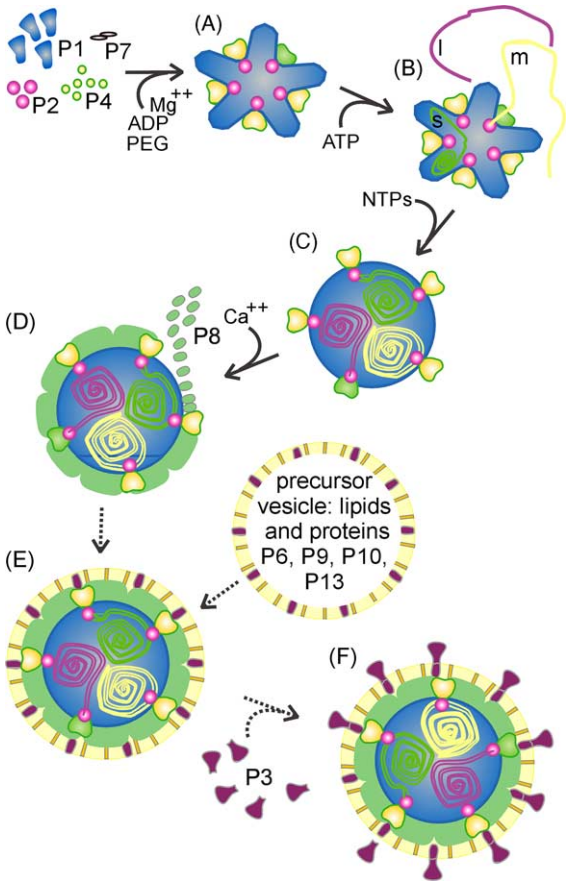


Fig. 2. Overview of the ϕ6 assembly pathway. Solid arrows indicate reactions for which in vitro systems have been established, while dotted arrows mark processes that can so far be studied only within a cell. For in vitro reactions the minimal chemical conditions that activate subsequent processes are marked beside the arrow. The empty procapsid of ϕ6 is composed of four proteins (A). P1 forms the structural skeleton of the particle, which is stabilized by P7. The enzymatic components, the packaging NTPase P4 hexamer and the polymerase P2 monomer, are located at or near the five-fold axes. One of the P4 hexamers differs from the others in its physical properties so that P4 is more tightly bound to the procapsid (special vertex, colored in green) than the others (colored in yellow). The special vertex is active during the packaging of the single-stranded genomic segments s, m and l (B). The packaging of the three genomic segments is followed by minus strand synthesis which take place within the particle (C). The empty packaging competent polymerase complex has dodecahedral appearance (A–B) while the mature, dsRNA filled, polymerase complex is roughly spherical (C).The nucleocapsid shell (P8) is subsequently assembled around the dsRNA filled polymerase complex (D) and the nucleocapsid is enveloped (E). Virions attain infectivity by acquisition of receptor binding spikes (P3) (F).

et al., 1988b). This indicates that the PC assembly is not dependent on additional factors encoded by the host cell or by the M and S genome segments. PC formation is also independent of natural viral RNA.

In addition to the complete polymerase complexes, it is possible to produce different incomplete PCs using similar heterologous expression systems. Soluble particles, missing either one of the proteins P2, P4, P7, or both P2 and P7, have been obtained. Thus, proteins P2, P4, and P7 can associate with the P1 skeleton independently of each other (Gottlieb et al., 1988b; Paatero et al., 1998; de Haas et al., 1999; Juuti and Bamford, 1995; Benevides et al., 2002). When P1 from ϕ6 is expressed alone, it forms spherical particles, which are very aggregative and partially unfolded (Gottlieb et al., 1988b; Paatero et al., 1998; Benevides et al., 2002). Similar aggregation has been observed for ϕ12 and ϕ13 P1 (D. Kainov, R. Tuma, unpublished) but limited solubility of ϕ8 P1 has been achieved (Kainov et al., 2003b).

2.2. In vitro assembly

The recently published in vitro assembly of the polymerase complexes of dsRNA bacteriophages ϕ6 and ϕ8 represented an important step in understanding the assembly of complex viruses and provided the first evidence for early intermediates on the assembly pathway of a dsRNA virus. The ϕ6 and ϕ8 assembly systems are very efficient; 30–50% of the precursor proteins are assembled into PCs when incubated under in vitro assembly conditions (Poranen et al., 2001; Kainov et al., 2003b). Based on their protein composition, sedimentation, morphology and enzymatic activities (ssRNA packaging, replication and transcription), the in vitro assembled structures are indistinguishable from the recombinant particles purified from bacteria (Poranen et al., 2001; Kainov et al., 2003b). Furthermore, the in vitro assembled polymerase complexes of ϕ6 are able to enter the host cell and to initiate a new round of productive infection (Poranen et al., 2001).

2.2.1. Characterization of building blocks

To set up an in vitro assembly system the component proteins have to be isolated in an active and soluble form. Table 1 compares properties of ϕ6 and ϕ8 PC building blocks. Despite negligible amino acid sequence homology each pair of the PC proteins has a similar monomeric size (Hoogstraten et al., 2000).

Protein P1 forms the dodecahedral framework of the viral core (Olkkonen and Bamford, 1987; Butcher et al., 1997). In addition to its structural role, P1 is also involved in the specific binding and recognition of the viral genome segments (Qiao et al., 2003). The P1 framework is rather malleable and undergoes substantial but cooperative conformational changes during maturation (see Fig. 2 and Mindich, this issue) (Butcher et al., 1997). These allosteric properties most likely result in the precise control of the polymerase complex multiple functions during genome encapsidation and replication (Poranen et al., 2004). We have purified protein P1 from two of the dsRNA bacteriophages, ϕ6 and ϕ8. P1 of ϕ6 is an elongated monomer exhibiting limited solubility (Poranen et al., 2001; Benevides et al., 2002), while P1 from ϕ8 can be obtained as a soluble tetramer (Kainov et al., 2003b) (Table 1).

Protein P2 contains the consensus sequence of known RNA-dependent RNA polymerases (Koonin et al., 1989). Particles missing P2 (P1P4P7 particles) can package ssRNA

Table 1
Properties of procapsid building blocks

Protein	Virus	Mass (kDa)[a]	Copies/virion[b]	Multimeric status	Hydrodynamic radius (nm)	Functions
P1	ϕ6	85.0	120	Monomer[c]	3.8[c]	Major structural protein
	ϕ8	86.8	120	Tetramer[d]	5.5[d]	Major structural protein
P2	ϕ6	74.8	12	Monomer[e]	3.3[c]	RNA-dependent RNA polymerase[e]
	ϕ8	71.5	12	Monomer[d]	2.6[d]	RNA-dependent RNA polymerase[h]
P4	ϕ6	35.0	72	Hexamer[f]	5.9[f]	ssRNA packaging, NTPase, transcription[j]
	ϕ8	34.1	72	Hexamer[d]	5.1[d]	NTPase, helicase[i]
P7	ϕ6	17.2	60	Dimer[g]	3.7[c]	Assembly cofactor[c], packaging factor[g]
	ϕ8	19.0	60	Dimer[d]	2.8[d]	Not known

[a] Calculated from the cDNA sequence (excluding N-terminal Met that is removed).
[b] Estimated copy numbers, based on Day and Mindich (1980) and the symmetry of the virion.
[c] Poranen et al. (2001).
[d] Kainov et al. (2003b).
[e] Makeyev and Bamford (2000).
[f] Juuti et al. (1998).
[g] Juuti and Bamford (1997).
[h] Yang et al. (2001).
[i] Kainov et al. (2003a).
[j] Pirttimaa et al. (2002).

but are unable to carry out genome replication and transcription (Juuti and Bamford, 1995). P2 has been purified as a monomer and is an active RNA-dependent RNA polymerase which can initiate complementary strand synthesis on the ssRNA template by a primer-independent mechanism (Makeyev and Bamford, 2000). The structure of ϕ6 P2 shares common features with the polymerases of hepatitis C virus, human immunodeficiency virus (Butcher et al., 2001), and reovirus (Tao et al., 2002). The polymerase P2 is located under the five-fold vertices in the particle interior (Ikonen et al., 2003).

Protein P4 of ϕ6, as well as that of ϕ8, contains sequence motifs of NTP binding proteins (Walker-motifs) (Gottlieb et al., 1992; Kainov et al., 2003a). Purified P4 self-assembles into ring-like hexamers, which are stable in the case of ϕ8 (Kainov et al., 2003b) but require ADP and divalent cations (Mg^{2+} and Ca^{2+}) for stability in the case of ϕ6 (Juuti et al., 1998). ϕ6 P4 is active only in the hexameric form and can hydrolyze all nucleotide triphosphates (rNTP, dNTP, and ddNTP). The NTPase activity is weakly stimulated by ssRNA (Paatero et al., 1995; Juuti et al., 1998). In contrast, the activity of ϕ8 P4 is tightly coupled to ssRNA binding and the hexameric protein also possesses an RNA helicase activity (Kainov et al., 2003a). Such activity has not been detected for purified P4 of ϕ6, although ϕ6 PC can displace short stretches of dsRNA during packaging (Kainov et al., 2003a). ϕ6 P4 is the viral packaging NTPase and interestingly, it is also needed for transcription (Pirttimaa et al., 2002).

Protein P7 is a minor component of the particle. Purified P7 is an elongated dimer (Juuti and Bamford, 1997) (Table 1). The N-terminal part of ϕ6 P7 forms the core of the dimer, while the C-termini constitute flexible, protease sensitive arms (S.J. Butcher, personal communication). The flexible part of the protein is not crucial for the stable incorporation of the P7 into the polymerase complex, but it is essential for the viability of the virus (M.M. Poranen, S.J. Butcher, unpublished).

P7-null particles have poor packaging and transcription efficiencies (Juuti and Bamford, 1997). In addition, P7 is an assembly cofactor (see Section 2.2.3). These diverse roles of P7 suggest that it might be either a particle-stabilizing (glue) protein or constitute a hinge which allows structural transitions within the P1 shell. The estimated copy number for P7 is 60 (Juuti and Bamford, 1997), and it has been suggested to be located near the two-fold axes of the PC, so that the elongated P7 dimer could span in the interface between two P1 dimers (Poranen et al., 2001).

2.2.2. Assembly conditions

The ϕ6 and ϕ8 PC can be assembled by incubating the component proteins in the presence of 3–6% (w/v) polyethyleneglycol 4000 or 6000 (Poranen et al., 2001; Kainov et al., 2003b). When high enough concentrations of ϕ8 PC proteins (>1 mg/ml) were used, PC-like particles were assembled also in the absence of polyethyleneglycol. The assembly of ϕ6 PC is also dependent on low concentrations of ADP and divalent cations that stabilize the P4 hexamers.

In addition to the complete PCs different incomplete particles can be produced in vitro by omitting one or more proteins from the assembly mixture (Poranen et al., 2001). At high concentrations ϕ8 P1 alone could assemble into PC-like dodecahedral cages (Kainov et al., 2003b). However, the high critical concentration (4.6 g/l for ϕ8) and low stability of these cages indicated that other proteins were likely to control PC assembly via stabilization of intermediates and by lowering the critical concentration for assembly nucleation.

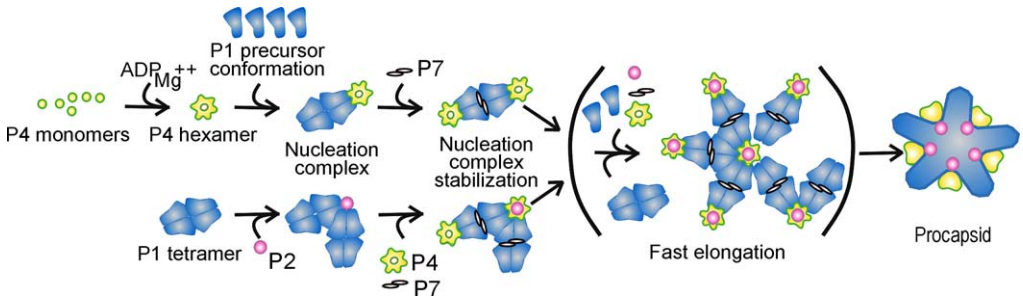


Fig. 3. Summary and comparison of the ϕ6 (above) and ϕ8 (below) in vitro assembly pathways. Both pathways share a P1 tetramer as an obligatory intermediate and proceed via nucleation-limited polymerization. P1 and P4 association into nucleation complex initiates ϕ6 procapsid assembly while nucleation complex of ϕ8 contains P1 and P2. In both cases the nucleation complex is then stabilized by P4 and P7 and the procapsid shell is rapidly completed by addition of individual building blocks. A hypothetical assembly intermediate, depicting how the symmetry mismatch at the five-fold symmetry position of the capsid could be built up, is shown in brackets.

2.2.3. Assembly pathway

An in vitro system allows one to define the minimal set of components required for particle assembly and also to identify rate-limiting nucleation steps. Such systems are also amenable to time-resolved studies, which can provide information about the structure of metastable intermediates that are often populated early on the assembly pathway.

The formation of ϕ6 PC-like structures is dependent on protein P4, suggesting that the P4 hexamer plays a crucial role in the nucleation step (Poranen et al., 2001). In ϕ8 below critical P1concentrations, the P2 polymerase was identified as necessary and sufficient for the formation of particle-like structures, although aberrant assemblies also appeared (Kainov et al., 2003b). The addition of P4 increased the assembly rate and prevented aberrant structure formation. Thus, P2 and P4 are essential for the initiation of assembly. This ensures incorporation of the two key enzymes into the PC (Fig. 3).

The kinetics of ϕ6 and ϕ8 PC formation exhibit a distinct lag phase typical of nucleation-limited assembly (Prevelige et al., 1993). Analyses of ϕ6 kinetics identified formation of a P1 tetramer-P4 hexamer complex as the rate-limiting, nucleation step (Fig. 3) (Poranen et al., 2001). In the ϕ8 system P1 forms stable pre-assembled tetrameric building blocks and the nucleation complex resulted from interaction of several P1 tetramers with a P2 monomer (Fig. 3) (Kainov et al., 2003b). When P7 was included in the ϕ6 assembly system the rate of assembly was significantly accelerated. This suggests that P7 acts as an assembly cofactor, albeit a nonessential one. The presence of P7 accelerated the rate-limiting step of ϕ6 PC assembly and an intermediate composed of a tetrameric P1 connecting two P4 hexamers was identified. Thus, P7 can stabilize the P1 tetramers and accelerate the assembly process (Poranen et al., 2001). In the ϕ8 assembly, P7 had only a small effect on the assembly rate but P4 hexamers significantly increased the assembly rate presumably playing a similar stabilizing role as ϕ6 P7 (Kainov et al., 2003b). Using time-resolved dynamic light scattering distinct intermediates with hydrodynamic radii between 10–12 nm (Poranen et al., 2001) and 8–10 nm (Kainov et al., 2003b) were observed for ϕ6 and ϕ8, respectively. These intermediates correspond to the stabilized nucleation complexes detected by kinetic methods (Fig. 3), that is one P1 tetramer and one P4 hexamer in the case of ϕ6 and 2–3 P1 tetramers together with P2 monomer and P4 hexamer in the case of ϕ8.

A tetrameric P1 intermediate appears to be obligatory for both ϕ6 and ϕ8. The tetramer forms a bridge between two neighboring five-fold vertices and thus facilitates propagation of the structure (Fig. 3). In the ϕ6 system it was possible to identify a nucleation complex containing two P4 hexamers which suggests that the assembly proceeds via a two-fold centered assembly intermediate.

2.2.4. Conformational changes accompanying procapsid assembly

Assembly-induced conformational rearrangements have been demonstrated for many viral systems either by high-resolution crystallographic studies or by spectroscopic methods. Such changes are often required to accommodate non-equivalent interactions in viral capsids (Dokland, 2000; Tuma et al., 2001). The availability of incomplete ϕ6 and ϕ8 PCs has allowed the study of conformational changes associated with the interaction of particular subunits within the PC. Large ordering of α-helices has been shown to be characteristic of P1-P4 interactions (Benevides et al., 2002). The C-terminal tail of P4, which is otherwise disordered in solution, most likely adopts the α-helical structure. The same region has been shown essential for P4 incorporation into the PC (Paatero et al., 1998). Such conformational switching may be necessary for accommodation of the symmetry mismatch between the five-fold vertex of the P1 shell and the P4 hexamer (de Haas et al., 1999).

In addition to P4, P7 undergoes ordering upon binding to PC. On the other hand, interactions of P2 do not lead to any significant changes in secondary and tertiary structures (Benevides et al., 2002). This is consistent with the fully ordered and compact three-dimensional structure of the polymerase (Butcher et al., 2001).

3. Maturation of procapsids to viral cores

PCs package the single-stranded copies of the viral genome segments, and subsequently replicate the packaged ssRNA segments into the mature double-stranded form within the particle (Ewen and Revel, 1988, 1990) (Fig. 2). During this process the particle undergoes substantial conformational changes and obtains its mature, relatively spherical, appearance (Fig. 2) (Butcher et al., 1997). The packaging and the replication of the genome are highly controlled processes and reviewed in detail elsewhere (Poranen et al., 2004).

4. Nucleocapsid shell assembly

The NC surface protein P8 can be purified from virions in a trimeric, highly α-helical, form (Tuma et al., 1999; Poranen et al., 2001). In the presence of millimolar concentrations of calcium, purified P8 can self-assembles into a polydisperse mixture of aberrant, shell-like, structures (M.M. Poranen and D.H. Bamford, unpublished observations). These assemblies are not topologically closed and appear as open spirals in electron micrographs. This suggests that the assembly of closed P8 shells is dependent on the core template. Indeed P8 reassembles onto both the virion-derived and the in vitro reconstituted, RNA-filled, polymerase complexes (Olkkonen et al., 1990, 1991; Poranen et al., 2001). The assembly is also calcium dependent, and no detectable conformational change is associated with the assembly of P8 shell (Tuma et al., 1999) suggesting that P8 subunits assemble in a quasi-equivalent manner.

The P8 shell assembly onto the in vitro assembled and packaged polymerase complexes yields in vitro assembled NC. Because NCs can directly interact with host plasma membrane and thus infect host cell spheroplasts (i.e. cells with the outer membrane and peptidoglycan layers partially removed), the in vitro assembled polymerase complexes can be introduced into the host cell (Ojala et al., 1990; Poranen et al., 2001). The infectivity of the assembly products indicates that the formation of bacteriophage ϕ6 NC does not require any phage or host encoded assembly factors.

5. Envelope assembly

The final step in the ϕ6 maturation is the envelope assembly around the NC. So far no in vitro based system has been devised to study this process, and the knowledge on envelope formation is based on analyses of ϕ6 infected cells and different expression systems.

A membrane protein P9 together with a nonstructural protein P12 constitute the minimum set required for envelopment (Mindich and Lehman, 1983; Stitt and Mindich, 1983a; Johnson and Mindich, 1994; Mindich et al., 1976). The envelope assembly takes place within the cell cytoplasm (Bamford et al., 1976; Ellis and Schlegel, 1974; Gonzalez et al., 1977) and does not involve budding out from the host plasma membrane (Stitt and Mindich, 1983a). The host recognition and attachment protein, P3, which is anchored to the envelope via protein P6, is expressed as a soluble protein and is the last component to be assembled onto the virions (Fig. 2) (Mindich et al., 1979; Stitt and Mindich, 1983b).

6. Comparison to Reoviridae

The eukaryotic Reoviridae and prokaryotic Cystoviridae (including ϕ6 and ϕ8) share a similar genome replication strategy in which the single-stranded genomic precursor molecules are replicated to their double-stranded forms within the viral capsid (Patton and Spencer, 2000; Poranen et al., 2004). Additionally, they share similar core architecture and it has been suggested that these two families may belong to the same evolutionary lineage (Bamford et al., 2002).

The arrangement of the 120 P1 molecules within the core of ϕ6 is similar to that of the bluetongue virus VP3 layer (Grimes et al., 1998) and the reovirus λ1 layer (Reinisch et al., 2000) (see Nibert and Harrison, this issue) as well as the rotavirus VP2 layer (Prasad et al., 1996; Lawton et al., 1997; Prasad et al., 2001). The arrangement of the five-fold vertices of the polymerase complex differs. While the ϕ6 has the P4 packaging hexamer at these positions the reovirus has a turret of the RNA-capping pentamer (λ2) in equivalent locations (Reinisch et al., 2000). By contrast, BTV and rotavirus do not contain similar turret-like structures (Grimes et al., 1998; Prasad et al., 1996). The rotavirus and BTV RNA processing and replication machinery has been localized within the core (Prasad et al., 1996; Gouet et al., 1999). Similarly, the P2 enzyme has been positioned inside the ϕ6 core under the five-fold vertex (Ikonen et al., 2003). Although the atomic structures of several eukaryotic dsRNA virus polymerase complexes are known, in the absence of an in vitro system assembly intermediates have only been postulated. A decamer of VP3 was proposed as a BTV core assembly intermediate (Grimes et al., 1998). It may well be an intermediate albeit a late one, which is preceded by several smaller species. Recent results from the ϕ6–ϕ8 system suggest that a tetramer of the major structural protein P1 in complex with one or more of the key enzymes (the polymerase and the packaging ATPase) is essential for nucleation (Poranen et al., 2001; Kainov et al., 2003b). Interestingly, it was suggested that the viral polymerase of rotavirus (VP3) may nucleate the core assembly (Patton and Spencer, 2000).

The NC shell of cystoviruses is built from P8 trimers in a fashion similar to the BTV VP7 and rotavirus VP6 layer (Grimes et al., 1995; Lawton et al., 1997). The ϕ P8 as well as the BTV VP7 need the core template for correct assembly (Grimes et al., 1997). The outer most capsid layers (a membrane envelope in Cystoviridae) accommodate host

specificity and are not structurally related among different dsRNA viruses.

Acknowledgements

This work was supported by the Finnish Centres of Excellence program 2000–2005 and Life 2000 (R.T.).

References

Bamford, D.H., Burnett, R.M., Stuart, D.I., 2002. Evolution of viral structure. Theor. Popul. Biol. 61, 461–470.

Bamford, D.H., Mindich, L., 1980. Electron microscopy of cells infected with nonsense mutants of bacteriophage ϕ6. Virology 107, 222–228.

Bamford, D.H., Palva, E.T., Lounatmaa, K., 1976. Ultrastructure and life cycle of the lipid-containing bacteriophage ϕ6. J. Gen. Virol. 32, 249–259.

Benevides, J.M., Juuti, J.T., Tuma, R., Bamford, D.H., Thomas Jr., G.J., 2002. Raman difference spectroscopy of the ϕ6 procapsid: Characterization of subunit-specific interactions in a dsRNA virus. Biochemistry 41, 11946–11953.

Butcher, S.J., Dokland, T., Ojala, P.M., Bamford, D.H., Fuller, S.D., 1997. Intermediates in the assembly pathway of the double-stranded RNA virus ϕ6. EMBO J. 16, 4477–4487.

Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H., Stuart, D.I., 2001. A mechanism for initiating RNA-dependent RNA polymerization. Nature 410, 235–240.

Day, L.A., Mindich, L., 1980. The molecular weight of bacteriophage ϕ6 and its nucleocapsid. Virology 103, 376–385.

de Haas, F., Paatero, A.O., Mindich, L., Bamford, D.H., Fuller, S.D., 1999. A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage ϕ6. J. Mol. Biol. 294, 357–372.

Dokland, T., 2000. Freedom and restraint, themes in virus capsid assembly. Structure Fold. Des. 8, R157-62.

Ellis, L.F., Schlegel, R.A., 1974. Electron microscopy of Pseudomonas ϕ6 bacteriophage. J. Virol. 14, 1547–1551.

Ewen, M.E., Revel, H.R., 1988. In vitro replication and transcription of the segmented double-stranded RNA bacteriophage ϕ6. Virology 165, 489–498.

Ewen, M.E., Revel, H.R., 1990. RNA-protein complexes responsible for replication and transcription of the double-stranded RNA bacteriophage ϕ6. Virology 178, 509–519.

Gonzalez, C.F., Langenberg, W.G., Van Etten, J.L., Vidaver, A.K., 1977. Ultrastructure of bacteriophage ϕ6: arrangement of the double-stranded RNA and envelope. J. Gen. Virol. 35, 353–359.

Gottlieb, P., Metzger, S., Romantschuk, M., Carton, J., Strassman, J., Bamford, D.H., Kalkkinen, N., Mindich, L., 1988a. Nucleotide sequence of the middle dsRNA segment of bacteriophage ϕ6: placement of the genes of membrane-associated proteins. Virology 163, 183–190.

Gottlieb, P., Strassman, J., Bamford, D.H., Mindich, L., 1988b. Production of a polyhedral particle in *Escherichia coli* from a cDNA copy of the large genomic segment of bacteriophage ϕ6. J. Virol. 62, 181–187.

Gottlieb, P., Strassman, J., Mindich, L., 1992. Protein P4 of the bacteriophage ϕ6 procapsid has a nucleoside triphosphate-binding site with associated nucleoside triphosphate phosphohydrolase activity. J. Virol. 66, 6220–6222.

Gottlieb, P., Strassman, J., Qiao, X.Y., Frucht, A., Mindich, L., 1990. In vitro replication, packaging, and transcription of the segmented double-stranded RNA genome of bacteriophage f6: Studies with procapsids assembled from plasmid-encoded proteins. J. Bacteriol. 172, 5774–5782.

Gouet, P., Diprose, J.M., Grimes, J.M., Malby, R., Burroughs, J.N., Zientara, S., Stuart, D.I., Mertens, P.P., 1999. The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. Cell 97, 481–490.

Grimes, J., Basak, A.K., Roy, P., Stuart, D.I., 1995. The crystall structure of bluetongue virus VP7. Nature 373, 167–170.

Grimes, J.M., Burroughs, J.N., Gouet, P., Diprose, J.M., Malby, R., Zientara, S., Mertens, P.P., Stuart, D.I., 1998. The atomic structure of the bluetongue virus core. Nature 395, 470–478.

Grimes, J.M., Jakana, J., Ghosh, M., Basak, A.K., Roy, P., Chiu, W., Stuart, D.I., Prasad, B.V., 1997. An atomic model of the outer layer of the bluetongue virus core derived from X-ray crystallography and electron cryomicroscopy. Structure 5, 885–893.

Hoogstraten, D., Qiao, X., Sun, Y., Hu, A., Onodera, S., Mindich, L., 2000. Characterization of ϕ8, a bacteriophage containing three double-stranded RNA genomic segments and distantly related to ϕ6. Virology 272, 218–224.

Ikonen, T., Kainov, D.E., Serimaa, R.E., Timmins, P., Tuma, R., 2003. Localization of replication machinery within the polymerase complex of dsRNA bacteriophage ϕ6. J. Appl. Cryst. 36, 525–529.

Johnson 3rd, M.D., Mindich, L., 1994. Isolation and characterization of nonsense mutations in gene 10 of bacteriophage ϕ6. J. Virol. 68, 2331–2338.

Juuti, J.T., Bamford, D.H., 1995. RNA binding, packaging and polymerase activities of the different incomplete polymerase complex particles of dsRNA bacteriophage ϕ6. J. Mol. Biol. 249, 545–554.

Juuti, J.T., Bamford, D.H., 1997. Protein P7 of phage ϕ6 RNA polymerase complex, acquiring of RNA packaging activity by in vitro assembly of the purified protein onto deficient particles. J. Mol. Biol. 266, 891–900.

Juuti, J.T., Bamford, D.H., Tuma, R., Thomas Jr., G.J., 1998. Structure and NTPase activity of the RNA-translocating protein (P4) of bacteriophage ϕ6. J. Mol. Biol. 279, 347–359.

Kainov, D.E., Pirttimaa, M., Tuma, R., Butcher, S.J., Thomas, G.J., Jr., Bamford, D.H., Makeyev, E.M., 2003a. RNA packaging devices of dsRNA bacteriophages: possibly as simple as hexamer of P4 protein. J. Biol. Chem. 278, 48084–48091.

Kainov, D.E., Butcher, S.J., Bamford, D.H., Tuma, R., 2003b. Conserved intermediates on the assembly pathway of dsRNA bacteriophages. J. Mol. Biol. 328, 791–804.

Koonin, E.V., Gorbalenya, A.E., Chumakov, K.M., 1989. Tentative identification of RNA-dependent RNA polymerases of dsRNA viruses and their relationship to positive strand RNA viral polymerases. FEBS Lett. 252, 42–46.

Ktistakis, N.T., Lang, D., 1987. The dodecahedral framework of the bacteriophage ϕ6 nucleocapsid is composed of protein P1. J. Virol. 61, 2621–2623.

Lawton, J.A., Zeng, C.Q., Mukherjee, S.K., Cohen, J., Estes, M.K., Prasad, B.V., 1997. Three-dimensional structural analysis of recombinant rotavirus-like particles with intact and amino-terminal-deleted VP2: implications for the architecture of the VP2 capsid layer. J. Virol. 71, 7353–7360.

Makeyev, E.V., Bamford, D.H., 2000. Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage ϕ6. EMBO J. 19, 124–133.

Mindich, L., Davidoff-Abelson, R., 1980. The characterization of a 120 S particle formed during ϕ6 infection. Virology 103, 386–391.

Mindich, L., Lehman, J., 1983. Characterization of ϕ6 mutants that are temperature sensitive in the morphogenetic protein P12. Virology 127, 438–445.

Mindich, L., Lehman, J., Huang, R., 1979. Temperature-dependent compositional changes in the envelope of ϕ6. Virology 97, 171–176.

Mindich, L., Qiao, X., Qiao, J., Onodera, S., Romantschuk, M., Hoogstraten, D., 1999. Isolation of additional bacteriophages with genomes of segmented double-stranded RNA. J. Bacteriol. 181, 4505–4508.

Mindich, L., Sinclair, J.F., Cohen, J., 1976. The morphogenesis of bacteriophage ϕ6: particles formed by nonsense mutants. Virology 75, 224–231.

Ojala, P.M., Romantschuk, M., Bamford, D.H., 1990. Purified ϕ6 nucleocapsids are capable of productive infection of host cells with partially disrupted outer membranes. Virology 178, 364–372.

Olkkonen, V.M., Bamford, D.H., 1987. The nucleocapsid of the lipid-containing double-stranded RNA bacteriophage ϕ6 contains a protein skeleton consisting of a single polypeptide species. J. Virol. 61, 2362–2367.

Olkkonen, V.M., Gottlieb, P., Strassman, J., Qiao, X.Y., Bamford, D.H., Mindich, L., 1990. In vitro assembly of infectious nucleocapsids of bacteriophage ϕ6: formation of a recombinant double-stranded RNA virus. Proc. Natl. Acad. Sci. U.S.A. 87, 9173–9177.

Olkkonen, V.M., Ojala, P.M., Bamford, D.H., 1991. Generation of infectious nucleocapsids by in vitro assembly of the shell protein on to the polymerase complex of the dsRNA bacteriophage ϕ6. J. Mol. Biol. 218, 569–581.

Paatero, A.O., Mindich, L., Bamford, D.H., 1998. Mutational analysis of the role of nucleoside triphosphatase P4 in the assembly of the RNA polymerase complex of bacteriophage ϕ6. J. Virol. 72, 10058–10065.

Paatero, A.O., Syvaoja, J.E., Bamford, D.H., 1995. Double-stranded RNA bacteriophage f6 protein P4 is an unspecific nucleoside triphosphatase activated by calcium ions. J. Virol. 69, 6729–6734.

Patton, J.T., Spencer, E., 2000. Genome replication and packaging of segmented double-stranded RNA viruses. Virology 277, 217–225.

Pirttimaa, M., Paatero, A.O., Frilander, M., Bamford, D.H., 2002. Nonspecific NTPase P4 of dsRNA bacteriophage ϕ6 polymerase complex is required for ssRNA packaging and transcription. J. Virol. 76, 10122–10127.

Poranen, M.M., Paatero, A.O., Tuma, R., Bamford, D.H., 2001. Self-assembly of a viral molecular machine from purified protein and RNA constituents. Mol. Cell 7, 845–854.

Poranen, M.M., Pirttimaa, M.J., Bamford, D.H., 2004. Bacteriophage ϕ6. In: Catalano, C. (Ed.), Viral Genome Packaging. Landes Bioscience, in press.

Prasad, B.V., Crawford, S., Lawton, J.A., Pesavento, J., Hardy, M., Estes, M.K., 2001. Structural studies on gastroenteritis viruses. Novartis Found. Symp. 238, 26–37.

Prasad, B.V., Rothnagel, R., Zeng, C.Q., Jakana, J., Lawton, J.A., Chiu, W., Estes, M.K., 1996. Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. Nature 382, 471–473.

Prevelige Jr., P.E., Thomas, D., King, J., 1993. Nucleation and growth phases in the polymerization of coat protein and scaffolding subunits into icosahedral procapsid shells. Biophys. J. 64, 824–835.

Qiao, X., Qiao, J., Mindich, L., 2003. Analysis of specific binding involved in genomic packaging of the double-stranded-RNA bacteriophage ϕ6. J. Bacteriol. 185, 6409–6414.

Reinisch, K.M., Nibert, M.L., Harrison, S.C., 2000. Structure of the reovirus core at 3.6 Å resolution. Nature 404, 960–967.

Sands, J.A., 1973. The phospholipid composition of bacteriophage ϕ6. Biochem. Biophys. Res. Commun. 55, 111–116.

Sinclair, J.F., Tzagoloff, A., Levine, D., Mindich, L., 1975. Proteins of bacteriophage ϕ6. J. Virol. 16, 685–695.

Stitt, B.L., Mindich, L., 1983a. Morphogenesis of bacteriophage ϕ6: a presumptive viral membrane precursor. Virology 127, 446–458.

Stitt, B.L., Mindich, L., 1983b. The structure of bacteriophage ϕ6: protease digestion of ϕ6 virions. Virology 127, 459–462.

Tao, Y., Farsetta, D.L., Nibert, M.L., Harrison, S.C., 2002. RNA synthesis in a cage – structural studies of reovirus polymerase 13. Cell III, 733–745.

Tuma, R., Bamford, J.H.K., Bamford, D.H., Thomas Jr., G.J., 1999. Assembly dynamics of the nucleocapsid shell subunit (P8) of bacteriophage ϕ6. Biochemistry 38, 15025–15033.

Tuma, R., Tsuruta, H., Benevides, J.M., Prevelige Jr., P.E., Thomas Jr., G.J., 2001. Characterization of subunit structural changes accompanying assembly of the bacteriophage P22 procapsid. Biochemistry 40, 665–674.

van Etten, J.V., Lane, L., Gonzalez, C., Partridge, J., Vidaver, A., 1976. Comparative properties of bacteriophage ϕ6 and ϕ6 nucleocapsid. J. Virol. 18, 652–658.

Yang, H., Makeyev, E.V., Bamford, D.H., 2001. Comparison of polymerase subunits from double-stranded RNA bacteriophages. J. Virol. 75, 11088–11095.

Microbial Cell Factories

BioMed Central

Review

Open Access

Folding and assembly of large macromolecular complexes monitored by hydrogen-deuterium exchange and mass spectrometry

Bohumila Suchanova[1] and Roman Tuma*[1,2]

Address: [1]Institute of Biotechnology, University of Helsinki, Finland and [2]Astbury Centre for Structural Molecular Biology, University of Leeds, UK

Email: Bohumila Suchanova - bohuna_suchanova@hotmail.com; Roman Tuma* - r.tuma@leeds.ac.uk

* Corresponding author

Published: 4 April 2008

Microbial Cell Factories 2008, **7**:12 doi:10.1186/1475-2859-7-12

Received: 14 January 2008
Accepted: 4 April 2008

This article is available from: http://www.microbialcellfactories.com/content/7/1/12

Abstract

Recent advances in protein mass spectrometry (MS) have enabled determinations of hydrogen deuterium exchange (HDX) in large macromolecular complexes. HDX-MS became a valuable tool to follow protein folding, assembly and aggregation. The methodology has a wide range of applications in biotechnology ranging from quality control for over-expressed proteins and their complexes to screening of potential ligands and inhibitors. This review provides an introduction to protein folding and assembly followed by the principles of HDX and MS detection, and concludes with selected examples of applications that might be of interest to the biotechnology community.

Introduction

Many essential cellular activities are mediated by large protein complexes. Similarly, viruses, the cellular parasites that are often detrimental to human and animal health, constitute a special but important class of large macromolecular assemblies. Such assemblies are composed of multiple protein subunits and may also contain nucleic acids (e.g. ribosome, viruses, chromatin) and lipids (certain viruses and membrane complexes like nuclear pore, receptors, respiratory complexes). As with any protein successful folding of the subunits and assembly into the final complex are the pre-requisites for attaining activity while non-specific aggregation interferes with function and often triggers a cascade of secondary responses that are detrimental to cells and organisms (e.g. apoptosis, neurodegeneration). In certain cases aggregation produces highly structured and stable assemblies (e.g. amyloid fibrils) that in effect represent an alternative outcome of folding and assembly under the altered conditions [1,2]. In all cases, subunit-subunit interactions determine the outcome of assembly and stability of the final complexes. Given the complexity of large assemblies new approaches to investigate structure and dynamics are needed.

Hydrogen deuterium exchange (HDX), which probes accessibility and local dynamics of polypeptide chains, is a powerful method to study protein folding. With the advent of mass spectrometric (MS) detection the method (further referred to as HDX-MS) became amenable to study assembly and structure of large macromolecular complexes. In this contribution, we first provide basic concepts of protein folding and assembly that are necessary for understanding HDX principles. Then we explain the theoretical basis and practical methods for HDX-MS. In the later part we illustrate how HDX-MS can be utilized for characterizing protein folding, misfolding, aggregation, and for probing interactions in large macromolecular complexes.

Protein folding and assembly

The process of macromolecular assembly starts when the subunit polypeptide chain is made on the ribosome and folds. During early stages of folding the number of accessible conformations is enormous. Experimental studies demonstrated that a rapid collapse of hydrophobic protein core and concomitant formation of secondary structure constitute the early events during folding [3-6]. Detailed tertiary (specific side chain contacts and orientations) and quaternary (subunit interfaces) structures usually develop later. These contacts further stabilize the core and exclude solvent (H_2O) from the protein interior making the hydrophobic core the most stable part of the structure. The formation of the core has been demonstrated by an early protection of the amide protons (NH) against isotope exchange (hydrogen deuterium exchange, HDX, see below for the detailed mechanism) during the folding process. The core also hosts the most stably hydrogen bonded and thus protected peptide groups which exhibit extremely slow amide exchange under equilibrium conditions [7-9]. This is because amide exchange rates are related to the local and global stability [10].

Protein folding can be best explained with the help of multidimensional energy landscape [11,12]. The folding process is described as successive decrease of polypeptide energy to a global minimum on a landscape which, for a typical globular protein, has a funnel-like shape (Fig. 1 shows a one dimensional cross-section projected along a specific folding parameter such as compactness or percentage of native structure) [13,14]. In vitro experiments and theoretical considerations demonstrated that most protein molecules in the ensemble rapidly adopt their lowest energy state following energetically preferred paths while minority is being slowed down by trapping in local minima on the landscape [15,16]. These kinetically trapped molecules are prone to aggregation [17-19]. Evolutionary pressure selects protein sequences against trapping in local minima and consequently most naturally occurring polypeptides exhibit relatively smooth folding landscapes. However, genetic engineering and overexpression of proteins in heterologous hosts might induce additional roughness into landscapes [20]. This has been manifested by numerous expression studies in which folding behavior was derailed by as little as a single point mutation or by a slight change in folding conditions or absence of correct post-translational modifications [21-27].

Many self-assembling and interacting proteins, however, attain their native fold only upon incorporating into the macromolecular complex [28-30]. The energy landscape for folding and association of self-assembling (or aggregating) proteins depends strongly on protein concentrations and ambient conditions (Fig. 1) [29,31-35]. This is

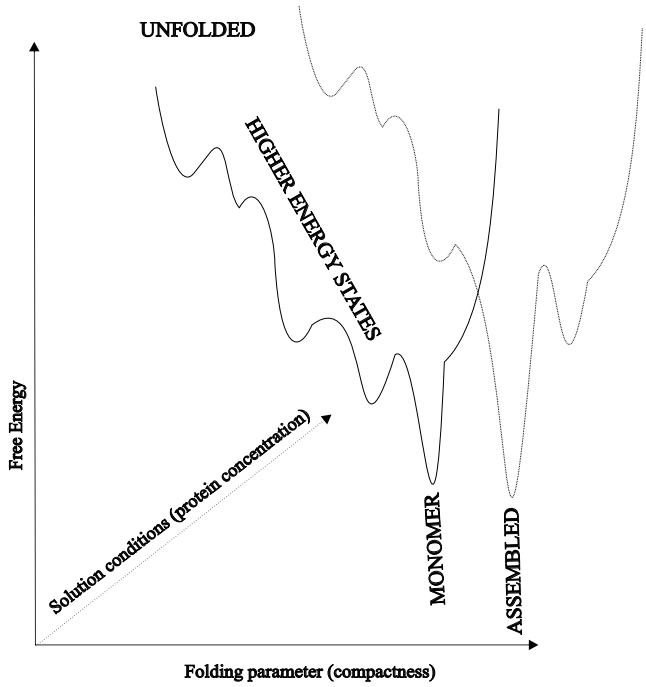


Figure 1
A schematic, one dimensional cross-section of the multidimensional energy landscape projected along a specific folding reaction coordinate.

because assembly is often cooperative and relies on multiple weak interactions between many subunits. Weak interactions are sensitive to changes in environmental conditions like temperature, ionic strength and composition and pH. Since amide protection is related to local protein stability [36] HDX constitutes excellent tool for mapping the assembly energy landscape. For example comparison of HDX for subunits in disassembled and assembled states yields appraisal of stabilization by subunit contacts and provides basis for mapping of subunit interfaces.

Structural characterization of macromolecular complexes

High resolution structure determination relies on experimental techniques such as X-ray crystallography and NMR spectroscopy both of which require substantial protein quantities. The proteins and their complexes are usually produced by expression in a suitable host cell. The most widely used expression hosts are various laboratory strains of *Escherichia coli*, yeast cells or insect cells infected with bacullovirus. Folding of eukaryotic proteins in *E. coli* is often compromised due to the lack of the adequate glycosylation apparatus and appropriate post-translational processing. In such cases one has to resort to expression in cells of higher eukaryots which may significantly decrease the yield to levels insufficient for crystallization or NMR

experiments and consequently lower resolution structural techniques, such as electron microscopy (EM), must be used. HDX is an emerging method belonging to this class. In addition, the stoichiometry and structure of the expressed complexes may differ from those occurring in the original cells and tissues. Hence there is a need for methods which allow comparing structure of the expressed complexes with the native, tissue-derived species. HDX and mass spectrometry are well positioned to fill this gap [37,38].

Electron cryo-microscopy offers a suitable alternative for the structure determination of larger complexes (>1 MDa) [39]. However, EM alone seldom provides high enough resolution (i.e. better than 4Å) that would allow building structural models in atomic details. Cryo-EM routinely achieves resolution as high as 7 Å but the resulting electron densities do not allow polypeptide chain tracing and the primary structure (sequence) cannot be mapped onto the structure. Consequently, this provides little information about the residues at subunit interfaces or about the subunit fold. This problem can be circumvented if high resolution structures of the subunits are known and could be fitted into the EM density [40-42]. However, in cases where the high-resolution structures are not available or substantial portions of the structure are not resolved in the models (e.g. intrinsically unfolded domains which become folded upon assembly) one would like to obtain some information about the disposition of amino acids between folding core, subunit interfaces and intrinsically unfolded regions. As indicated above and discussed in more detail in the next section HDX can provide such information. HDX and mass spectrometry (MS) can also help to tackle sample heterogeneity and classification of EM images [43] and thus provides supportive tool to EM [44].

Theory and Methods

Principles of hydrogen-deuterium exchange

Physical chemistry of amide exchange

HDX probes the exchange kinetics of the main-chain amide (NH) hydrogens for deuterium in samples exposed to D_2O (Fig. 2A). The advantage is that, in principle, it provides site-specific probes along the whole polypeptide chain (except prolines and the N-terminal amino group). Under extreme pH conditions (pH<1 or pH>14) a free amide group can be either protonated or ionized [45]. Amide hydrogens can exchange via those states albeit slowly under physiologically relevant conditions. The reaction is greatly accelerated by an acid or a base. In aqueous solutions the hydronium ion (H_3O^+) or the hydroxide anion (OH^-) plays the role of the acid or the base, respectively [45]. The acid- or base-catalyzed substitution of the unprotected (free) amide hydrogen, i.e. the chemical step of the exchange reaction, is strongly pH- and temperature-dependent (Fig. 2A right panel). This means that an exposed amide group would exchange on millisecond time scale at room temperature and neutral pH. Conversely, it would take about an hour to exchange the same exposed NH at low pH = 2.5 and 0°C [46]. The strong pH dependence of the chemical step allows preserving the pattern of labeling that was attained under native, neutral pH conditions by shifting the sample to pH 2.5 and low temperature. This, in turn provides a window of opportunity for the analysis of isotope incorporation by mass spectrometry [46].

Steric hindrance by the neighboring side chains (R in Fig. 2A) also affects the rate of the chemical step. This dependence was calibrated using short, unfolded peptides as model compounds and constitutes relatively small correction to the intrinsic exchange rates of unprotected peptide groups (further designated as k_2 rate) [47-49]. This correction becomes important when quantitative interpretation for site-specific rates is being sought e.g. determination of the energetics for individual peptide bonds during folding.

Effects of protein secondary and tertiary structures

In folded proteins most of the amide hydrogens participate in hydrogen bonding within secondary structure elements, e.g. α-helices and β-sheets (Fig. 2B). This considerably slows down the exchange since it requires temporal breaking of hydrogen bonds. In the folded state secondary structure is further stabilized by tertiary contacts and thus exchange requires partial or complete unfolding of the protein and exposure of the buried segments to solvent. This process is sometimes called structural opening and is a prerequisite for the chemical step of exchange [45]. The open states in Fig. 2B correspond to higher-energy, partially unfolded intermediates on the energy landscape in Fig. 1. Quaternary, subunit contacts provide additional stabilization and protect NH sites on or close to the subunit interfaces [50,51]. In a typical well-folded protein the exchange rates may differ by many orders of magnitude reflecting local and global stability [46].

Two step model of amide exchange

The exchange rates are governed by the folding energy landscape. However, the correspondence is not straightforward and it is necessary to consider all factors contributing to the overall exchange rate. In order to simplify such analysis several reasonable assumptions and approximations are made. For amides within protected regions (secondary structure, hydrophobic core, subunit interfaces) structural opening is the rate-limiting step and the exchange process may be considered as a two step reaction (Fig. 2C): (1) structural opening which is governed by rate constants k_1 and k_{-1}. (2) chemical exchange step which is

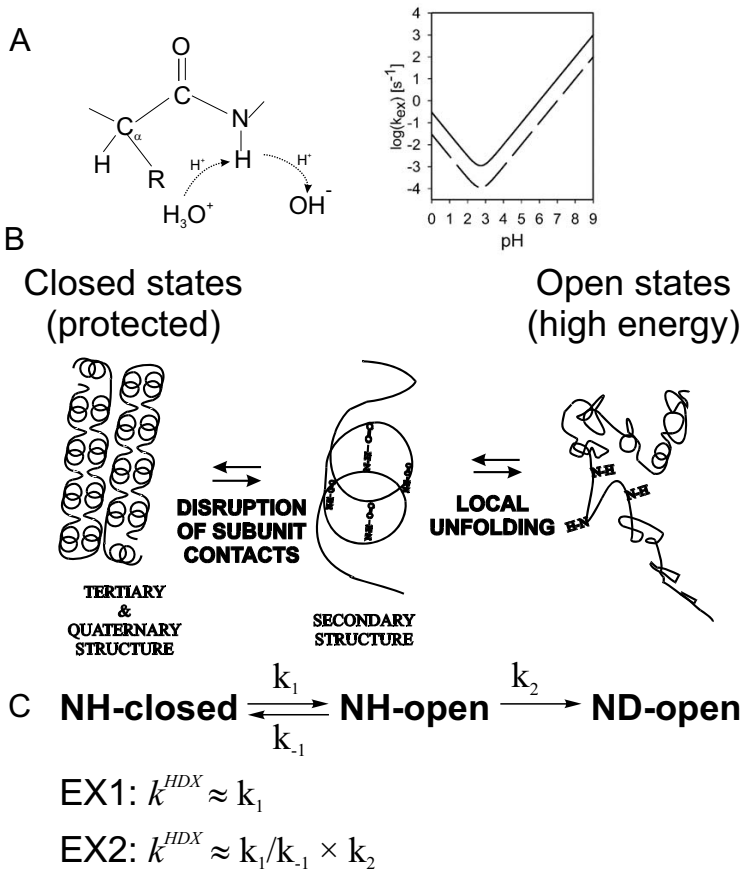


Figure 2
Mechanism of HDX. (A) Mechanism of the chemical step. Inset on the right illustrates the pH and temperature (solid line 25°C, dashed line 0°C) dependence of the intrinsic exchange rate (B) Influence of quaternary, tertiary and secondary structure on HDX. (C) Overall reaction scheme and the two limiting cases.

governed by rate k_2 and considered irreversible under the condition of excess D_2O. Then the overall HDX rate can be described by:

$$k_{ex} = k_1k_2/(k_2+k_{-1}) \quad (1)$$

There are two limiting cases that are considered when interpreting HDX exchange rates:

(EX1) $k_2 >> k_{-1}$, i.e. the chemical exchange is fast compared to the structural opening step:

$$k_{ex} = k_1 \quad (2)$$

The EX1 limit is typical for exchange via global unfolding under strongly denaturing conditions or for subunit dissociation in large assemblies. The rate does not depend on pH. The temperature dependence of the rate yields the activation energy for the unfolding event, i.e. provides information about the heights of barriers (roughness) on the energy landscape.

(EX2) $k_{-1} >> k_2$, i.e. the open state is short lived compared to the rate of the chemical step:

$$k_{ex} = k_1k_2/k_{-1} = K_{op}k_2 \quad (3)$$

where K_{op} is the apparent equilibrium constant for structural opening. EX2 is typical for exchange from the folded state under native or mildly denaturing conditions and is strongly pH dependent (through k_2). EX2 is a product of the local unfolding equilibrium constant and the chemical step rate constant. When site-specific exchange rates in unfolded state are available it is possible to factor them out and to obtain the protection factor:

$$p = k_2/k_{ex} \quad (4)$$

This factor is then related to the local stability, i.e. depths of "wells" on the energy landscape:

$$\Delta G_{local} = -R \cdot T \cdot \ln(p) \quad (5)$$

where R is the universal gas constant and T is absolute temperature.

Because different mechanisms of local opening usually operate simultaneously, neighboring amides within the same element of secondary structure may exchange with different rates under native conditions. Upon destabilization under progressively more denaturing conditions, these rates converge to a value that is characteristic of the cooperative unfolding unit to which these sites belong [36,52]. This approach identifies structural characteristics of partially unfolded states and enables to map the energy landscape in structural terms [10].

HDX detection by mass spectrometry

We first compare various methods for HDX detection and then explain practical issues involved in HDX-MS measurements. More detailed discussion of experimental implementation has been provided in recent reviews [46,53,54].

Comparison of different methods for HDX detection

Several experimental techniques for measurement of HDX are available. NMR, which determines HDX through disappearance/appearance of amide proton resonances in D_2O/H_2O solutions, is the most widely used and provides site-specific probe of protection. Presently, NMR is not applicable to large protein complexes (>200 kD) and requires high concentrations (>10 mg/ml). HDX dynamics of large macromolecular complexes can be readily examined by Fourier transform infrared (FTIR) [55-60] and Raman spectroscopy [29,32,61,62]. These techniques probe incorporation of deuterium by monitoring shifts of frequencies arising from collective vibrations of the peptide bond and the amide NH group (so called amide bands, deuterium incorporation increases the local mass and hence these bands shift to lower frequencies). These techniques differentiate the exchange rates only by the type of secondary structure and do not provide mapping onto the protein sequence.

Mass spectrometry measures deuterium incorporation as an increase of apparent mass, more specifically as a change in the isotopic composition. Advances in protein MS enabled routine resolution of isotopic composition for oligopeptides with masses up to several kDa. Similarly, shifts due to deuterium incorporation are readily discernible in mass spectra of intact proteins. This in turn allowed to probe HDX kinetics for large proteins [63,64] and macromolecular complexes [33,65-68] and map subunit-subunit interactions in assemblies [51,69-72]. Clear advantages of MS detection over NMR are the lower protein concentrations and sample quantities needed. Although, in principle, MS allows for residue-specific resolution of HDX kinetics this is seldom achieved and only region-specific information is obtained [53,73,74].

Data acquisition and instrumentation

The extent and rate of HDX is measured from mass increases of peptide fragments after enzymatic cleavage of a protein (Fig. 3). The exchange is usually initiated by diluting the protein into D_2O exchange buffer (Fig. 3A).

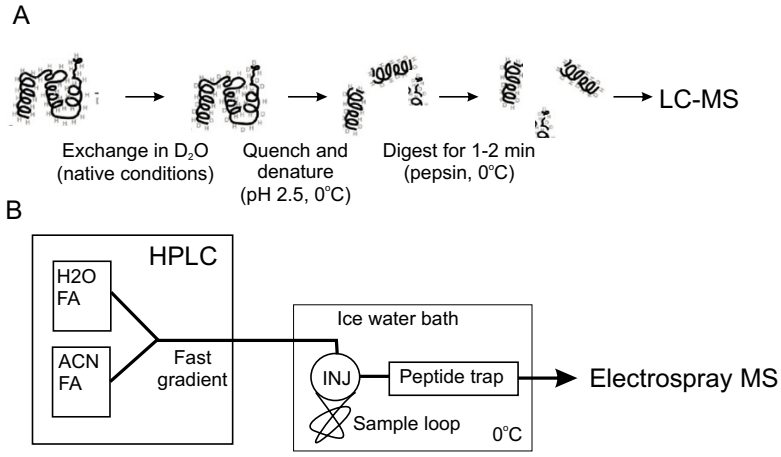


Figure 3
Practical aspects of HDX measurement. (A) Simplified exchange and protease digestion protocol. (B) Cooled LC-MS setup, INJ = injector [64].

The sample is incubated under the exchange conditions for the desired exchange period (usually 30 s to 10 h) and then quenched by rapid acidification to pH 2.5 on ice (usually done with formic acid). This effectively slows down any further exchange (so called exchange-in) and minimizes pickup of hydrogens during subsequent handling in H_2O solutions (so called back-exchange). The quenching solution may be supplemented with a denaturant (e.g. guanidine hydrochloride or urea) to facilitate dissociation of stable complexes prior to proteolytic digestion. The quenched sample may be analyzed immediately or flash-frozen and stored for up to several weeks in liquid nitrogen for off-line analysis.

Prior to MS analysis the sample is digested with a non-specific acid protease (e.g. pepsin, Aspergillus type XIII, Rhizhopus type XVIII, or their mixture [75]) on ice. The resulting digest is loaded onto a reverse-phase column (usually short C-8 or peptide trap) and rapidly separated by a steep gradient. As shown in Fig. 3B it is essential that the HPLC column and the injector are cooled and kept at constant temperature (e.g. in an ice-water bath) to minimize the back-exchange and maintain reproducible conditions [76]. With the advent of high-resolution MS, good HPLC separation is not essential. The chromatography step effectively provides rapid desalting and removal of undigested protein before sample introduction into the mass spectrometer. This is usually done through a standard or micro-capillary electrospray ionization interface (ESI) [77,78]. Any mass spectrometer with resolving power 5000 or better (e.g. a quadrupole-time of flight (Q-TOF), an ion trap) is in principle suitable for detecting HDX. Higher resolution enables a more accurate measurement of multiply charged peptides and is essential for resolving overlapping isotopic envelopes that are commonly found in MS spectra of large proteins and their assemblies. High mass accuracy and reproducibility, usually provided by calibration with a suitable standard, are essential. The highly repetitive nature of HDX-MS experiment (i.e. collection of many time points) makes it suitable for automation and several setups were reported [53,79].

Note that the exchangeable side-chain hydrogens (e.g. tyrosine or serine OH groups, the amino group of lysine etc.) exchange rapidly back during digestion and HPLC separation performed in H_2O [80]. On the other hand, deuterium at main-chain amide sites persists much longer (see Fig. 2A). However, back-exchange of main-chain amides, which inevitably occurs during sample handling after quenching (i.e. peptic digestion and reverse-phase separation), interferes with HDX data analysis by diminishing the differences between exchanged populations and needs to be kept as low as practically possible. Recently, several attempts were made to alleviate the back-exchange problems during separation. In principle ultra performance liquid chromatography (UPLC) provides faster and superior resolution. Comparison between UPLC and conventional HPLC in HDX/MS experiments using cytochrome c as the benchmark protein failed to provide convincing evidence in favor of UPLC mainly due to higher back-exchange. This was caused by inadequate cooling of the injector in the commercial UPLC system [81]. A radically different approach using a supercritical fluid chromatography (SFC) that employs non-aqueous CO_2 mobile phase and fast flow rates yielded substantial reduction of back-exchange compared to the HDX-optimized HPLC separation. The slightly impaired peptide separation in SFC was offset by the superior resolving power of the Fourier transform ion cyclotrone resonance mass spectrometer (FTICR-MS) which was used in this study [82].

Data analysis-from mass spectra to structural interpretations

Digestion with pepsin or other acidic proteases is rather non-specific and under given conditions (e.g. digestion time, denaturant concentration, protease-to-protein ratio) yields overlapping fragments with lengths ranging from 5 to 15 amino acids. The same set of peptide fragments is obtained for a given protein and digestion conditions [75]. This assures that consistent and reproducible data sets are obtained in independent runs. Each of the fragments acquires one or more positive charges during ionization and produces a typical isotopic envelope in the mass spectrum (Fig 4A). The relative abundance of different ionization states and fragments strongly depends on the particular ionization interface and mass spectrometer design. Hence, it is advisable to use one MS instrument throughout the whole study.

Each peak in the envelope corresponds to an increasing number of heavier isotopes (naturally occurring ^{13}C, ^{34}S and the HDX introduced ^{2}H) and spacing between them is inversely related to the acquired charge. Even the non-exchanged control (t = 0 s) contains series of peaks corresponding, in succession, to the monoisotopic fragment mass (the lowest m/z, all atoms are of the lightest isotope) followed by the peaks from species containing one or more of the heavier, naturally occurring isotopes. Note, that due to 1% abundance of ^{13}C and other naturally occurring isotopes the monoisotopic peak is usually weak for longer fragments (>10 residues).

The fragments in non-deuterated sample are assigned to the amino acid sequence by tandem MS (MS^2; i.e. fragmentation of the selected peptide in the mass spectrometer, usually performed by collision with gas or application of strong electric field, followed by analysis of the resulting ions; see [77] for details) [83]. In favorable cases, especially when using ultra-high resolution mass

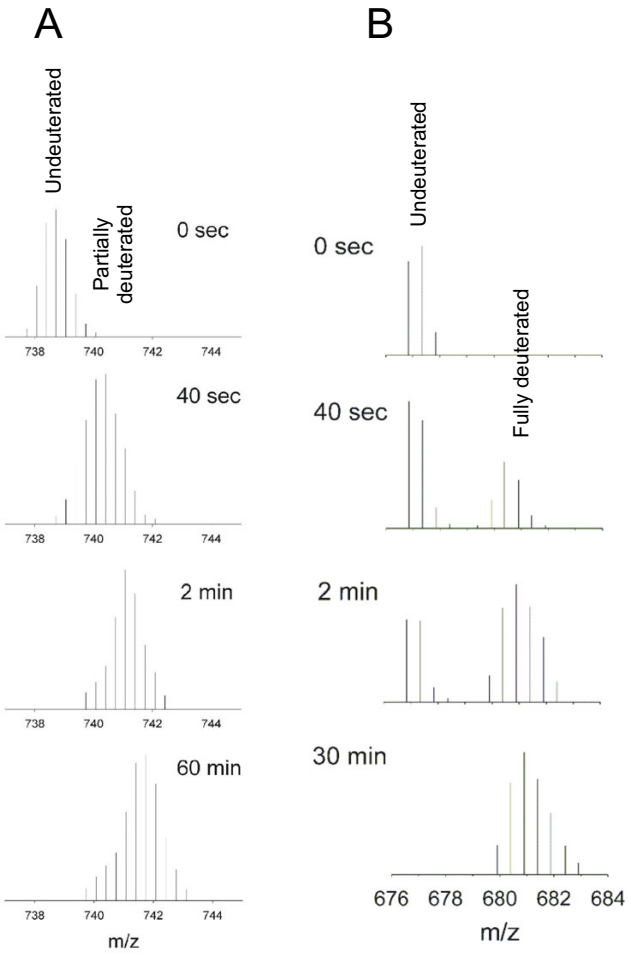


Figure 4
Example of HDX detected by mass spectra. m/z isotopic envelopes for EX2 (A) and EX1 (B) limit of exchange. Data obtained for a region that is situated within the subunit interface in φ8 P4 hexamer [114].

spectrometers (e.g. FT-ICR), assignments may be based on the exact mass alone. However, a confirmation by MS^2 is still desirable [84].

As expected, deuteration shifts the isotopic pattern to higher m/z values (Fig. 4A). As seen in the second spectrum from top the monoisotopic peak may not be resolved in the shifting isotopic envelopes and hence the patterns are usually identified by the expected progression of m/z peaks for the given charge state. One of the main advantages of MS detection is that MS spectra exhibit distinctly different patterns for EX1 and EX2 mechanisms. This provides basis for distinguishing between the two limits without measuring the tedious and often experimentally inaccessible pH dependence.

In the EX1 regime amides remain either all unexchanged or become all exchanged at once and thus the mass spectra show a bimodal pattern (c.f. t = 0 and 40 sec in Fig. 4B). The apparent rate constant, k_{ex} = k1 in Eq. 2, corresponds to the rate of structural opening or a conformational change, and it is determined from the time-dependence of the integrated intensity belonging to the deuterated envelope.

In the EX2 regime exchange causes progressive shift and binomial broadening of the isotopic envelope (Fig. 4A). In this case the centroid of the isotopic envelope or the average of the peak positions weighted by their intensities are both reasonable and frequently used measures of the degree of deuteration [46,51,64,85,86]. In this case the HDX rate constant, k_{ex} in Eq. 3, is obtained from the dependence of the centroid/average position on exchange time. This analysis provides kinetics averaged over the whole fragment and consequently corrections for the effects of primary structure (sequence) on k_2 in the EX2 limit are impossible. Such correction is not necessary when comparing two states of the same protein, e.g. for mapping subunit interfaces in protein assemblies.

In certain cases (e.g. high resolution spectra like those shown in Fig. 4) the true deuteration distribution may be obtained from the measured envelope by deconvolution with the naturally occurring isotope distribution [87]. Such procedure allows analyzing exchange under conditions when part of the fragment is exchanging via EX1 mechanism while the rest belongs to an EX2 limit [88].

Site-specific resolution of exchange will be essential for quantitative study of enzymes and molecular motors and for direct comparison with theoretical results [89-91]. Several attempts to improve the resolution have been made. One approach was to analyze simultaneously the exchange kinetics of overlapping fragments and separate the individual contributions computationally. In favorable cases this may produce site-specific resolution for few residues within the entire sequence. An attempt at more reliable experimental solution was made by using MS^2 for sequencing of deuterated fragments. However, it was found that the collision-induced fragmentation step led to significant scrambling of deuterium among the amide sites and hence this approach proved impractical [74]. Recent experiments using gentler fragmentation methods demonstrated significant preservation of the deuteration pattern [73,92].

In all cases the measured degree of exchange is always lower than the actual value because some of the label is lost during sample handling (e.g. digestion and HPLC) due to the back-exchange. The true extent of exchange is an important parameter for interpretation. Take the fol-

lowing example: In the absence of site-specific resolution the exchange is averaged over the whole peptide fragment. Hence, situations, in which half of the fragment residues are exposed and exhibits fast exchange while the rest is buried and does not exchange at all, are common. Ignoring the true extent of the exchange, which amounts to only half of the available sites in this example, and taking into account only the fast kinetics would lead to an erroneous conclusion that the whole peptide is exposed.

In order to obtain the true extent of exchange one must correct for the back-exchange. The degree of back-exchange depends on the experimental setup and sample handling and varies from as little as 10% for ESI-MS to about 40% in MALDI-TOF MS detection. There is also variation between different fragments, presumably due to sequence effects on the back-exchange. Therefore, the best way to correct for back-exchange is to measure a fully deuterated control sample under identical conditions, e.g. a protein which was first denatured and then re-folded in D_2O buffer. For some proteins refolding in D_2O is impossible and one must resort to approximations. For example, it is fair to qualitatively compare the relative extent of exchange for two states of the same protein (e.g. assembled and free subunit of a virus) if the two data sets were collected under identical conditions [70].

Automated data processing

Mass spectrometry is venerable for generating large data sets and HDX-MS is no exception. A typical 35 kD protein may yield about 100 assigned fragments which contribute to the mass spectra. In addition, the spectra contain plethora of unassigned peaks e.g. from oxidized protein or from pepsin. In HDX-MS the useful MS data is spread over several (4–10) scans within the LC-MS run. Usually, the full kinetics contains at least 15 time-points collected in three independent replicas. Taken together one needs to extract and process at least 4000 isotopic envelopes from about 200 spectra per each kinetic curve. Such a task calls for a considerable automation and deployment of data-mining tools.

MS^2 peptide assignments can be done using either commercial proteomics add-on packages for automated database searches and subsequently validated by spectral prediction tools when necessary [93]. Despite the apparent popularity of HDX-MS, specific tools are not to be found in any of the commercial software packages that are currently supplied with instruments. This spurred in-house development of various HDX-MS-specific software packages [54,79,94] and eventually led to applications ranging from semi-automated to fully-automated processing that are now freely available [88,95,96]. The utility of some packages is still hindered by the plethora of proprietary data formats used by different manufacturers but could be overcome by adopting the mzXML standard for data interfacing [96,97]. Recently, Pascal and colleagues launched a web application ("Deuterator") compatible with multiple file formats for automated HDX-MS data analysis [96,98]. Although promising, the utility of the web-based approach is somewhat compromised by the need to supply large datasets to a remote server.

Applications

In the following sections we present selected examples of HDX-MS use in detection and characterization of folding intermediates, association of small oligomers and dynamics in large assemblies. Later we provide illustrations of HDX-MS utility for structural characterization of aggregates and intrinsically unstructured proteins. Finally, we discuss applications to characterize functional dynamics in protein complexes and molecular machines.

Detection of folding and association intermediates in vitro

Native state exchange reveals folding intermediates

Multi-state and two-state protein folding cannot be often distinguished due to instability of partially folded intermediates. Meta-stable intermediates can be detected via HDX and play a crucial role in the determination of folding kinetic and native state dynamics [85]. HDX experiments are also considered most effective in teasing out the structural details of protein folding intermediates, often with amino-acid resolution [99].

The three-state unfolding of ubiquitin (highly structured, partially unstructured A-state, and fully unstructured state) was examined by HDX-MS [100]. The highly dynamic A-state consisted of flexible, rapidly exchanging C-terminal region while the N-terminal domain adopted less dynamic, native-like β-strand configuration.

Time-resolved techniques and characterization of folding and assembly intermediates

Pulse-labeling HDX-MS techniques enable detection and characterization of transient folding intermediates which are not readily populated under equilibrium conditions [101]. An on-line HDX pulse-labeling ESI MS apparatus was developed and tested on myoglobin folding and heme incorporation [86,102] (Fig. 5). The method is based on tandem mixing chambers: the first one serves to initiate folding/assembly reaction while the second one is used to stop the folding reaction and transiently expose the products to deuterium label. Subsequently, the exchange is quenched and the incorporated label is quantified by MS. This provides a "snapshot" of protection at a particular stage of the reaction. Usually, this is applied to intact proteins but there is no principal obstacle in carrying out pepsin digestion and LC-MS analysis and obtain region-specific exchange kinetics.

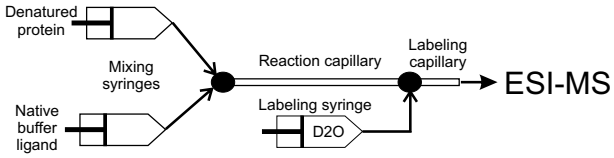


Figure 5
Schematics of on-line folding and pulse labeling apparatus (based on [102]).

The method enabled millisecond resolution of the folding dynamics for a well-studied model protein ubiquitin and led to the discovery of a previously uncharacterized intermediate [103]. The method was then used to follow folding and association of dimeric protein, S100A11, and revealed three different kinetic species – a relatively unfolded monomer, a more compact folded monomer and the native dimer [104].

Pulse labeling can also reveal the interplay between folding and assembly in multidomain, multimeric proteins. Rabbit muscle aldolase is a homotetramer encompassing subunits with three domains each. A locally unfolded aldolase was pulse-labeled with deuterium after destabilization in urea for a defined period of time (1 min to 48 h). Isotope patterns revealed four populations corresponding to distinct conformations: one native and three distinct, partially unfolded intermediates. The intermediates, which were further characterized using pepsin digestion, corresponded to aldolase with one, two or all three domains unfolded, respectively. Remarkably, the completely unprotected intermediate, which also lacked secondary structure, still remained tetrameric, suggesting strong coupling between folding and assembly of the tetramer. The observed HDX kinetics supported sequential and cooperative mechanism of aldolase unfolding [51,80,105].

Structural characterization of expressed proteins

When there is no high resolution structure available, HDX-MS experiment can provide valuable information about the domain structure and may be used to validate structural models which are built solely on the basis of limited sequence homology. This approach was applied to a protein kinase scaffolding protein, D-AKAP2, which augments interactions of signaling molecules inside cells [106]. HDX-MS revealed two regions of low exchange located in the midst of fast exchanging regions and identified two distinctly folded regions (Fig. 6). Fig. 6A illustrates a plausible representation of region-specific HDX kinetics in the absence of three-dimensional structure. The sequence, which was found homologous to regulator of G-protein signaling (RGS), mapped within the first folded region. The second folded region encompassed a highly protected protein kinase A (PKA) binding site and a less protected PDZ-binding motif (PDZ domain is a potential target of D-AKAP2). HDX-MS thus confirmed the multi-domain architecture of D-AKAP2. HDX-MS was also used to validate a homology-based structural model of the RGS domain (Fig. 6B–C) [106].

An interesting application of HDX-MS has been developed to follow protein folding during in vitro, cell-free, transcription-translation [107]. This is a fast and easy way to use method based on MALDI-TOF and yields global degree of folding using minute protein quantities without purification. In addition, it affords rapid identification of the expressed protein. HDX-MS was also used to refine protein constructs for crystallization trials [108,109]. A somewhat similar MALDI-MS approach was used to characterize stability of an expressed protein in the cell extracts i.e. under conditions close to those found in the cytoplasm [38]. The method, abbreviated by the authors as SUPREX (stability of unpurified proteins from rates of H/D exchange) revealed that the model protein, bacteriophage λ repressor, exhibited identical stability in the cell lysate and in the dilute solution of pure protein, respectively. This study was later extended to perform and detect HDX of the λ repressor in the cytoplasm of intact *E. coli* cells, i.e. *in vivo* [37]. This method exploited permeability of lipid membranes for small molecules like water and urea which allowed for rapid equilibration of D_2O between the cytoplasm and the deuterated medium containing increasing amounts of denaturant. As in their previous study the authors found no difference between repressor stability under *in vivo* and *in vitro* conditions, respectively. However, the *in vivo* stability was significantly enhanced by administering hyperosmotic shock to the cells prior to the exchange experiment.

Assembly and dynamics of large complexes

Macromolecular complexes, some of which are indeed fairly sophisticated molecular machines, ensure consecutiveness of cellular processes such as macromolecular synthesis, transport, and metabolism. Their functions rely on self-assembly, subunit rearrangements and conformational changes throughout the duty cycle. Viruses represent special class of such machines and, in effect they could be considered smart containers for targeted delivery of macromolecular cargo. They are programmed for controlled replication, encapsidation, transport and release of their genomes into new host cells. This is accomplished by series of concerted structural changes within viral capsids. The utility of HDX-MS in virus research and in characterization of macromolecular complexes in general was recently reviewed [71] and here we present selected examples, mostly from the virus field, to illustrate the type of problems this method may help to answer.

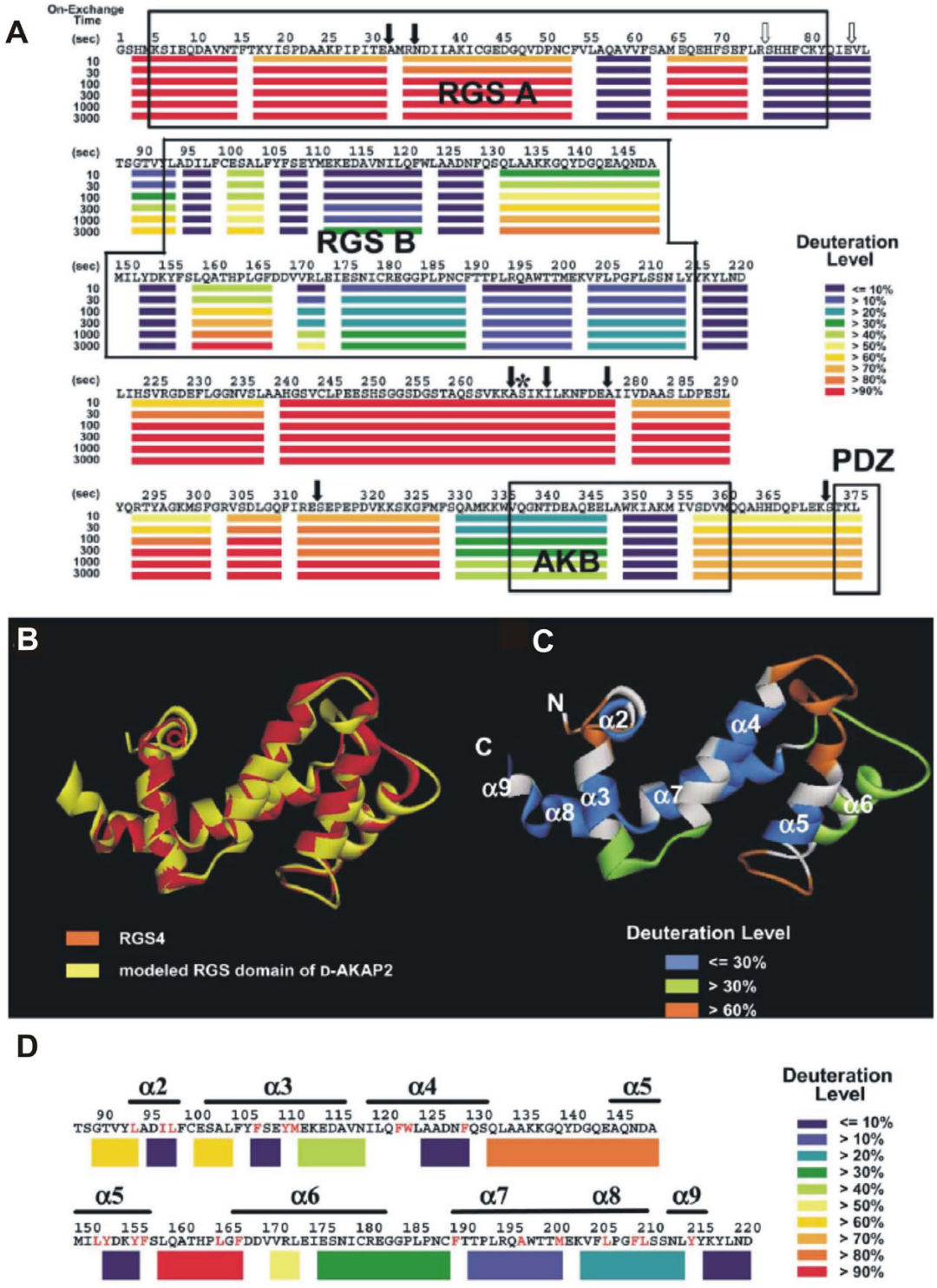


Figure 6

Domain structure of D-AKAP2 scaffolding protein. (A) Graphical representation of deuteration levels at different time points is shown in color (scale on the right) below the sequence. (B) Backbone homology-based model of the RGS domain (C) Color representation of deuteration levels after 3000 s mapped onto the modeled structure. (D) Deuteration levels after 3000 s mapped onto the primary structure. From [106] with permission.

Mapping subunit interfaces and association dynamics

The first step towards functional macromolecular complex is assembly of subunits guided by subunit-subunit interactions. In many instances only the high-resolution structures of subunits are available while a medium resolution electron density of the whole assembly is readily obtained by cryo-EM. The fitting of the subunit structures into the EM density is greatly facilitated by knowing the subunit interfaces within the complex. This strategy was adopted for bacteriophage φ12 which is a dsRNA virus belonging to the *Cystoviridae* family and is structurally related to members of the *Reoviridae* family [110]. Although the high-resolution structure of the virus-associated packaging ATPase, a hexameric molecular motor P4, had been known [111] and the hexameric structure had been resolved by cryo-EM [112,113] it was not possible to determine which way the hexamer interacted with the viral procapsid (Fig. 7A). HDX-MS was used to map the subunit interfaces [94]. Fig. 7B shows comparison of the isotopic patterns for the C-terminal helix in the free hexamer (red) and the PC-bound motor (blue). A substantial increase in protection (lower final plateau in Fig. 7C) and slower kinetic (smaller initial slope) is clearly apparent for the assembled state. Mapping of the average exchange rates for all resolved fragments onto the surface representation of the hexamer confirmed that P4 associates with the procapsid using its C-terminal facet (Fig. 7D). This information was subsequently used to fit the atomic model of the hexamer into a refined asymmetric cryo-EM reconstruction of a related virus φ6 [113] (Fig. 7E).

HDX-MS also revealed that the association of the motor with the viral capsid stabilized subunit interfaces (c.f. exchange of helices IH1 and IH2 in Fig. 7F) and prevented spontaneous opening of the hexameric ring (Fig. 7G). Ring opening had been detected by HDX-MS and associated with loading of ssRNA into the central channel (Fig. 7G) [114]. The exchange dynamics of the PC-associated hexamer closely followed temperature B-factors derived from the crystal structure while the free, solution state was more dynamic [94] (Fig. 7F).

HDX-MS and crosslinking was used to identify new contacts between subunits in the capsid of human immunodeficiency virus (HIV) [70]. The subunit of HIV capsid (CA) encompasses N-terminal domain connected via a linker to C-terminal domain. The capsid assembly process was thought to be driven by association (dimerization) of C-terminal domains and hexamer formation by six N-terminal domains. While the C-terminal domain was known to form stable dimers in solution there was no evidence for the oligomerization of the N-terminal domain. Hence, the assembly mechanism was far from being certain. By comparing the exchange of the full-length CA in the monomeric and the assembled state Lanman and colleagues were able to delineate a new interface between the N-terminal and C-terminal domains of neighboring subunits [70,115]. This contact was shown to play essential role in stabilizing hexamers during assembly.

HDX-MS can also provide information about subunit association kinetics in functional assemblies. Small heat shock proteins (sHSP) belong to a family of molecular chaperones, which transiently bind partially unfolded proteins and prevent their aggregation. One member of this family, HSP16.9, assembles into a dodecamer at room temperature. Surprisingly, subunit interfaces exhibited no protection against HDX after incubation for 5 s in D_2O [50]. Note, that under these conditions most oligomeric proteins show significant protection of their subunit interfaces. Almost complete exchange suggested that large conformational motions were taking place within the assembly, leading to the disruption of subunit interfaces. When pulse-labeled for just 10 ms, the subunit interfaces showed significantly lower exchange comparing to the 5 s experiment. Thus, the HSP16.9 oligomer underwent fast association-dissociation dynamics on a sub-second time scale. At 42°C, HSP16.9 forms a dimer with the same exchange pattern and kinetics as the dodecamer at room temperature. It is believed that HSP16.9 is heat-activated by shifting the equilibrium between the two forms [50] and the fast association-dissociation dynamics plays essential role in the process. This example also shows that it is important to consider HDX kinetics on millisecond time scale when dealing with metastable, highly dynamic complexes.

Structure and dynamics of subunits within large assemblies

EM and X-ray diffraction yield structures of well-defined, stable conformations. However, even highly symmetric, icosahedral viral capsids are dynamic entities. Essential biological processes, like delivery of viral genomes, are mediated by structural transformations. HDX-MS enables monitoring of these dynamic events.

Many viruses require a maturation step in which a capsid precursor undergoes large scale structural rearrangements in response to e.g. protease processing of the viral polyprotein. HIV maturation is triggered by a virion-associated protease which specifically cleaves the assembled Gag polyprotein to releases the individual structural proteins (NC, nucleocapsid; CA, capsid; MA matrix; ENV, envelope glycoprotein). During the process the assembled CA collapses from a spherical form into a conical core. This step is essential for infectivity and hence the protease has been successfully targeted by several antiviral drugs. However, in structural terms, the maturation process has been elusive [115]. HDX-MS was used to compare CA in the mature and immature form of virus-like particles (VLP). The intact CA protein in the mature form exhibited signif-

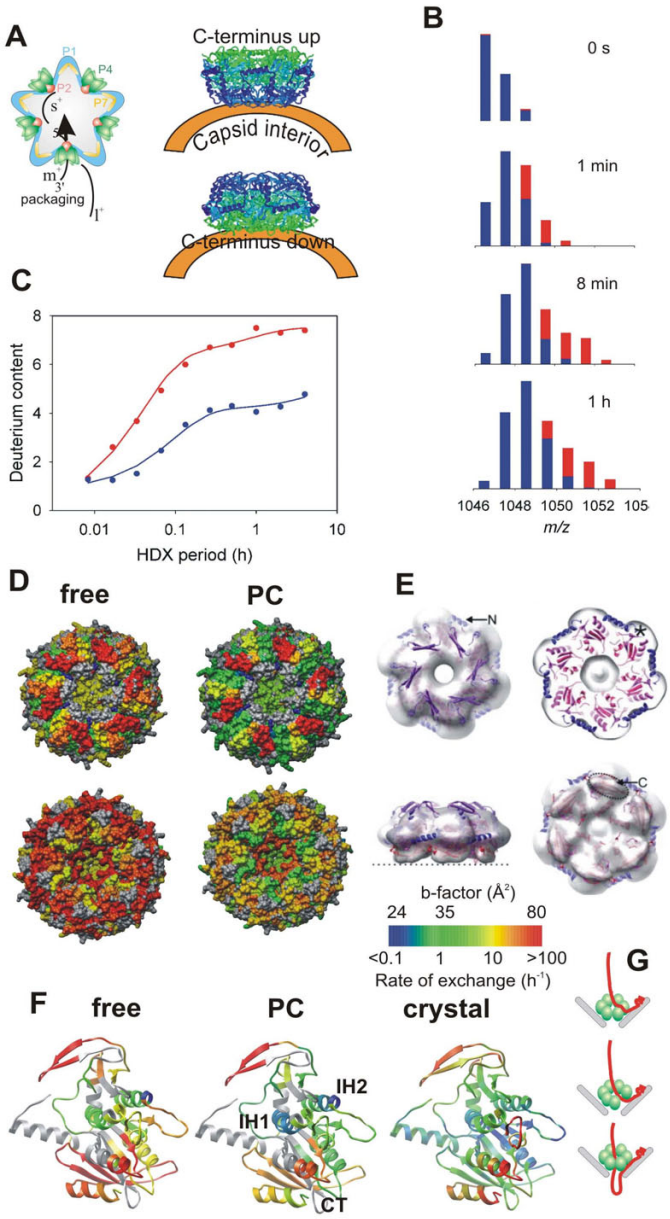


Figure 7

Mapping subunit interfaces and dynamics in large viral assembly- P4 hexamer within φ12 procapsid [94]. (A) Schematics of the viral procapsid (PC) and packaging of ssRNA precursors by the P4 hexamers associated with the capsid vertices (left). The right panel illustrates the problem of determining the orientation of the hexamer with respect to the procapsid. (B) Bar representation of the isotopic envelopes during progressive deuteration, free hexamer in red and PC-bound in blue. (C) The HDX kinetics derived from centroid positions of the isotopic envelopes in panel B (and from additional data not shown). (D) Mapping of average exchange rates onto P4 structure in surface color representation (color scale on the bottom right). (E) Fitting of the P4 hexamer into the EM-derived electron density using the correct orientation from HDX experiment (adopted from [113]). (F) Average exchange rates mapped onto the ribbon diagram of P4 subunit for free (left) and PC-bound (middle) hexamer together with the scaled crystallographic temperature factors (B-factors, right panel). Color scale as in panel D. (G) Schematics of ssRNA loading into the hexameric packaging motor via ring opening mechanism. Top: A specific viral RNA structure (packaging signal) is recognized by the major capsid protein P1 which brings the polynucleotide strand to the vicinity of P4. Middle: P4 ring opens and lets the RNA slip in between subunit interfaces into the central channel. Bottom: The ring topologically encloses the bound RNA and translocates RNA in 5' to 3' direction into the capsid at the expense of ATP hydrolysis (packaging).

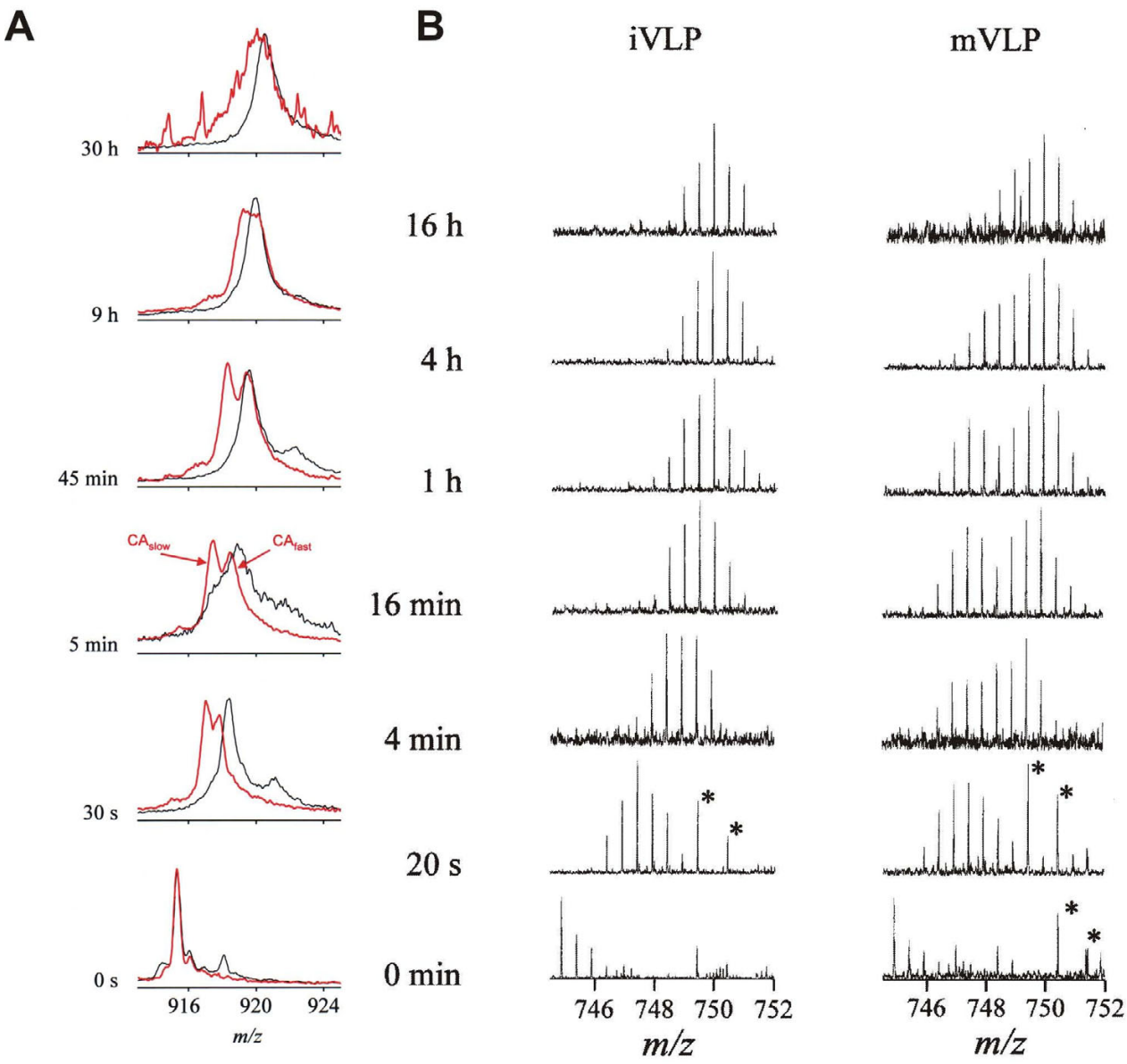


Figure 8

Structural changes during HIV maturation [71, 115, 116]. (A) Progress of deuteration for the intact capsid protein (+28 charge state, 25 601 Da) in mature virus-like-particles (mVLP, red) and as disassembled monomer in solution (black). (B) Time-resolved isotopic envelopes for the peptic fragment encompassing residues 55–68 of the CA protein in immature (left) and mature (right) VLPs. The peaks marked by asterisk do not belong to the CA fragment envelopes. From [71] with permission.

icant protection with respect to the immature state (Fig. 8A) [116]. A bimodal pattern in the m/z spectra of the intact, mature CA indicated two populations of CA conformers. The first exhibited protection similar to the immature CA while the second was significantly more protected. Similar bimodal HDX pattern was observed for a peptide fragment from the interface between N- and C-terminal domain (Fig. 8B). This demonstrated that only half of the subunits in HIV-1 capsid matured into the conical cores and that the formation of the heterotypic N-C interface is a key feature of the maturation process. On the other hand, the formation of new interfaces between N-terminal domains, which had been proposed on the basis of EM to mediate maturation, was not detected [116].

Unlike HIV dsDNA and dsRNA bacteriophages (e.g. *Cystoviridae* discussed above) mature via packaging of their genomes into empty procapsids which simultaneously expand and undergo substantial stabilization [117,118]. HDX-MS application to viral maturation was pioneered using dsDNA bacteriophages P22 and Raman spectroscopy for detection [29]. This initial investigation was subsequently extended by mass spectrometry [33] and demonstrated that the observed stabilization of the viral lattice was due to refolding of coat protein domains. Further study of P22 capsids by FT-ICR HDX-MS identified the N-terminal domain as being substantially stabilized upon maturation while the C-terminal dimerization domain remained unaffected. Together with crosslinking experiments HDX results assisted mapping of the refolded domains into the EM-derived electron density [119,120].

Flexibility of virions has been demonstrated as an important factor in the delivery of genomes by plus-sense ssRNA viruses [121]. Capsid dynamics of human rhinovirus was studied by HDX-MS [122]. A fast and intermediate exchange pattern was observed for the interfaces around the five-fold axis. These sites may serve as the initiation site for uncoating and release of the viral RNA.

Native-state HDX of whole protein subunits within the intact large and small ribosomal subunit was detected by MALDI-MS [66]. While this method did not resolve subunit interfaces it provided first glimpse at the overall dynamics for almost all ribosomal proteins. The protection correlated well with the assembly pathway i.e. subunits that were incorporated early during ribosomal assembly exhibited significantly higher protection. On the other hand, faster HDX was observed for subunits implicated in tRNA translocation between sites or those involved in pivoting of large and small ribosomal subunits. These functions presumably require higher flexibility.

Aggregation, amyloid and inclusion bodies

Fluids in cells contain macromolecules at high total concentration, causing molecular crowding [123]. In most cases crowding favors protein folding as well as formation of functional complexes relative to the less compact, non-native structures. However, under such conditions even brief loss of native protein structure or exposure to extreme conditions (e.g. oxidative stress, heat shock) may lead to aggregation. Aggregation produces plethora of species, ranging from soluble oligomers, amorphous aggregates, to fibrils, amyloid plagues and inclusion bodies [124,125]. HDX-MS has played important role in characterizing structure and dynamics of protein aggregates which are often refractory to other methods [126]. HDX-MS also constitutes an indispensable tool in characterizing intrinsically unfolded proteins [127].

Prion proteins are prone to adopt different structures and some mutants readily oligomerise and aggregate to form amyloid plagues. Amyloid and inclusion body formation is associated with severe human disorders such as Alzheimer's or Huntington's diseases. The oligomers are currently thought to constitute the toxic species.

One of the proteins that are often associated with amyloid formation is α-synuclein. HDX-MS was used to delineate structural differences between two states of α-synuclein, the natively unstructured soluble monomer and the aggregated insoluble amyloid [128]. The monomer exchanged with rates corresponding to an unstructured random coil. In the amyloid state the long N-terminal and the C-terminal segments remained mostly unprotected while the central β-sheet was significantly protected. The protected β-sheet segments exceeded the length expected for an amyloid ribbon and no exposed amides corresponding to the putative interconnecting turns were observed. These results indicate that the α-synuclein amyloid adopts a structure similar to that of amyloid-β [129].

Ovine prion protein oligomerization was investigated by a combination of size-exclusion chromatography, circular dichroism and HDX-MS [130]. Three different oligomeric species were detected and structurally characterized. Surprisingly, HDX detected increased flexibility of certain regions that were shown to play essential role in aggregation. The heterogeneity and the increased dynamic character of these oligomeric precursors pointed to the existence of multiple aggregation pathways.

Overexpression and inclusion body formation may facilitate high recovery of bioactive protein provided the protein attains native structure after solubilization [131]. Over-expression and purification from inclusion bodies was employed to study the cytoplasmic region of tyrosine kinase-interacting protein (TIP) from Herpes virus saimiri. Intrinsic disorder was predicted for TIP and confirmed by HDX-MS but this did not compromise the enzyme activity and binding [132].

Protein dynamics and function

Protein function is often intimately linked to protein dynamics and involves, for example, conformational changes during enzyme activation or segment immobilization imposed by ligand binding. Conformational changes require transient population of higher energy states on the folding landscape and thus they can be detected by HDX as in the case of folding intermediates [133-135]. The importance of thermally activated protein dynamics was demonstrated for a thermophilic alcohol dehydrogenase by measuring temperature-dependent HDX rates [136]. The activation energies for HDX rates of peptides in the vicinity of the NAD^+ cofactor and the sub-

strate binding site underwent two transitions, at 30 and 45°C, respectively. These transitions correlated with changes in NAD^+ binding kinetics and enzymatic activity.

HDX-MS was used for mapping the interactions of proteins with ligands [137,138] and enabled to unravel subtle structural changes in troponin C upon Ca^{2+} binding. These changes are important for regulation but escaped detection by other techniques [139]. The concept is illustrated in Fig. 9A–D for binding of nucleotide di- and triphosphates and RNA to the hexameric viral packaging motor, protein P4 (see above). The exchange of a peptide fragment originating from the vicinity of the ATP binding site exhibited slower rate in the presence of ATP or ADP but is was largely unaffected by RNA binding (Fig. 9B–D) [114]. Further insight was obtained by computing the distribution of exchange rates by a maximum entropy method [140] (Fig. 9C) and by classification of the amide sites into three groups: fast (exchanged before the first time point, i.e. 30 sec), intermediate (exchanging on the time scale of the experiment), and protected (not exchanged during the whole duration of the experiment, i.e. 8 h) (Fig. 9D) [114]. The three-class analysis revealed that reversible ligand binding affected mostly the fast exchanging amides and provided only partial protection. This is understandable given that ATP and ADP binding and dissociation both happen on millisecond-to-second time scale.

The three-class analysis proved valuable in visualizing changes in local stability during RNA binding and ATP-driven translocation by the hexameric packaging motor P4. HDX information for each peptic fragment was color-coded and mapped onto the three-dimensional structure. For each fragment the applied color was an RGB (red-green-blue) blend which was weighted by the relative fraction of protected (blue), intermediate (green) and fast (red) amide sites (Fig. 9F). Such representation is superior to the simple average rate coloring scheme used in Fig. 7 because it contains more information and enables to distinguish cases in which the average rate is not affected, e.g. when fast and slow sites shift to intermediate class. The three-color representation revealed increased flexibility of the hexamer upon RNA binding (Fig. 9E). During translocation the protein subunits cycle between stable and flexible states and this resulted in purple color of many regions (purple = blue + red). The only exception is the interface helix (IH, green) for which the intermediate exchange is of the EX1 type (see the primary MS data in Fig. 4) and represents opening of the ring during RNA loading.

HDX-MS proved extremely useful in uncovering allosteric activation networks in protein kinases [141,142]. The extracellular regulated protein kinase-2 (ERK-2) is a MAP kinase which is activated by phosphorylation. The influence of phosphorylation on conformational flexibility was probed by HDX-MS [143]. HDX demonstrated that phosphorylation induced flexibility in the hinge region between two domains. Subsequently, conformational changes upon AMP-PNP binding to the inactive and the activated (phosphorylated) ERK-2 revealed that the DFG motif within the catalytic site was stabilized in the presence of AMP-PNP but only in the active form [144]. Hence phosphorylation facilitates interdomain closure that is necessary for the precise alignment between ATP and the polypeptide substrate.

In another example HDX-MS was used to probe allosteric activation of coagulation factor VIIa (FVIIa) by a tissue factor (TF), both of which play essential role in blood clotting [145]. Observed HDX kinetics demonstrated stabilization of the activation domain and the 170-loop in FVIIa upon TF binding. Interestingly, the two protected regions are distal to the TF recognition helix within FVIIa sequence. Comparison of HDX results with molecular dynamics simulations identified a key interaction between Leu305 and Phe374 which is likely to transmit the stabilizing effect from the recognition site to the activation domain. This is an example in which region specific HDX was augmented by molecular dynamics simulations to reach conclusions at atomic level.

Conclusion

It is clear from the above illustrative but largely incomplete survey of HDX-MS applications that this method has gained popularity within the community. As suitable mass spectrometers become more affordable they will be acquired by individual labs for dedicated HDX-MS use. This will generate vast amounts of data and hopefully stimulate further software developments. However, at the moment there is no public depository for such a vast amount of potentially useful data. The results are scattered throughout literature in various formats which are generally not amenable to quantitative comparisons and searches. Therefore, there is an urgent need to develop a uniform way of how to present and deposit HDX-MS results and time is ripe to create a fully searchable public database.

In addition, more advanced instrumentation will enable developments of new methods and improve sequence coverage and site-specific resolution by e.g. tandem MS. Another avenue is to improve the overall throughput by decreasing the time necessary for data collection by e.g. lab-on-chip implementation of fully automated protocols. This in turn would significantly enhance temporal resolution of HDX and enable probing faster, functional dynamics which is often associated with enzyme action.

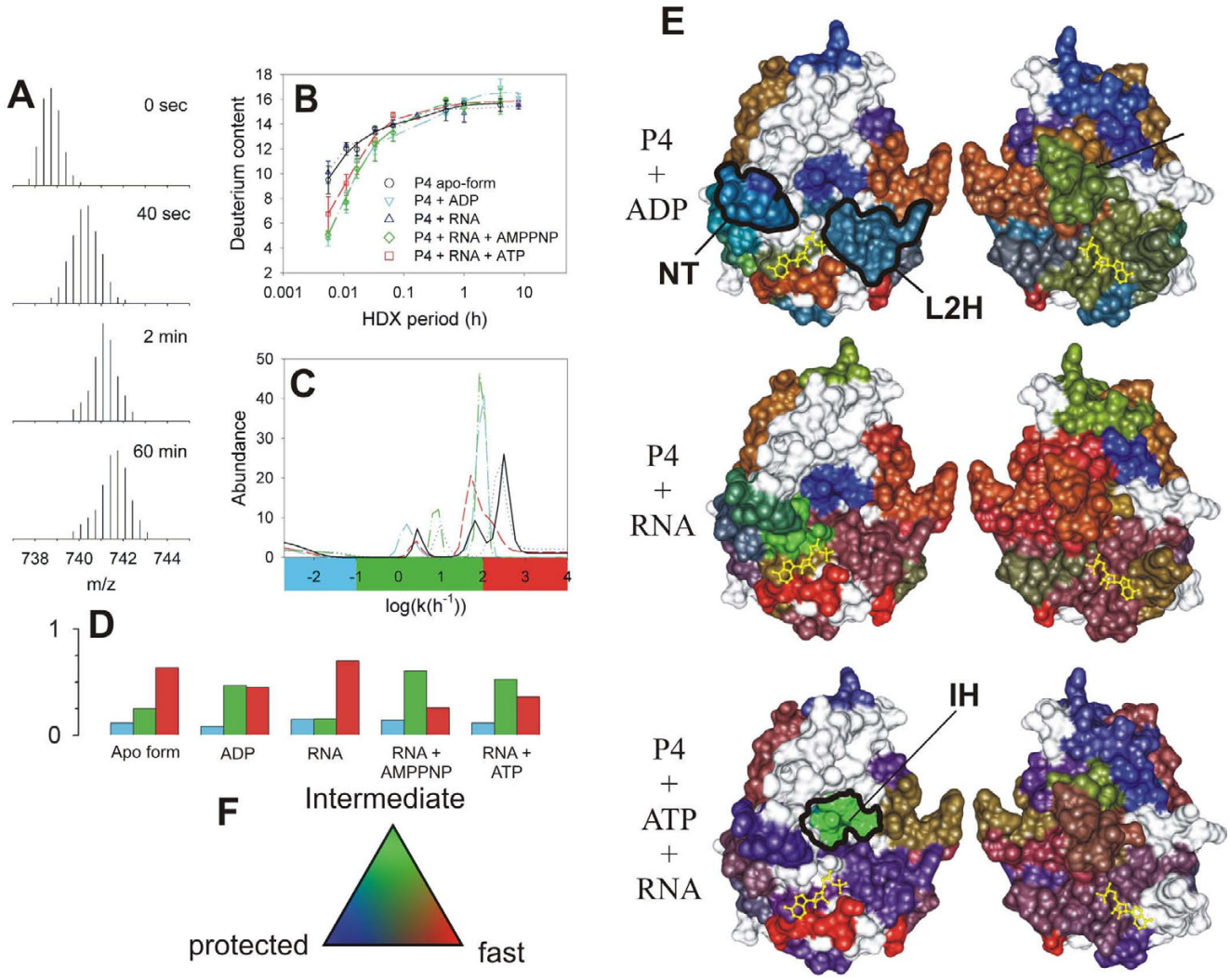


Figure 9

Mapping functional dynamics in the hexamer of P4 from bacteriophage φ8 [114]. Mass/charge (*m*/*z*) spectra corresponding to the peptic fragment encompassing residues 139–158 (*m* = 2210.14 Da, z = 3) during H/D exchange (only interval 0 to 60 min shown). (B) Increase of deuterium content in the peptide (symbols) and the corresponding maximum entropy fit (MEM) for P4 alone (black circles, solid line), P4+1 mM poly(C) (blue triangles, dotted), P4 + 1 mM poly(C) + 1 mM ATP (red squares, dashed line), P4 + 1 mM poly(C) + 1 mM AMP-PNP (green diamonds, dash-dot-dot) and P4 + 1 mM ADP (cyan inverted triangles, dash-dot). Standard deviations (error bars) were estimated from three independent experiments. (C) Rate distributions obtained by MEM fitting of data in panel B. The color bar under the panel indicates the three integration regions which were used to obtain the number of sites within each rate class (blue = slow/protected, green = intermediate, red = fast/unprotected). (D) Number of amide sites in the three classes and under different conditions (nucleotide di/triphosphates, RNA binding) obtained from data in panel C, bar colors as in panel C. (E) RGB representation of the HDX kinetics for subunit interfaces. The two facets (left and right) represent the facing interfaces from the neighboring subunits in a surface representation. Bound ATP molecule is shown in yellow ball-and-stick representation. Several regions of interest are delineated: NT-nucleotide binding pocket; L2H-loop 2 and α-helix 6 which constitute the moving lever of the motor; IH-interfacial helix which becomes transiently exposed during ring opening and RNA loading. (F) Three-color, RGB scale for number of amides exchanging in the three classes.

These later developments will make HDX-MS valuable companion to structural genomics.

Authors' contributions

BS selected the reviewed papers and drafted the manuscript. RT wrote the introduction and edited the final version. All authors read and approved the final version.

Acknowledgements

This work was supported by EUROCORES Programmes EuroSCOPE of the European Science Foundation, which is also funded by the European Commission, Contract no. ERAS-CT-2003-980409 of the Sixth Framework Programme. We would like to thank Dr. Jiri Lisal for assistance with preparation of the illustrations.

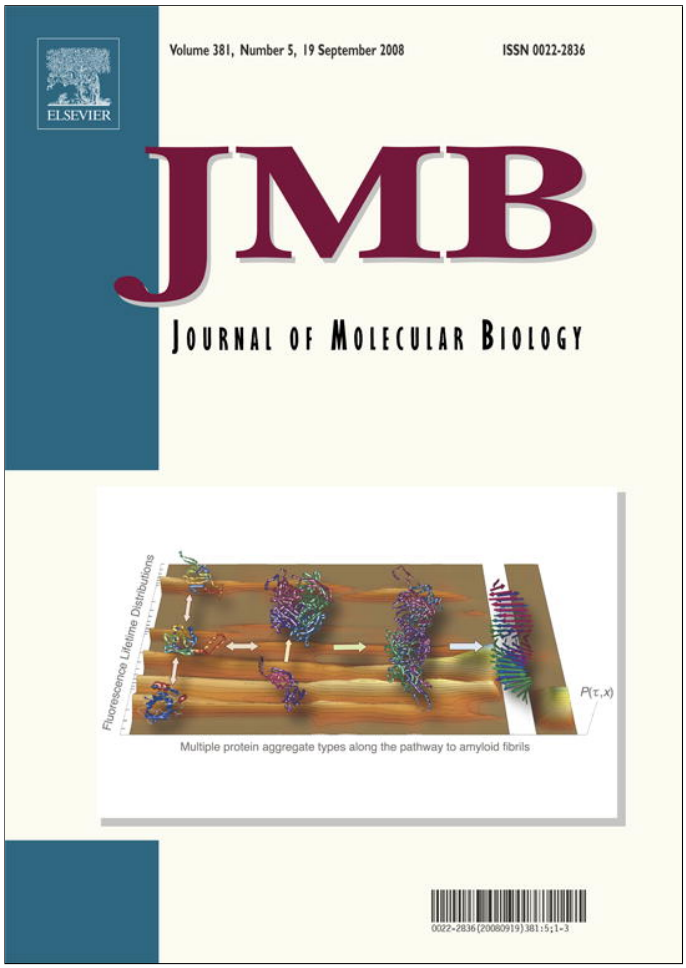
References

1. Xu S: **Aggregation drives "misfolding" in protein amyloid fiber formation.** *Amyloid* 2007, **14:**119-131.
2. Perczel A, Hudaky P, Palfi VK: **Dead-end street of protein folding: thermodynamic rationale of amyloid fibril formation.** *J Am Chem Soc* 2007, **129:**14959-14965.
3. Sosnick TR, Mayne L, Englander SW: **Molecular collapse: the rate-limiting step in two-state cytochrome c folding.** *Proteins* 1996, **24:**413-426.
4. Sosnick TR, Mayne L, Hiller R, Englander SW: **The barriers in protein folding.** *Nat Struct Biol* 1994, **1:**149-156.
5. Kubelka J, Hofrichter J, Eaton WA: **The protein folding 'speed limit'.** *Curr Opin Struct Biol* 2004, **14:**76-88.
6. Eaton WA, Munoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J: **Fast kinetics and mechanisms in protein folding.** *Annu Rev Biophys Biomol Struct* 2000, **29:**327-359.
7. Li R, Woodward C: **The hydrogen exchange core and protein folding.** *Protein Sci* 1999, **8:**1571-1590.
8. Woodward C: **Is the slow exchange core the protein folding core?** *Trends Biochem Sci* 1993, **18:**359-360.
9. Woodward C, Carulla N, Barany G: **Native state hydrogen-exchange analysis of protein folding and protein motional domains.** *Methods Enzymol* 2004, **380:**379-400.
10. Weinkam P, Zong C, Wolynes PG: **A funneled energy landscape for cytochrome c directly predicts the sequential folding route inferred from hydrogen exchange experiments.** *Proc Natl Acad Sci U S A* 2005, **102:**12401-12406.
11. Wolynes PG: **Recent successes of the energy landscape theory of protein folding and function.** *Q Rev Biophys* 2005, **38:**405-410.
12. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG: **Funnels, pathways, and the energy landscape of protein folding: a synthesis.** *Proteins* 1995, **21:**167-195.
13. Onuchic JN, Wolynes PG: **Theory of protein folding.** *Curr Opin Struct Biol* 2004, **14:**70-75.
14. Hardin C, Eastwood MP, Prentiss M, Luthey-Schulten Z, Wolynes PG: **Folding funnels: the key to robust protein structure prediction.** *J Comput Chem* 2002, **23:**138-146.
15. Baker D: **A surprising simplicity to protein folding.** *Nature* 2000, **405:**39-42.
16. Dobson CM, Karplus M: **The fundamentals of protein folding: bringing together theory and experiment.** *Current Opinion in Structural Biology* 1999, **9:**92-101.
17. Speed MA, Wang DI, King J: **Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition.** *Nat Biotechnol* 1996, **14:**1283-1287.
18. Yang M, Yordanov B, Levy Y, Bruschweiler R, Huo S: **The sequence-dependent unfolding pathway plays a critical role in the amyloidogenicity of transthyretin.** *Biochemistry* 2006, **45:**11992-12002.
19. Teplow DB, Lazo ND, Bitan G, Bernstein S, Wyttenbach T, Bowers MT, Baumketner A, Shea JE, Urbanc B, Cruz L, Borreguero J, Stanley HE: **Elucidating amyloid beta-protein folding and assembly: A multidisciplinary approach.** *Acc Chem Res* 2006, **39:**635-645.
20. Xia Y, Levitt M: **Simulating protein evolution in sequence and structure space.** *Curr Opin Struct Biol* 2004, **14:**202-207.
21. Taylor DJ, Johnson JE: **Folding and particle assembly are disrupted by single-point mutations near the autocatalytic cleavage site of Nudaurelia capensis {omega} virus capsid protein.** *Protein Sci* 2005, **14:**401-408.
22. Volgina VV, Sun T, Bozek G, Martin TE, Storb U: **Scarcity of [lambda]1 B cells in mice with a single point mutation in C[lambda]1 is due to a low BCR signal caused by misfolded lambda1 light chain.** *Molecular Immunology* 2007, **44:**1417-1428.
23. Mogensen JE, Ipsen H, Holm J, Otzen DE: **Elimination of a Misfolded Folding Intermediate by a Single Point Mutation.** *Biochemistry* 2004, **43:**3357-3367.
24. He HW, Feng S, Pang M, Zhou HM, Yan YB: **Role of the linker between the N- and C-terminal domains in the stability and folding of rabbit muscle creatine kinase.** *The International Journal of Biochemistry & Cell Biology* 2007, **39:**1816-1827.
25. Fan J, Perry SJ, Gao Y, Schwarz DA, Maki RA: **A Point Mutation in the Human Melanin Concentrating Hormone Receptor 1 Reveals an Important Domain for Cellular Trafficking.** *Mol Endocrinol* 2005, **19:**2579-2590.
26. Capriotti E, Compiani M: **Diffusion-collision of foldons elucidates the kinetic effects of point mutations and suggests control strategies of the folding process of helical proteins.** *Proteins: Structure, Function, and Bioinformatics* 2006, **64(1):**198-209.
27. Borgia A, Bonivento D, Travaglini-Allocatelli C, Di Matteo A, Brunori M: **Unveiling a Hidden Folding Intermediate in c-Type Cytochromes by Protein Engineering.** *J Biol Chem* 2006, **281:**9331-9336.
28. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN: **Flexible nets. The roles of intrinsic disorder in protein interaction networks.** *Febs J* 2005, **272:**5129-5148.
29. Tuma R, Prevelige PE Jr., Thomas GJ Jr.: **Mechanism of capsid maturation in a double-stranded DNA virus.** *Proc Natl Acad Sci U S A* 1998, **95:**9885-9890.
30. Hurley TD, Yang J, Zhang L, Goodwin KD, Zou Q, Cortese M, Dunker AK, DePaoli-Roach AA: **Structural basis for regulation of protein phosphatase 1 by inhibitor-2.** *J Biol Chem* 2007, **282:**28874-28883.
31. Tuma R, Tsuruta H, Benevides JM, Prevelige PE Jr., Thomas GJ Jr.: **Characterization of subunit structural changes accompanying assembly of the bacteriophage P22 procapsid.** *Biochemistry* 2001, **40:**665-674.
32. Tuma R, Bamford JH, Bamford DH, Russell MP, Thomas GJ Jr.: **Structure, interactions and dynamics of PRD1 virus I. Coupling of subunit folding and capsid assembly.** *J Mol Biol* 1996, **257:**87-101.
33. Tuma R, Coward LU, Kirk MC, Barnes S, Prevelige Jr PE: **Hydrogen-deuterium Exchange as a Probe of Folding and Assembly in Viral Capsids,.** *Journal of Molecular Biology* 2001, **306:**389-396.
34. Lefebvre BG, Robinson AS: **Pressure treatment of tailspike aggregates rapidly produces on-pathway folding intermediates.** *Biotechnol Bioeng* 2003, **82:**595-604.
35. Carrio MM, Villaverde A: **Protein aggregation as bacterial inclusion bodies is reversible.** *FEBS Lett* 2001, **489:**29-33.
36. Bai Y, Sosnick TR, Mayne L, Englander SW: **Protein folding intermediates: native-state hydrogen exchange.** *Science* 1995, **269:**192-197.
37. Ghaemmaghami S, Oas TG: **Quantitative protein stability measurement in vivo.** *Nat Struct Biol* 2001, **8:**879-882.
38. Ghaemmaghami S, Fitzgerald MC, Oas TG: **A quantitative, high-throughput screen for protein stability.** *Proc Natl Acad Sci U S A* 2000, **97:**8296-8301.
39. Fuller SD, Butcher SJ, Cheng RH, Baker TS: **Three-dimensional reconstruction of icosahedral particles--the uncommon line.** *J Struct Biol* 1996, **116:**48-55.
40. Martin CS, Burnett RM, de Haas F, Heinkel R, Rutten T, Fuller SD, Butcher SJ, Bamford DH: **Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions.** *Structure* 2001, **9:**917-930.
41. Martin AG, Depoix F, Stohr M, Meissner U, Hagner-Holler S, Hammouti K, Burmester T, Heyd J, Wriggers W, Markl J: **Limulus polyphemus hemocyanin: 10 A cryo-EM structure, sequence analysis, molecular modelling and rigid-body fitting reveal the interfaces between the eight hexamers.** *J Mol Biol* 2007, **366:**1332-1350.

42. Wikoff WR, Conway JF, Tang J, Lee KK, Gan L, Cheng N, Duda RL, Hendrix RW, Steven AC, Johnson JE: **Time-resolved molecular dynamics of bacteriophage HK97 capsid maturation interpreted by electron cryo-microscopy and X-ray crystallography.** *J Struct Biol* 2006, **153:**300-306.
43. Fu J, Gao H, Frank J: **Unsupervised classification of single particles by cluster tracking in multi-dimensional space.** *J Struct Biol* 2007, **157:**226-239.
44. Poliakov A, van Duijn E, Lander G, Fu CY, Johnson JE, Prevelige PE Jr., Heck AJ: **Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies.** *J Struct Biol* 2007, **157:**371-383.
45. Englander SW, Kallenbach NR: **Hydrogen exchange and structural dynamics of proteins and nucleic acids.** *Q Rev Biophys* 1983, **16:**521-655.
46. Englander SW: **Hydrogen Exchange and Mass Spectrometry: A Historical Perspective.** *Journal of the American Society for Mass Spectrometry* 2006, **17:**1481-1489.
47. Bai Y, Englander SW: **Hydrogen bond strength and beta-sheet propensities: the role of a side chain blocking effect.** *Proteins* 1994, **18:**262-266.
48. Bai Y, Milne JS, Mayne L, Englander SW: **Primary structure effects on peptide group hydrogen exchange.** *Proteins* 1993, **17:**75-86.
49. Bai Y, Milne JS, Mayne L, Englander SW: **Protein stability parameters measured by hydrogen exchange.** *Proteins* 1994, **20:**4-14.
50. Wintrode PL, Friedrich KL, Vierling E, Smith JB, Smith DL: **Solution structure and dynamics of a heat shock protein assembly probed by hydrogen exchange and mass spectrometry.** *Biochemistry* 2003, **42:**10667-10673.
51. Zhang Z, Post CB, Smith DL: **Amide hydrogen exchange determined by mass spectrometry: application to rabbit muscle aldolase.** *Biochemistry* 1996, **35:**779-791.
52. Hoang L, Bedard S, Krishna MM, Lin Y, Englander SW: **Cytochrome c folding pathway: kinetic native-state hydrogen exchange.** *Proc Natl Acad Sci U S A* 2002, **99:**12173-12178.
53. Hamuro Y, Coales SJ, Southern MR, Nemeth-Cawley JF, Stranz DD, Griffin PR: **Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry.** *J Biomol Tech* 2003, **14:**171-182.
54. Hoofnagle AN, Resing KA, Ahn NG: **Practical methods for deuterium exchange/mass spectrometry.** *Methods Mol Biol* 2004, **250:**283-298.
55. Heimburg T, Marsh D: **Investigation of secondary and tertiary structural changes of cytochrome c in complexes with anionic lipids using amide hydrogen exchange measurements: an FTIR study.** *Biophys J* 1993, **65:**2408-2417.
56. Goormaghtigh E, Cabiaux V, Ruysschaert JM: **Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. II. Experimental aspects, side chain structure, and H/D exchange.** *Subcell Biochem* 1994, **23:**363-403.
57. Vigneron L, Ruysschaert JM, Goormaghtigh E: **Fourier transform infrared spectroscopy study of the secondary structure of the reconstituted Neurospora crassa plasma membrane H(+)-ATPase and of its membrane-associated proteolytic peptides.** *J Biol Chem* 1995, **270:**17685-17696.
58. Baenziger JE, Methot N: **Fourier transform infrared and hydrogen/deuterium exchange reveal an exchange-resistant core of alpha-helical peptide hydrogens in the nicotinic acetylcholine receptor.** *J Biol Chem* 1995, **270:**29129-29137.
59. Vigano C, Manciu L, Buyse F, Goormaghtigh E, Ruysschaert JM: **Attenuated total reflection IR spectroscopy as a tool to investigate the structure, orientation and tertiary structure changes in peptides and membrane proteins.** *Biopolymers* 2000, **55:**373-380.
60. Li T, Talvenheimo J, Zeni L, Rosenfeld R, Stearns G, Arakawa T: **Changes in protein conformation and dynamics upon complex formation of brain-derived neurotrophic factor and its receptor: investigation by isotope-edited Fourier transform IR spectroscopy.** *Biopolymers* 2002, **67:**10-19.
61. Reilly KE, Thomas GJ Jr.: **Hydrogen exchange dynamics of the P22 virion determined by time-resolved Raman spectroscopy. Effects of chromosome packaging on the kinetics of nucleotide exchanges.** *J Mol Biol* 1994, **241:**68-82.
62. Xu M, Shashilov V, Lednev IK: **Probing the cross-beta core structure of amyloid fibrils by hydrogen-deuterium exchange deep ultraviolet resonance Raman spectroscopy.** *J Am Chem Soc* 2007, **129:**11002-11003.
63. Zhang Z, Smith DL: **Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation.** *Protein Sci* 1993, **2:**522-531.
64. Hoofnagle AN, Resing KA, Ahn NG: **Protein analysis by hydrogen exchange mass spectrometry.** *Annual Review of Biophysics and Biomolecular Structure* 2003, **32:**1-25.
65. Wang L, Lane LC, Smith DL: **Detecting structural changes in viral capsids by hydrogen exchange and mass spectrometry.** *Protein Sci* 2001, **10:**1234-1243.
66. Yamamoto T, Izumi S, Gekko K: **Mass spectrometry of hydrogen/deuterium exchange in 70S ribosomal proteins from E. coli.** *FEBS Letters* 2006, **580:**3638-3642.
67. Anand GS, Law D, Mandell JG, Snead AN, Tsigelny I, Taylor SS, Ten Eyck LF, Komives EA: **Identification of the protein kinase A regulatory RIalpha-catalytic subunit interface by amide H/2H exchange and protein docking.** *Proc Natl Acad Sci U S A* 2003, **100:**13264-13269.
68. Lee T, Hoofnagle AN, Kabuyama Y, Stroud J, Min X, Goldsmith EJ, Chen L, Resing KA, Ahn NG: **Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry.** *Mol Cell* 2004, **14:**43-55.
69. Baerga-Ortiz A, Hughes CA, Mandell JG, Komives EA: **Epitope mapping of a monoclonal antibody against human thrombin by H/D-exchange mass spectrometry reveals selection of a diverse sequence in a highly conserved protein.** *Protein Sci* 2002, **11:**1300-1308.
70. Lanman J, Lam TT, Barnes S, Sakalian M, Emmett MR, Marshall AG, Prevelige PE Jr.: **Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry.** *J Mol Biol* 2003, **325:**759-772.
71. Lanman J, Prevelige PE Jr.: **High-sensitivity mass spectrometry for imaging subunit interactions: hydrogen/deuterium exchange.** *Curr Opin Struct Biol* 2004, **14:**181-188.
72. Mandell JG, Baerga-Ortiz A, Falick AM, Komives EA: **Measurement of solvent accessibility at protein-protein interfaces.** *Methods Mol Biol* 2005, **305:**65-80.
73. Kweon HK, Hakansson K: **Site-specific amide hydrogen exchange in melittin probed by electron capture dissociation Fourier transform ion cyclotron resonance mass spectrometry.** *Analyst* 2006, **131:**275-280.
74. Nazabal A, Bonneu M, Saupe SJ, Schmitter JM: **High-resolution H/D exchange studies on the HET-s218-295 prion protein.** *J Mass Spectrom* 2005, **40:**580-590.
75. Cravello L, Lascoux D, Forest E: **Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins.** *Rapid Commun Mass Spectrom* 2003, **17:**2387-2393.
76. Lam TKT, Lanman JK, Emmett MR, Hendrickson CL, Marshall AG, Prevelige PE: **Mapping of protein: protein contact surfaces by hydrogen/deuterium exchange, followed by on-line high-performance liquid chromatography-electrospray ionization fourier-transform ion-cyclotron-resonance mass analysis.** *Journal of Chromatography A* 2002, **982:**85-95.
77. Dass C: **Principles and Practice of Biological Mass Spectrometry.** New York, John Wiley & Sons; 2001.
78. Andren PE, Emmett MR, DaGue BB, Steulet AF, Waldmeier P, Caprioli RM: **Blood-brain barrier penetration of 3-aminopropyl-n-butylphosphinic acid (CGP 36742) in rat brain by microdialysis/mass spectrometry.** *J Mass Spectrom* 1998, **33:**281-287.
79. Chalmers MJ, Busby SA, Pascal BD, He Y, Hendrickson CL, Marshall AG, Griffin PR: **Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry.** *Anal Chem* 2006, **78:**1005-1014.
80. Deng Y, Smith DL: **Identification of Unfolding Domains in Large Proteins by Their Unfolding Rates.** *Biochemistry* 1998, **37:**6256-6262.
81. Wu Y, Engen JR, Hobbins WB: **Ultra performance liquid chromatography (UPLC) further improves hydrogen/deuterium exchange mass spectrometry.** *J Am Soc Mass Spectrom* 2006, **17:**163-167.
82. Emmett MR, Kazazic S, Marshall AG, Chen W, Shi SDH, Bolanos B, Greig MJ: **Supercritical Fluid Chromatography Reduction of**

Hydrogen/Deuterium Back Exchange in Solution-Phase Hydrogen/Deuterium Exchange with Mass Spectrometric Analysis. *Anal Chem* 2006, **78:**7058-7060.
83. Engen JR, Smith DL: **Investigating protein structure and dynamics by hydrogen exchange MS.** *Anal Chem* 2001, **73:**256A-265A.
84. Marshall AG, Hendrickson CL, Emmett MR, Rodgers RP, Blakney GT, Nilsson CL: **Fourier transform ion cyclotron resonance: state of the art.** *Eur J Mass Spectrom (Chichester, Eng)* 2007, **13:**57-59.
85. Mayne L, Englander SW: **Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange.** *Protein Sci* 2000, **9:**1873-1877.
86. Simmons DA, Dunn SD, Konermann L: **Conformational Dynamics of Partially Denatured Myoglobin Studied by Time-Resolved Electrospray Mass Spectrometry with Online Hydrogen-Deuterium Exchange.** *Biochemistry* 2003, **42:**5896-5905.
87. Zhang Z, Guan S, Marshall AG: **Enhancement of the effective resolution of mass spectra of high-mass biomolecules by maximum entropy-based deconvolution to eliminate the isotopic natural abundance distribution.** *Journal of the American Society for Mass Spectrometry* 1997, **8:**659-670.
88. Hotchko M, Anand GS, Komives EA, Ten Eyck LF: **Automated extraction of backbone deuteration levels from amide H/2H mass spectrometry experiments.** *Protein Sci* 2006, **15:**583-601.
89. Yang L, Song G, Jernigan RL: **How well can we understand large-scale protein motions using normal modes of elastic network models?** *Biophys J* 2007, **93:**920-929.
90. Bahar I, Wallqvist A, Covell DG, Jernigan RL: **Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model.** *Biochemistry* 1998, **37:**1067-1075.
91. Li G, Van Wynsberghe A, Demerdash O N., A., Cui Q: **Normal mode analysis of macromolecules: From enzyme active sites to molecular machines.** In *Normal Mode Analysis Theory and Applications to Biological and Chemical Systems* Edited by: Cui Q and Bahar I. Boca Raton, Chapman & Hall/CRC; 2006:65-90.
92. Ferguson PL, Pan J, Wilson DJ, Dempsey B, Lajoie G, Shilton B, Konermann L: **Hydrogen/Deuterium Scrambling during Quadrupole Time-of-Flight MS/MS Analysis of a Zinc-Binding Protein Domain.** *Anal Chem* 2007, **79:**153-160.
93. Sun S, Meyer-Arendt K, Eichelberger B, Brown R, Yen CY, Old WM, Pierce K, Cios KJ, Ahn NG, Resing KA: **Improved validation of peptide MS/MS assignments using spectral intensity prediction.** *Mol Cell Proteomics* 2007, **6:**1-17.
94. Lisal J, Kainov DE, Lam TT, Emmett MR, Wei H, Gottlieb P, Marshall AG, Tuma R: **Interaction of packaging motor with the polymerase complex of dsRNA bacteriophage.** *Virology* 2006, **351:**73-79.
95. Weis DD, Engen JR, Kass IJ: **Semi-automated data processing of hydrogen exchange mass spectra using HX-Express.** *J Am Soc Mass Spectrom* 2006, **17:**1700-1703.
96. Pascal BD, Chalmers MJ, Busby SA, Mader CC, Southern MR, Tsinoremas NF, Griffin PR: **The Deuterator: software for the determination of backbone amide deuterium levels from H/D exchange MS data.** *BMC Bioinformatics* 2007, **8:**156.
97. Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti RH, Apweiler R, Cheung K, Costello CE, Hermjakob H, Huang S, Julian RK, Kapp E, McComb ME, Oliver SG, Omenn G, Paton NW, Simpson R, Smith R, Taylor CF, Zhu W, Aebersold R: **A common open representation of mass spectrometry data and its application to proteomics research.** *Nat Biotechnol* 2004, **22:**1459-1466.
98. **Deuterator** [http://deuterator.florida.scripps.edu/]
99. Krishna MM, Hoang L, Lin Y, Englander SW: **Hydrogen exchange methods to study protein folding.** *Methods* 2004, **34:**51-64.
100. Hoerner JK, Xiao H, Kaltashov IA: **Structural and Dynamic Characteristics of a Partially Folded State of Ubiquitin Revealed by Hydrogen Exchange Mass Spectrometry.** *Biochemistry* 2005, **44:**11286-11294.
101. Konermann L, Simmons DA: **Protein-folding kinetics and mechanisms studied by pulse-labeling and mass spectrometry.** *Mass Spectrometry Reviews* 2003, **22:**1-26.
102. Simmons DA, Konermann L: **Characterization of Transient Protein Folding Intermediates during Myoglobin Reconstitution by Time-Resolved Electrospray Mass Spectrometry with On-Line Isotopic Pulse Labeling.** *Biochemistry* 2002, **41:**1906-1914.
103. Pan J, Wilson DJ, Konermann L: **Pulsed Hydrogen Exchange and Electrospray Charge-State Distribution as Complementary Probes of Protein Structure in Kinetic Experiments: Implications for Ubiquitin Folding.** *Biochemistry* 2005, **44:**8627-8633.
104. Pan J, Rintala-Dempsey AC, Li Y, Shaw GS, Konermann L: **Folding Kinetics of the S100A11 Protein Dimer Studied by Time-Resolved Electrospray Mass Spectrometry and Pulsed Hydrogen-Deuterium Exchange.** *Biochemistry* 2006, **45:**3005-3013.
105. Deng Y, Smith DL: **Hydrogen exchange demonstrates three domains in aldolase unfold sequentially.** *Journal of Molecular Biology* 1999, **294:**247-258.
106. Hamuro Y, Burns LL, Canaves JM, Hoffman RC, Taylor SS, Woods VL: **Domain Organization of -AKAP2 Revealed by Enhanced Deuterium Exchange-Mass Spectrometry (DXMS).** *Journal of Molecular Biology* 2002, **321:**703-714.
107. Jungbauer LM, Cavagnero S: **Characterization of protein expression and folding in cell-free systems by maldi-tof mass spectrometry.** *Anal Chem* 2006, **78:**2841-2852.
108. Spraggon G, Pantazatos D, Klock HE, Wilson IA, Woods VL Jr., Lesley SA: **On the use of DXMS to produce more crystallizable proteins: structures of the T. maritima proteins TM0160 and TM1171.** *Protein Sci* 2004, **13:**3187-3199.
109. Pantazatos D, Kim JS, Klock HE, Stevens RC, Wilson IA, Lesley SA, Woods VL Jr.: **Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS.** *Proc Natl Acad Sci U S A* 2004, **101:**751-756.
110. Poranen MM, Tuma R: **Self-assembly of double-stranded RNA bacteriophages.** *Virus Res* 2004, **101:**93-100.
111. Mancini EJ, Kainov DE, Grimes JM, Tuma R, Bamford DH, Stuart DI: **Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation.** *Cell* 2004, **118:**743-755.
112. de Haas F, Paatero AO, Mindich L, Bamford DH, Fuller SD: **A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage phi6.** *J Mol Biol* 1999, **294:**357-372.
113. Huiskonen JT, Jaalinoja HT, Briggs JA, Fuller SD, Butcher SJ: **Structure of a hexameric RNA packaging motor in a viral polymerase complex.** *J Struct Biol* 2007, **158:**156-164.
114. Lisal J, Lam TT, Kainov DE, Emmett MR, Marshall AG, Tuma R: **Functional visualization of viral molecular motor by hydrogen-deuterium exchange reveals transient states.** *Nat Struct Mol Biol* 2005, **12:**460-466.
115. Lanman J, Prevelige PE Jr.: **Kinetic and mass spectrometry-based investigation of human immunodeficiency virus type 1 assembly and maturation.** *Adv Virus Res* 2005, **64:**285-309.
116. Lanman J, Lam TT, Emmett MR, Marshall AG, Sakalian M, Prevelige PE Jr.: **Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange.** *Nat Struct Mol Biol* 2004, **11:**676-677.
117. Jiang W, Li Z, Zhang Z, Baker ML, Prevelige PE Jr, Chiu W: **Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions.** *Nat Struct Biol* 2003, **10:**131-135.
118. Galisteo ML, King J: **Conformational transformations in the protein lattice of phage P22 procapsids.** *Biophys J* 1993, **65:**227-235.
119. Kang S, Hawkridge AM, Johnson KL, Muddiman DC, Prevelige PE Jr.: **Identification of subunit-subunit interactions in bacteriophage P22 procapsids by chemical cross-linking and mass spectrometry.** *J Proteome Res* 2006, **5:**370-377.
120. Kang S, Prevelige PE Jr.: **Domain study of bacteriophage p22 coat protein and characterization of the capsid lattice transformation by hydrogen/deuterium exchange.** *J Mol Biol* 2005, **347:**935-948.
121. Johnson JE: **Virus particle dynamics.** *Adv Protein Chem* 2003, **64:**197-218.
122. Wang L, Smith DL: **Capsid structure and dynamics of a human rhinovirus probed by hydrogen exchange mass spectrometry.** *Protein Sci* 2005, **14:**1661-1672.
123. Minton AP: **Implications of macromolecular crowding for protein assembly.** *Current Opinion in Structural Biology* 2000, **10:**34-39.

124. Kodali R, Wetzel R: **Polymorphism in the intermediates and products of amyloid assembly.** *Current Opinion in Structural Biology* 2007, **17:**48-57.
125. Idicula-Thomas S, V. BP: **Protein aggregation: A perspective from amyloid and inclusion-body formation.** *Current Science* 2007, **92:**758-767.
126. Kheterpal I, Wetzel R: **Hydrogen/Deuterium Exchange Mass Spectrometry-A Window into Amyloid Structure.** *Acc Chem Res* 2006, **39:**584-593.
127. Uversky VN, Fink AL: **Conformational constraints for amyloid fibrillation: the importance of being unfolded.** *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* 2004, **1698(2):**131-153.
128. Del Mar C, Greenbaum EA, Mayne L, Englander SW, Woods VL Jr.: **Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level.** *Proc Natl Acad Sci U S A* 2005, **102:**15477-15482.
129. Del Mar C, Greenbaum EA, Mayne L, Englander SW, Woods VL Jr.: **Structure and properties of {alpha}-synuclein and other amyloids determined at the amino acid level.** *PNAS* 2005, **102:**15477-15482.
130. Eghiaian F, Daubenfeld T, Quenet Y, van Audenhaege M, Bouin AP, van der Rest G, Grosclaude J, Rezaei H: **Diversity in prion protein oligomerization pathways results from domain expansion as revealed by hydrogen/deuterium exchange and disulfide linkage.** *Proc Natl Acad Sci U S A* 2007, **104:**7414-7419.
131. Singh SM, Panda AK: **Solubilization and refolding of bacterial inclusion body proteins.** *Journal of Bioscience and Bioengineering* 2005, **99(4):**303-310.
132. Mitchell JL, Trible RP, Emert-Sedlak LA, Weis DD, Lerner EC, Applen JJ, Sefton BM, Smithgall TE, Engen JR: **Functional Characterization and Conformational Analysis of the Herpesvirus saimiri Tip-C484 Protein.** *Journal of Molecular Biology* 2007, **366:**1282-1293.
133. Ptitsyn OB, Bychkova VE, Uversky VN: **Kinetic and equilibrium folding intermediates.** *Philos Trans R Soc Lond B Biol Sci* 1995, **348:**35-41.
134. Eyles SJ, Kaltashov IA: **Methods to study protein dynamics and folding by mass spectrometry.** *Methods* 2004, **34:**88-99.
135. Wales TE, Engen JR: **Hydrogen exchange mass spectrometry for the analysis of protein dynamics.** *Mass Spectrometry Reviews* 2006, **25:**158-170.
136. Liang ZX, Lee T, Resing KA, Ahn NG, Klinman JP: **Thermal-activated protein mobility and its correlation with catalysis in thermophilic alcohol dehydrogenase.** *Proc Natl Acad Sci U S A* 2004, **101:**9556-9561.
137. Wang F, Blanchard JS, Tang X: **Hydrogen Exchange/Electrospray Ionization Mass Spectrometry Studies of Substrate and Inhibitor Binding and Conformational Changes of Escherichia coli Dihydrodipicolinate Reductase.** *Biochemistry* 1997, **36:**3755-3759.
138. Wang F, Scapin G, Blanchard JS, Angeletti RH: **Substrate binding and conformational changes of Clostridium glutamicum diaminopimelate dehydrogenase revealed by hydrogen/deuterium exchange and electrospray mass spectrometry.** *Protein Sci* 1998, **7:**293-299.
139. Wang F, Li W, Emmett MR, Marshall AG, Corson D, Sykes BD: **Fourier transform ion cyclotron resonance mass spectrometric detection of small Ca2+-induced conformational changes in the regulatory domain of human cardiac troponin C.** *Journal of the American Society for Mass Spectrometry* 1999, **10:**703-710.
140. Zhang Z, Li W, Logan TM, Li M, Marshall AG: **Human recombinant [C22A] FK506-binding protein amide hydrogen exchange rates from mass spectrometry match and extend those from NMR.** *Protein Sci* 1997, **6:**2203-2217.
141. Shi Z, Resing KA, Ahn NG: **Networks for the allosteric control of protein kinases.** *Curr Opin Struct Biol* 2006, **16:**686-692.
142. Emrick MA, Lee T, Starkey PJ, Mumby MC, Resing KA, Ahn NG: **The gatekeeper residue controls autoactivation of ERK2 via a pathway of intramolecular connectivity.** *Proc Natl Acad Sci U S A* 2006, **103:**18101-18106.
143. Hoofnagle AN, Resing KA, Goldsmith EJ, Ahn NG: **Changes in protein conformational mobility upon activation of extracellular regulated protein kinase-2 as detected by hydrogen exchange.** *Proc Natl Acad Sci U S A* 2001, **98:**956-961.
144. Lee T, Hoofnagle AN, Resing KA, Ahn NG: **Hydrogen exchange solvent protection by an ATP analogue reveals conformational changes in ERK2 upon activation.** *J Mol Biol* 2005, **353:**600-612.
145. Olsen OH, Rand KD, Ostergaard H, Persson E: **A combined structural dynamics approach identifies a putative switch in factor VIIa employed by tissue factor to initiate blood coagulation.** *Protein Sci* 2007, **16:**671-682.



doi:10.1016/j.jmb.2008.06.020 J. Mol. Biol. (2008) 381, 1395–1406

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Detection of Intermediates and Kinetic Control during Assembly of Bacteriophage P22 Procapsid

Roman Tuma[1,2,3]*, Hiro Tsuruta[4], Kenneth H. French[1] and Peter E. Prevelige[1]

[1]*Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA*

[2]*Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK*

[3]*Institute of Biotechnology, University of Helsinki, Helsinki 00014, Finland*

[4]*Stanford Synchrotron Radiation Laboratory, Stanford Linear Accelerator Center, Stanford University, Menlo Park, CA 94025-7015, USA*

Received 21 January 2008; received in revised form 4 June 2008; accepted 10 June 2008
Available online 14 June 2008

Edited by M. Moody

Bacteriophage P22 serves as a model for the assembly and maturation of other icosahedral double-stranded DNA viruses. P22 coat and scaffolding proteins assemble *in vitro* into an icosahedral procapsid, which then expands during DNA packaging (maturation). Efficient *in vitro* assembly makes this system suitable for design and production of monodisperse spherical nanoparticles (diameter ≈50 nm). In this work, we explore the possibility of controlling the outcome of assembly by scaffolding protein engineering. The scaffolding protein exists in monomer–dimer–tetramer equilibrium. We address the role of monomers and dimers in assembly by using three different scaffolding proteins with altered monomer–dimer equilibrium (weak dimer, covalent dimer, monomer). The progress and outcome of assembly was monitored by time-resolved X-ray scattering, which allowed us to distinguish between closed shells and incomplete assembly intermediates. Binding of scaffolding monomer activates the coat protein for assembly. Excess dimeric scaffolding protein resulted in rapid nucleation and kinetic trapping yielding incomplete shells. Addition of monomeric wild-type scaffold with excess coat protein completed these metastable shells. Thus, the monomeric scaffolding protein plays an essential role in the elongation phase by activating the coat and effectively lowering its critical concentration for assembly.

Keywords: self-assembly; virus; X-ray; kinetics; structure

Introduction

A typical viral capsid is composed of an outer protein shell that contains and protects the viral nucleic acid. Simple viral shells may be made of multiple copies of a single coat protein.[1] Virus assembly is a process in which well-organized, usually highly symmetric (icosahedral or helical), viral capsids are built by association of the protein subunits.[2] The advent of nanotechnology has brought renewed interest in viruses and their assembly.[3] While understanding virus assembly is crucial to tackling viral diseases,[4,5] it also provides a basis for the exploitation of viruses in the design of new nanostructures.[6,7] Current nanotechnology applications rely on modified viral coat proteins that assemble native-like icosahedral shells and serve as attachment scaffolds for other molecules or assemblies. However, advanced applications of viral capsids are likely to require purposeful control over the assembly process. For example, one would like to assemble capsids within nanofabricated compartments (e.g., lab-on-chip applications) in a controllable fashion. This will require careful control of assembly initiation by chemical or physical means, e.g., gradual buildup of concentration or surface-immobilized nucleation sites. Similarly, control over the outcome of assembly may be needed when partially assembled shells are desired or if incorporation of a bulky cargo inside the viral shell would require slower, gradual polymerization. Achieving such a level of control depends on a detailed understanding of the assembly process.

**Corresponding author.* Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK. E-mail address: r.tuma@leeds.ac.uk.

Abbreviations used: PC, procapsid; TR-SAXS, time-resolved small-angle X-ray scattering; SVD, singular value decomposition; wt, wild type; CTF, C-terminal fragment.

Several bacteriophages and viruses had been reconstituted *in vitro* from purified components and the morphogenetic pathways for many bacteriophages had been delineated using ingenious genetic schemes.[8] However, detailed kinetic analyses have been achieved for only a few of the simplest capsids, which are usually composed of one type of subunit.[9,10] Among the more complex systems, in which two or more proteins co-assemble, bacteriophage P22 assembly is perhaps the best understood. The assembly of bacteriophage P22 has been extensively studied *in vivo* by both genetic and structural means.[11–13] These studies have revealed that the capsid assembles in two steps. First, an empty procapsid (PC) results from co-assembly of the major coat protein (Gp5, 47 kDa, abbreviated CP) with the scaffolding protein (Gp8, 34 kDa, abbreviated SP) and several minor capsid proteins.[14] Then the proapsid shell undergoes a large conformational change and expands during DNA packaging.[15,16] An *in vitro* system for procapsid reconstitution had been established[17] and made possible detailed kinetic studies of the assembly process.[18] The *in vitro* studies demonstrated that CP and SP are essential as well as sufficient for the formation of isometric procapsid shells (Fig. 1a). Kinetic analyses indicated that PC assembly proceeds in two phases, a rate-limiting nucleation phase is followed by fast growth leading to completion of the shells. The putative nucleation complex was proposed to consist of a coat protein pentamer together with several SP molecules. Subsequent studies with different scaffolding protein mutants pointed out the prominent role of scaffolding dimer in assembly.[19]

Scaffolding protein exists in a monomer–dimer–tetramer equilibrium in solution,[20] and their respective roles in assembly were proposed on the basis of the final PC structure.[21–23] The scaffolding protein can be extracted from the procapsid by treatment with either mild denaturant (GuHCl, urea)[17], salt[24] or heat[25] without affecting the structure.[23] This suggests that the scaffolding protein's role is in the formation of transient complexes during nucleation and growth phases. Given that the final structure provides only limited information on the early steps of assembly, the role of different scaffolding protein forms in these steps remains elusive. In particular, the role of scaffolding protein monomer has not been elucidated.

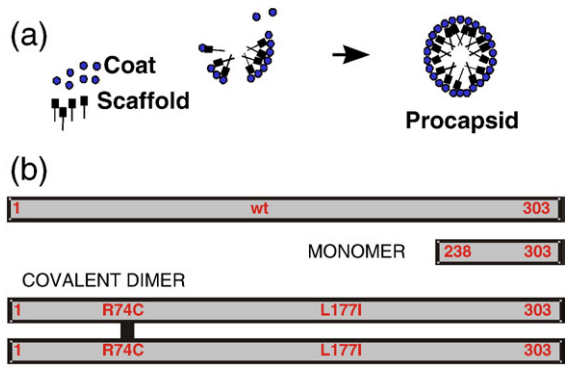


Fig. 1. (a) Scaffolding-protein-assisted *in vitro* assembly of bacteriophage P22 procapsid. (b) Different scaffolding protein constructs as used in this study.

This study assesses the role of different SP oligomers in assembly by time-resolved small-angle X-ray scattering (TR-SAXS). This technique, together with singular value decomposition (SVD),[26] enabled us to resolve complete shells from intermediates and aberrant products and to follow assembly under stoichiometric as well as substoichiometric conditions. In addition, we take advantage of various engineered scaffolding proteins (Fig. 1b) and delineate the role of the monomeric and the dimeric scaffolding protein in assembly. The results demonstrate that assembly rate and outcome can be controlled by manipulating the scaffolding protein dimerization equilibrium and coat protein concentrations.

Results

Wild-type scaffolding protein

SAXS-based assembly assay

Monomeric coat protein and wild-type (wt) scaffolding protein were manually mixed in two molar ratios (CP–SP 1:1 and 2:1, respectively) and their assembly process was monitored by SAXS (Figs. 2a and 3). The manual mixing procedure and activation of the X-ray hutch interlock system resulted in about 1-min dead time. The time dependence of the scattering revealed the gradual emergence of maxima and minima characteristic of closed shells. After 25 min, the reaction reached a stationary phase with little further change. The observed scattering curve was modeled by calculating the scattering from hollow spheres with outer and inner radii of 310 and 160 Å, respectively (Fig. 2b). The model fits well for q <0.04 Å^{-1} above which the departure of the icosahedral shell from the spherical model becomes apparent. The thickness of the shell (150 Å) includes the density from both coat and scaffolding protein. The overall dimensions are consistent with those derived from cryoelectron microscopy.[22,23]

Detection and structural characterization of assembly intermediates

The TR-SAXS data were plotted as a three-dimensional surface (Fig. 2a). Given the relatively high concentrations of CP and SP required for SAXS experiments, any intermediates would be expected to appear early during the assembly and would quickly grow into complete shells that would dominate scattering during the later stages of assembly. Due to their low abundance and small size any scattering arising from intermediates would be quite weak.[18] In order to examine whether the TR-SAXS data contain contributions from any putative intermediates, we employed SVD analysis. This mathematical procedure yields the number of linearly independent and orthogonal scattering components in TR-SAXS at a given noise level[27] (see Supplementary Methods for

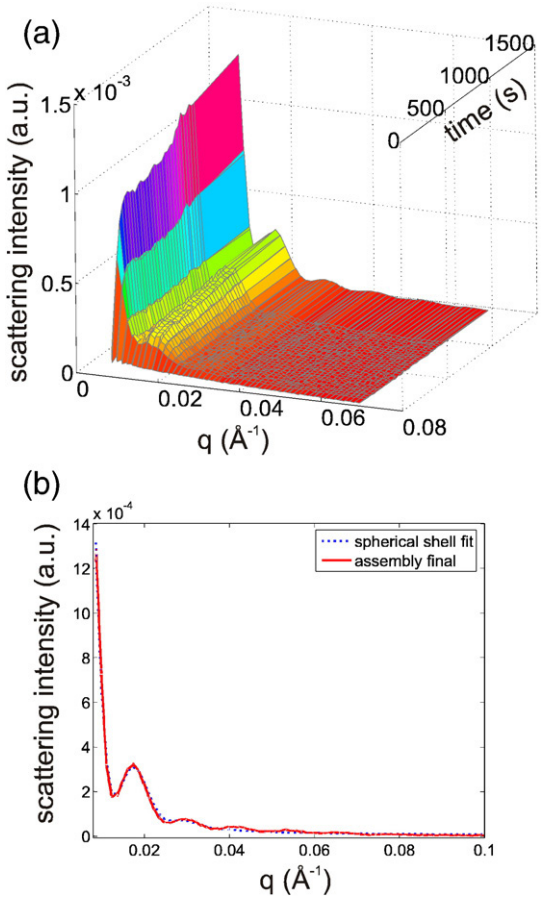


Fig. 2. Manual mixing (dead time about 1 min) and TR-SAXS using wt scaffold (14.7 μM) and coat (24.9 μM). (a) TR-SAXS. (b) Final product scattering (after 25 min, continuous red line) fitted to a hollow shell model (dotted blue): R_{inner} = 160 Å, R_{outer} = 311 Å.

detailed description). If there were no intermediates then the data set would be fully described (to within experimental noise) by two components, namely, scattering from the initial mixture of CP and SP and from the final shells, respectively, Such least-squares procedures failed to reproduce the data and significant, time-resolved residuals remained (Supplementary Methods Fig. S1). Hence, the data contained a significant scattering contribution from one or more assembly intermediates in addition to the scattering contributed by the initial and final species. In fact, three components (corresponding to the three highest SVD eigenvalues, Supplementary Methods Fig. S2) were necessary to reproduce the experimental data to within the noise level (Supplementary Methods Fig. S3). Because SVD yields orthogonal scattering and concentration factors, they alternate between positive and negative values (Supplementary Methods Fig. S4) and do not directly represent physically meaningful scattering curves or concentrations. However, the SVD scattering vectors represent a convenient orthogonal basis for obtaining a plausible and physically meaningful set of scattering curves as a linear combination. Constrained least-squares method and manual rotation were used to obtain scattering factors and concentration loadings that were non-negative to within the noise level (Fig. 3, see Ref. 28 and Supplementary Methods for further details).

Semi-quantitative scaling of scattering intensities (i.e., relative to each other) could be achieved by extrapolating the scattering factors to zero angle using the Guinier approximation.[29,30] In order to hold for shell-sized scattering objects, the Guinier approximation would require a significant portion of the experimental data to extend below $q < 0.005$ Å^{-1} (i.e., within the so called Guinier region, $R_g \cdot q \leq 2$). This was precluded by the limited experimental range ($q > 0.008$ for manual mixing data and $q > 0.012$ for stopped-flow experiments). In order to overcome this limitation, we assumed that the observed assembly intermediates were partial shells with rotational symmetry around the *z*-axis (i.e., spherical section inscribed by a cone with vertex angle 2Θ, Fig. 4a). This is a reasonable approximation, since other smaller species would not dominate in the scattering signal due to their sizes and would be less stable.[31] Then the scattering curves for such partial spherical shells were simulated for various degrees of completion (i.e., Θ from 0° to 180°, Fig. 4b). The true and apparent R_g values for each degree of completion were computed from these curves using the two experimentally accessible q ranges (Fig. 4c). Comparison of the simulated curves show that the extended Guinier region may be used for intermediates up to 45% completion, i.e., apparent R_g values up to 190 Å. Hence, we used the extended Guinier region to estimate the R_g and the degree of completeness, i.e., the relative apparent mass of the intermediate (from Fig. 4c). This relative mass was then used to scale the intensities of the scattering factors and amplitudes of relative mass concentrations (Fig. 3).

Note that the assignment of single apparent R_g and mass to a given scattering factor does not mean this represents a single intermediate. This could hardly be expected for a reaction of such complexity and, consequently, the values characterize an ensemble of intermediates. The apparent mass represents a weight-average value, while the apparent R_g is the *Z*-average across the distribution.[32] Both of these averages are biased towards larger species and, thus, these parameters effectively represent the upper bound of the distribution.

An apparent R_g of 155±25 Å (lower initial coat concentration, 1:1 CP–SP ratio) and 125±30 Å (higher coat concentration with 2:1 CP–SP ratio, i.e., similar to the stoichiometry found in native procapsids) was obtained for intermediates from the data in Fig. 3a and b, respectively. The apparent R_g values as well as the scattering curves agreed within experimental error for the two conditions. This suggests that in both cases, the largest intermediates were approximately of the same size corresponding to about 25–35% completeness (Fig. 4c). The scattering curves of the intermediates (red curves in Fig. 3) are also similar to the 30%-complete shell model (Fig. 4b). However, the kinetics exhibited distinct behavior for the two CP concentrations, namely, the maximum

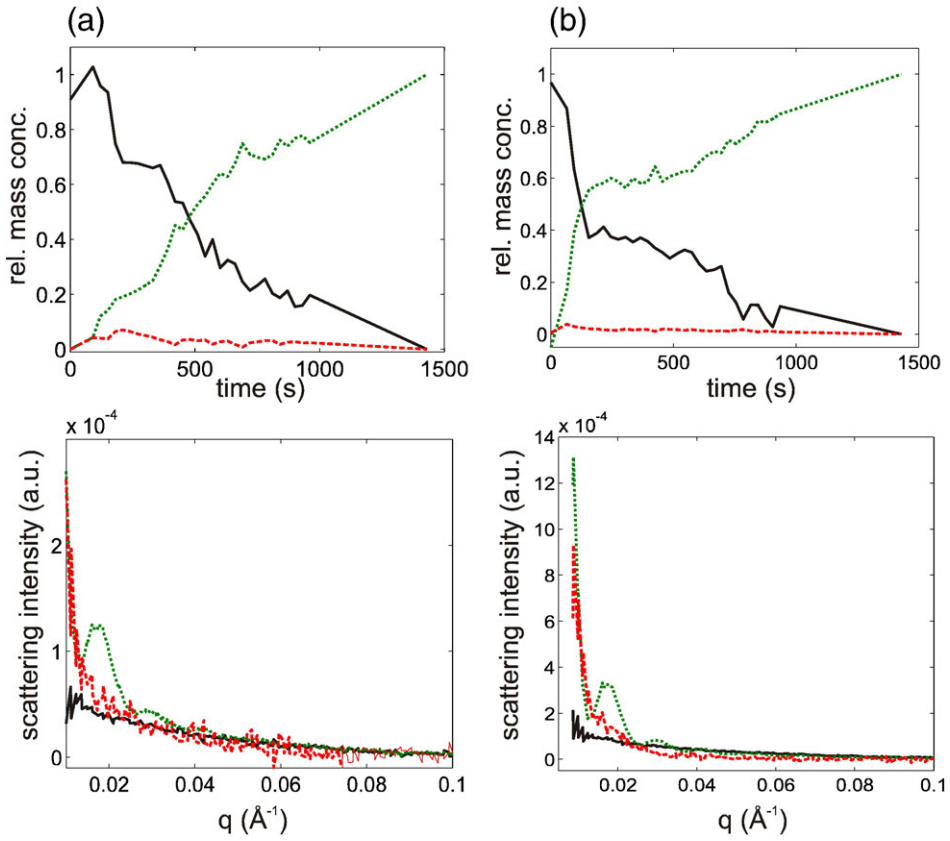


Fig. 3. Analysis of TR-SAXS from assembly reactions by SVD. Top panels are the factor loadings that represent relative distribution of protein mass between the species corresponding to the scattering factors in the bottom panels. (a) 14 μM CP+14.7 μM SP, (b) 24.9 μM CP+14.7 μM SP. Black traces (continuous line) correspond to the relative abundance or scattering factor of free coat and scaffolding protein (reactants). Red traces (dashed line) correspond to those of the intermediates. Green traces (dotted line) indicate the product concentrations and scattering factors.

in the intermediate concentration (representing 5% total mass fraction) was attained at about 65–100 s for the higher CP concentration, while the peak (8% total mass fraction) was delayed to 210 s for the lower concentration. Because of their relatively large size and preferential sensitivity of SAXS for larger species, these intermediates are still detectable despite their low abundance. On the other hand, their occurrence precluded detection of smaller species.

The 30%-complete shells contain roughly 86 CP subunits, i.e., a much larger number than the pentameric nucleation complex.[18] Similarly, the R_g values are also larger than those expected for the nucleation complex (~60–70 Å, estimated R_g for a pentameric complex). This indicates that smaller intermediates are either not significantly populated during assembly or populated only during the dead time of manual mixing experiments. Hence, we resorted to rapid mixing techniques to resolve early complexes of coat and scaffolding proteins that may lead to the formation of assembly nuclei.

Stopped-flow kinetics

Stopped-flow mixing was achieved using a two-syringe apparatus with an incorporated scattering flow cell.[33] In order to achieve the necessary sensitivity towards weakly scattering early intermediates (e.g., pre-nucleation complexes such as CP_2–SP_2 dimer), high concentrations of both CP (27.2 μM, the maximum achievable concentration for the monomeric, assembly-competent CP) and SP (22.9 μM) were used. The first 5 min of the assembly kinetics were sampled using 5- to 10-s exposures (Fig. 5a). SVD analysis revealed three significant factors that were then rotated to obtain non-negative, physically meaningful scattering factors (Fig. 5b, top) and the associated relative concentrations (Fig. 5b, bottom). The time course for the appearance of completed shells exhibited a noticeable lag time (70 s) previously assigned to the nucleation phase.[18] During this phase a considerable amount of coat protein was incorporated into intermediates. An apparent R_g of 170±50 Å was obtained from the initial slope (q values, 0.012–0.016 $Å^{-1}$) and the scattering factor in Fig. 5b resembled that of shells with completeness between 20% and 40%. Again, the size of these intermediates represents the upper bound of intermediate ensemble due to the Z-averaging nature of scattering intensities. This suggests that the rate-limiting nucleation step is followed by a rapid addition of coat/scaffold subunits to the partial shells until the number of sites for addition becomes large and the decreasing concentration of CP limits the rate of addition of coat protein to

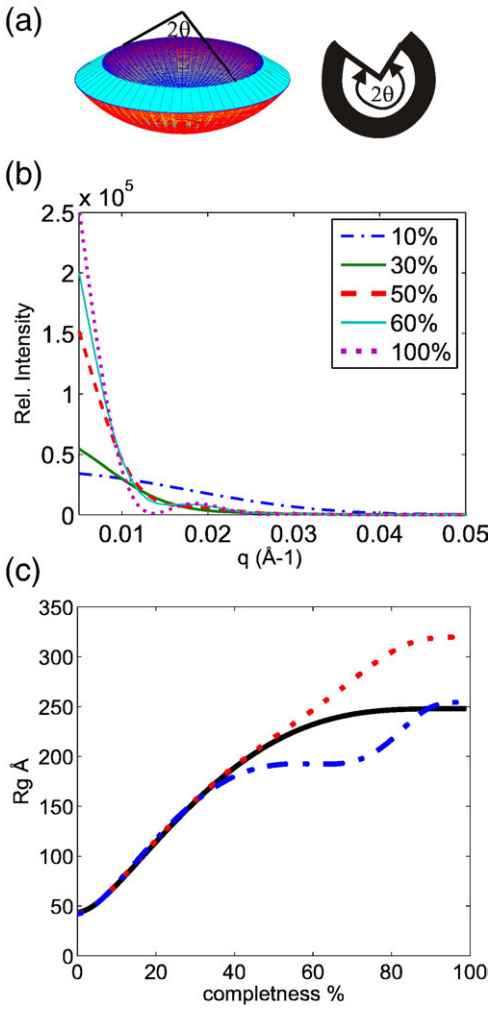


Fig. 4. Dependence of SAXS on the completeness of the shell. (a) Schematic illustration of the partial shell model (conical wedge, left; two-dimensional projection, right) with rotational symmetry used for simulating the scattering curves in panel b. The inner and outer radii obtained for the final state from the experimental data in Fig. 2b were used in the computations. The spherical Θ was varied from 0° to 180°, i.e., from 0% to 100% completeness (b) Representative simulated scattering curves for various degrees of completion. (c) Dependence of the radius of gyration (R_g) on completeness. The continuous black trace corresponds to the exact value computed directly from the model in panel a. The red dotted curve corresponds to the apparent R_g estimated from the simulated scattering curves in panel b using the q range 0.008–0.012 Å^{-1}. The blue dash–dot line shows the apparent R_g estimated from the simulated scattering curves using the q range 0.011–0.016 Å^{-1}.

the existing partial shells (elongation rate) and prevents shell closure. Under those conditions, the reaction is prone to kinetic trapping. Substantial accumulation of these intermediates suggests that at the initial high CP and SP concentrations, nucleation dominates over the elongation.

After the lag phase, the production of shells increased linearly over time until approximately 250 s. During this phase, the intermediate concentration peaked at about 150 s (at 10% total mass abundance) after which the intermediates were completed into shells, while nucleation of new assemblies was limited by the decreased CP and SP concentrations. Finally, the pools of free coat protein and partially assembled intermediates become exhausted and the assembly reaction enters a steady-state phase after approximately 250 s.

Comparison of the TR-SAXS data collected under different protein concentrations (Figs. 3 and 5) revealed that an increase in CP concentration speeds up the rate of elongation, while increasing SP concentration leads to a longer lag phase for complete shell formation and thus favors nucleation.

Scaffolding protein dimer

The scaffolding protein dimer has been implicated as the active species in P22 assembly. Here we took

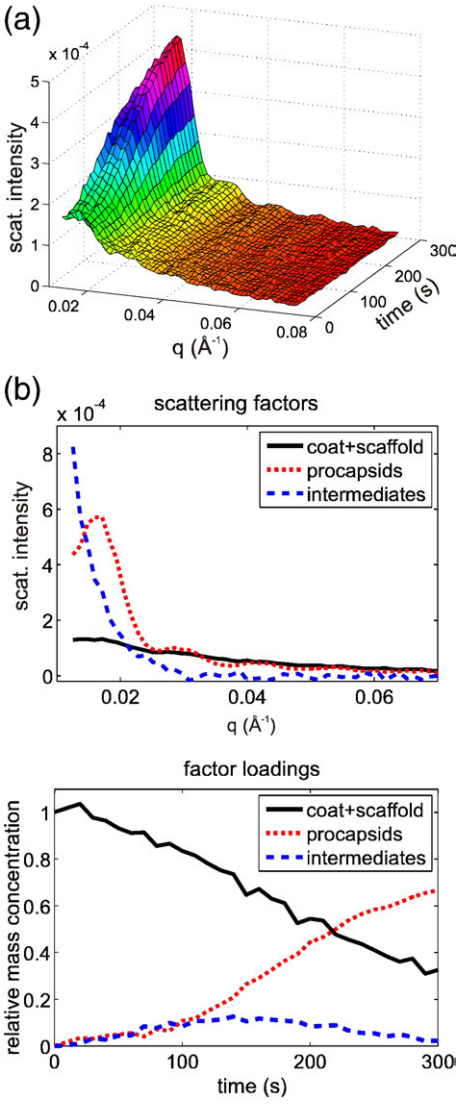


Fig. 5. P22 assembly kinetics monitored by stopped-flow TR-SAXS. (a) TR-SAXS curves rendered to 10-s resolution, 24.9 μM CP + 14.7 μM SP. (b) The scattering factors (top) and relative mass concentrations (bottom) computed from the data in (a) by SVD and factor rotation as described in Materials and Methods.

advantage of a reversible disulfide cross-link in an SP mutant (R74C/L177I, Fig. 1b) that leads to a covalent dimer formation.[19] We address the question of whether the SP dimer alone was sufficient for assembly. Although the assembly reactions were done under similar concentrations as in the case of wt SP, it is evident that the dimer alone is not sufficient for closed shell assembly *in vitro* (Fig. 6a). In a control experiment, the reduced form of the mutant SP exhibited assembly similar to that of the wt SP (Fig. 6d). This indicates that the defect is due to SP dimerization and not due to the mutation alone.

From the analysis of the assembly reaction, the final products had apparent R_g ranging from 80 to 180 Å depending on the initial CP concentration (Fig. 6b and c). Although higher CP concentrations led to apparently larger assemblies and kinetics exhibiting apparent intermediates (Fig. 6c), the final product did not contain any closed shells. The maxima and minima typical of scattering from spherical shells were not detected in any of the scattering factors. Under both CP concentrations, the largest assembly product exhibited an apparent R_g of 180 Å, which corresponds to 40% complete shells (Fig. 4c) provided these products were intermediates on the procapsid assembly pathway. Alternatively, these products could be aberrant assemblies.

To distinguish between these two possibilities, the distribution of assembled species obtained under different conditions were compared using sucrose gradients and electron microscopy (Fig. 7). In comparison with wt SP, the dimer was more efficient in inducing coat protein assembly (cf. CP amount at the top, 5% sucrose gradient fractions for rows 1 and 2 on the gel). However, the dimer induced assembly of broadly distributed, smaller (lighter) products, which appeared to be incomplete shells in a negative-stain electron micrograph (Fig. 7, right). Addition of wt SP to the dimer reaction restored the product distribution to almost the wild-type pattern (cf. rows 1 and 3). The addition of excess CP and simultaneous decrease of SP dimer concentration had a similar effect (row 5), while reducing both CP and SP concentrations yielded less complete shells (row 4). The incomplete assemblies were rescued by the addition of fresh CP after the stationary phase was reached with the dimer SP (not shown). This demonstrates that the products of assembly reaction with R74C dimer are PC assembly intermediates.

The SAXS and sucrose gradient results suggest that the SP dimer is more effective in nucleating assembly than the wt SP (monomer/dimer equilibrium) and decreases the critical concentration of coat protein (most CPs had assembled in the presence of R74C SP). Hence, the incomplete shells constitute kinetically trapped intermediates that are produced by overnucleation of assembly by the dimeric scaffold. From the rescue of assembly by wt SP, it is clear that there is a role for the scaffold monomer, most likely during the elongation phase.

Scaffolding protein monomer

The C-terminal 38-amino-acid residues of the scaffolding protein encompass the coat protein binding site.[34] The C-terminal fragment (CTF) lacks the dimerization domain and thus remains monomeric.[19,35] Despite the lack of dimerization, the monomeric CTF induced rapid assembly (Fig. 8a). However, the products were polydisperse and lacked scattering features typical of complete shells, i.e., maxima and minima. A broad range of products was found by sucrose gradient analysis (Fig. 8b). This shows that SP monomer switches the coat protein into an assembly-competent form and effectively decreases the critical concentration for assembly.

Discussion

Three phases in assembly

In our TR-SAXS study, we have resolved a nucleation phase during which rapid initiation of assembly leads to accumulation of partial shell intermediates. For wt SP this phase leads to the establishment of a steady state during which the nucleation rate matches the rate of shell completion and the amount of PC increases approximately linearly with time. Eventually, coat protein becomes depleted and the reaction enters a stationary phase during which there is no net increase in assembly products. The stationary phase is particularly pronounced for the kinetically trapped assembly reactions of SP covalent dimer (Fig. 6). This phase should not be confused with the final equilibrium, since kinetically trapped intermediates can be completed either by addition of coat protein (as shown here in Fig. 7) or by slow disassembly and reassembly.[36,37]

Experimental limitations (poor signal-to-noise ratio at low protein concentrations) precluded examination of broader protein concentration range. However, qualitative conclusions can be drawn from the concentrations sampled. The duration of the nucleation phase, measured as the time at which the concentration of intermediates peaks, decreases with increasing initial concentration of CP (cf. Fig. 3a and b) and increases upon doubling the SP concentration (i.e., increase in the fraction of dimer from 10% to 18%,[20] cf. Figs. 3b and 5b). The covalent SP dimer accelerated nucleation, completely eliminated the steady-state phase and yielded only kinetically trapped partial shells. This could be due to its higher affinity for the shells. However, the covalent SP dimer also decreased the apparent critical coat concentration for assembly (note the lower amount of unassembled coat on the top of sucrose gradients in Fig. 7 for the cross-linked R74C), pointing out its direct role in nucleation.[14]

The partial shells could be completed by addition of excess coat protein or by including even a small amount of wt SP in the reaction (i.e., monomeric SP). This suggests that efficient elongation is accelerated

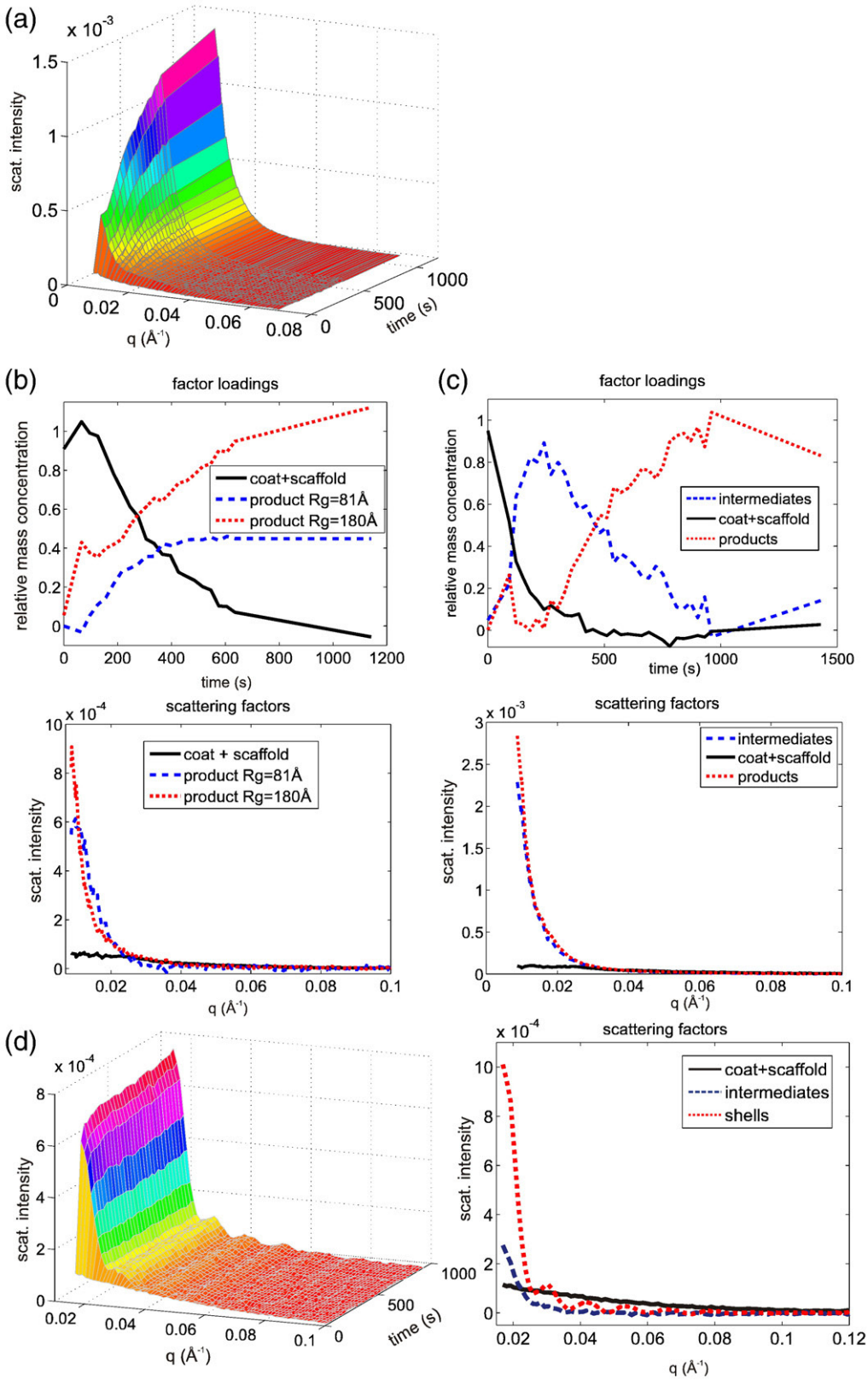


Fig. 6. Assembly of CP with R74C/L177I disulfide dimer SP. (a) TR-SAXS at 24.5 μM CP, 7.4 μM SP dimer (monomer concentration 14.7 μM). (b and c) Scattering factors (bottom) and relative concentrations (top) for two initial conditions: (b) 12.8 μM CP, 7.4 μM SP dimer; (c) 24.5 μM CP, 7.4 μM SP dimer. The dashed lines indicate that only the end point at 1500 s was collected after the finish of the time-resolved series at 900 s. (d) Assembly with reduced R74C: TR-SAXS on the left, scattering factors on the right; 16.6 μM CP, 18.5 μM SP (monomer concentration).

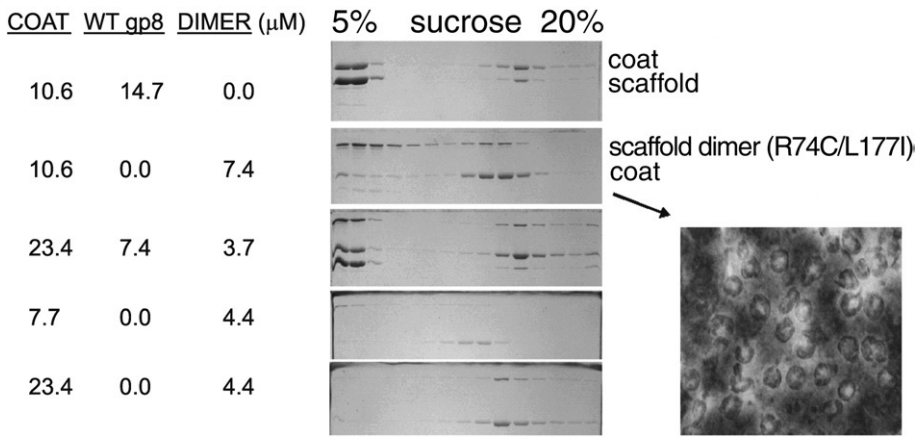


Fig. 7. Scaffolding monomer is required for assembly of complete shells. R74C dimer promotes nucleation. Left: SDS-PAGE analysis of fractions after sedimentation of assembly products on a 5–20% sucrose gradient. Right: negatively stained assemblies from the peak fraction of the scaffold dimer reaction. Protein concentrations (micromolar) are shown in columns on the left.

by monomeric SP. Our observation that the monomeric scaffolding fragment CTF promotes coat aggregation suggests that SP monomer may activate CP for assembly.

Activation of coat for assembly by scaffolding monomer

The activation of CP by SP constitutes a step common to both the nucleation and elongation phases. This activation relies on a direct CP–SP interaction that is predominantly electrostatic.[24] Low salt concentration would be expected to promote CP–SP interactions and thus lead to accelerated assembly. In fact, excessive nucleation and rapid assembly of partial shells was observed under low-salt conditions.[38] These authors concluded that strengthening the electrostatic interactions between CP and SP promoted rapid nucleation and growth of shells.[38] In addition, they showed that low ionic strength led to conformational changes in CP but not in SP. The altered conformation may, perhaps, be more amenable to assembly. PC assembly is accompanied by significant change in CP secondary structure.[39] This and the partially unfolded character of CP monomer[40] suggest that CP monomer may exist in solution as an ensemble of conformations of which only a fraction is assembly competent. Binding of SP monomer may sway the equilibrium in favor of assembly-competent conformations and thus activate the coat protein. This is reminiscent of bacteriophage MS2 coat activation for assembly by specific RNA binding.[41] However, in this case the conformational switching is involved in early stages of assembly and RNA packaging.

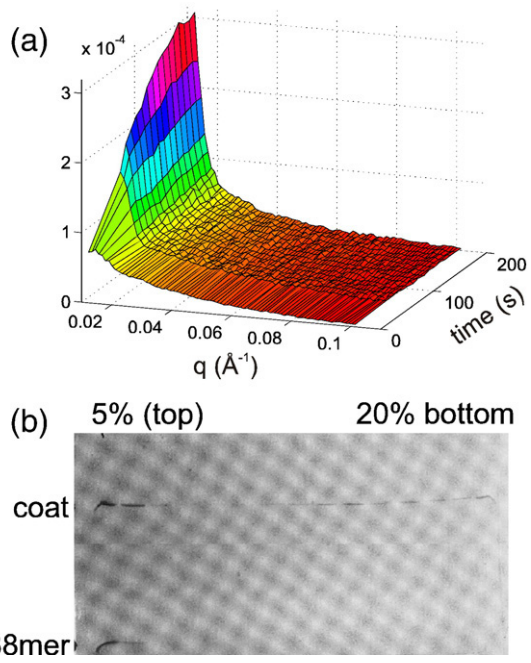


Fig. 8. Assembly of CP with SP CTF. (a) Stopped-flow TR-SAXS. (b) Assembly products after 60 min were separated by a sucrose gradient and fractions were analyzed by 13% SDS-PAGE.

The assembly phenotypes observed by Parent *et al.*[38] at low ionic strength are remarkably similar to those observed here for the SP covalent dimer under standard ionic strength. This suggests that perhaps, in addition to affecting CP–SP interactions, the low-salt conditions may also affect the monomer–dimer equilibrium of the highly polar and charged SP and promote nucleation by excess of scaffolding dimer.

Role of scaffolding monomer–dimer equilibrium in assembly

It is clear from the above discussion that monomeric SP is required for assembly, most likely in the later stages. In fact, addition of wt SP to the reaction mixture (as low as 0.05 mg/ml, i.e., 95% monomeric) also alleviated the overnucleation-induced kinetic trap caused by the covalent dimer (Fig. 7 and data not shown). This substoichiometric effect suggests that the monomeric SP may transiently bind to CP,

activate it and then dissociate instead of remaining stably associated with the resulting PC shell.

The monomeric CTF activates CP for assembly, but unlike the dimer, it is not stably incorporated into the products (cf. Fig. 7 and Fig. 8b). This suggests that the affinity of the coat protein for the monomeric SP is low or may even decrease during the course of assembly, while the affinity for the dimer is higher. Indeed, the procapsid exhibits two distinct types of binding sites for wt SP with different affinities that were assigned to monomeric and dimeric states, respectively.[42] Thus, the monomeric SP most likely transiently interacts with CP and activates it for assembly, effectively increasing the concentration of assembly-competent CP and increasing the elongation rate (Fig. 9). This also competes with nucleation, which is promoted by the SP dimer. Thus, the SP monomer–dimer equilibrium is ultimately linked to the nucleation–elongation balance and may be effective in steering the course and outcome of the assembly reaction (Fig. 9).

P22 assembly differs from the structurally related bacteriophage HK97

P22 coat protein has a fold similar to that of another double-stranded DNA bacteriophage HK97.[16] However, the latter utilizes a two-step assembly pathway in which relatively stable pentamers and hexamers serve as the building blocks for the prohead (procapsid equivalent in HK97) formation.[1] Thus, coat protein fold similarity and identical capsid architecture may not automatically result in the same assembly pathway. The different assembly pathways may be due to the radically different interaction between coat and scaffolding proteins. While P22 SP interacts with CP non-covalently, the HK97 has its scaffolding element, the N-terminal 110-residue Δ-domain, integrated within the coat protein. The Δ-domain is proteolytically cleaved within the assembled prohead and subsequently released. Very little is known about the interactions between the putative scaffolding domains in HK97. On the other hand, coat–coat interactions, which are weak for P22, dominate the early assembly of HK97 (i.e., pentamer–hexamer formation). Hence, the final structure of a virus may give us little information about the assembly mechanism unless a quantitative survey of protein–protein interactions is performed and related to the solution conditions. The latter are important because assembly is governed by a set of relatively weak protein–protein interactions,[43] and these in turn are extremely sensitive to solution conditions such as ionic strength and pH.

Kinetic control over assembly *in vitro* and implications for *in vivo*

The detectable amounts of P22 assembly intermediates also demonstrate that under the conditions used here the assembly reactions are far from equilibrium. Even under those conditions viral shells are being produced when wt SP is present, indicating that the assembly process is quite robust. However, the results presented here and elsewhere[38,31] suggest that virus assembly is prone to kinetic trapping under conditions of strong protein–protein interactions. The common cause is either high concentration of subunits, which leads to overnucleation, or imbalance of intermolecular interactions (e.g., mutations). Specifically, in a system with several co-assembling components, such as P22, balance of the interactions and concentrations for all components plays an essential role in successful assembly. Hence, the scaffolding protein association is relatively weak and dynamic, and coat protein associates only upon interaction with SP. In addition, the scaffolding protein concentration is being maintained inside the infected cell by a translational control in which the N-terminal domain acts as an autorepressor.[44]

Recent theoretical studies demonstrated that the assembly mechanisms could abruptly switch from a productive mode to a non-productive one upon a slight shift in concentrations or strength of subunit interactions.[45] In general, conditions of slower growth are less prone to kinetic trapping,[31] and weak protein–protein interactions represent one way to achieve this.[46] Alternatively, low protein concentrations need to be maintained, but such conditions

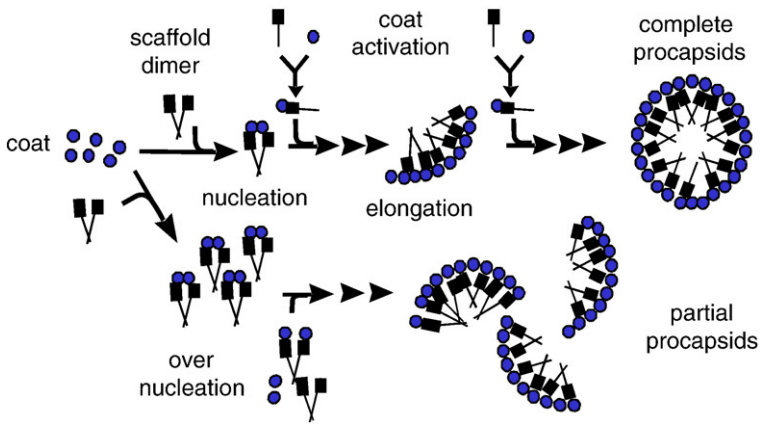


Fig. 9. Role of different scaffolding protein oligomers in assembly of bacteriophage P22 procapsid.

would interfere with the need for high protein production in a short time during infection, a condition applicable to most lytic viruses.

Although not desirable during the course of productive infection, kinetic trapping may be useful in the manufacture of virus-based nanoparticles. The stepwise completion of the shells *in vitro* as demonstrated here could be exploited for, e.g., encapsidating guest material or incorporating selectively labeled subunits that may otherwise interfere with the nucleation step of assembly. In the P22 system, the kinetic trap may be "programmed" at ease by manipulating the self-association of the highly soluble scaffolding protein while using the hard-to-obtain coat protein in its wild-type, assembly-competent, form.

Materials and Methods

Protein purification

Coat protein was purified fresh just before the assembly experiments to ensure monodispersity as described previously.[39] The wt scaffolding protein was extracted from procapsids expressed in *Escherichia coli* and purified as described.[20] Engineered scaffolding proteins were expressed in *E. coli* and purified as described.[19,20] All proteins were centrifuged for 15 min at 14,000*g* at 4 °C just before use. The R74C/L177I disulfide dimer was reduced by 3.3 mM glutathione.

In vitro assembly reactions

Assembly reactions were started by mixing (either manually or by a stopped-flow mixer) a small amount (ca 5–10 μl) of concentrated scaffolding protein into excess of coat protein. All reactions were done in buffer B (50 mM Tris, 25 mM NaCl, 1 mM ethylenediaminetetraacetic acid, pH 7.6) at 20 °C. If needed, assembly products were analyzed by sedimentation velocity on 5–20% (w/v) sucrose gradient (50-μl loading per tube). The gradient was run in an SW55 rotor at 20 °C for 55 min and fractionated into 250-μl fractions, which were subsequently analyzed on SDS-PAGE.

TR-SAXS data collection and processing

All SAXS data collection was performed at SSRL beam line 4-2 as previously described.[39] Stopped-flow time-resolved experiments were conducted using the experimental procedures previously published[47] with the exceptions of different exposure times and slightly different q range covered. The magnitude of momentum transfer q is defined as $4\pi \sin(\theta)/\lambda$, where θ is one-half the scattering angle and λ is the X-ray wavelength.

A limited region of q values containing most information (0.008–0.15 $Å^{-1}$ for the manual mixing runs and 0.011–0.15 $Å^{-1}$ for stopped-flow experiments) was selected for SVD analysis of time-resolved scattering. SVD was performed in Matlab by a standard protocol.[27] The number of significant scattering factors was determined by ranking the singular values[48] and by comparing the residuals between the reconstructed and experimental data with the experimental noise level[26] (see Supplementary Methods, Fig. S2 and S3).

The raw eigenvectors (scattering curves, scattering factors) and their cognate loading factors (time-dependent concentrations) as produced by SVD are orthogonal and thus generally do not have a direct physical meaning, e.g., often giving negative values (see Supplementary Methods for further discussion). Thus, linear combinations (rotations) of the SVD eigenvectors were generated to obtain physically meaningful scattering curves and their corresponding time-dependent concentrations.[28,49,50] First, the scattering curves of the initial (I) and the final (F) state were approximated by linear combinations of the SVD eigenvectrors in such a way that their loading factors were non-negative. Then an appropriate linear combination approximately orthogonal to both I and F was found to yield a non-negative scattering curve and the corresponding non-negative loading factors (Fig. 3). This procedure yielded similar TR-SAXS results at two different CP concentrations (cf. Fig. 3a and b). After scaling of intensities and concentrations, we obtained a unique solution to the problem (see Supplementary Methods for additional details).

Modeling of incomplete shells

Incomplete hollow shells with inner and outer radii of 160 and 310 Å, respectively, were modeled as spherical sections inscribed by a cone with angle θ. The distance distributions $H(R, \theta)$ for these partial shells were obtained computationally by direct enumeration for each value of θ (using steps of 0.9°). Scattering curves were obtained from the distance distributions by Fourier transform.[29]

Acknowledgements

R.T. was in part supported by a research fellowship from Academy of Finland (1118462). P.E.P. was supported by grant GM47980 from the National Institutes of Health (NIH) and grant DBI-9726698 from the National Science Foundation. X-ray scattering studies were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the NIH, National Center for Research Resources (NCRR), Biomedical Technology Program Grant 5 P41 RR001209. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.06.020

References

1. Hendrix, R. W. (2005). Bacteriophage HK97: assembly of the capsid and evolutionary connections. *Adv. Virus Res.* **64**, 1–14.

2. Reddy, V. S. & Johnson, J. E. (2005). Structure-derived insights into virus assembly. *Adv. Virus Res.* **64**, 45–68.
3. Wang, Q., Lin, T., Johnson, J. E. & Finn, M. G. (2002). Natural supramolecular building blocks. Cysteine-added mutants of cowpea mosaic virus. *Chem. Biol.* **9**, 813–819.
4. Stray, S. J., Bourne, C. R., Punna, S., Lewis, W. G., Finn, M. G. & Zlotnick, A. (2005). A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. *Proc. Natl Acad. Sci. USA*, **102**, 8138–8143.
5. Sakalian, M., McMurtrey, C. P., Deeg, F. J., Maloy, C. W., Li, F., Wild, C. T. & Salzwedel, K. (2006). 3-*O*-(3′,3′-dimethysuccinyl) betulinic acid inhibits maturation of the human immunodeficiency virus type 1 Gag precursor assembled *in vitro*. *J. Virol.* **80**, 5716–5722.
6. Wang, Q., Lin, T., Tang, L., Johnson, J. E. & Finn, M. G. (2002). Icosahedral virus particles as addressable nanoscale building blocks. *Angew. Chem. Int. Ed. Engl.* **41**, 459–462.
7. Falkner, J. C., Turner, M. E., Bosworth, J. K., Trentler, T. J., Johnson, J. E., Lin, T. & Colvin, V. L. (2005). Virus crystals as nanocomposite scaffolds. *J. Am. Chem. Soc.* **127**, 5274–5275.
8. Abendon, S. T. & Calendar, R. L. (2005). The Bacteriophages, 2nd edit. Oxford University Press, New York.
9. Zlotnick, A., Aldrich, R., Johnson, J. M., Ceres, P. & Young, M. J. (2000). Mechanism of capsid assembly for an icosahedral plant virus. *Virology*, **277**, 450–456.
10. Zlotnick, A., Johnson, J. M., Wingfield, P. W., Stahl, S. J. & Endres, D. (1999). A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry*, **38**, 14644–14652.
11. Casjens, S. & King, J. (1974). P22 morphogenesis: I. Catalytic scaffolding protein in capsid assembly. *J. Supramol. Struct.* **2**, 202–224.
12. Earnshaw, W., Casjens, S. & Harrison, S. C. (1976). Assembly of the head of bacteriophage P22: X-ray diffraction from heads, proheads and related structures. *J. Mol. Biol.* **104**, 387–410.
13. King, J., Botstein, D., Casjens, S., Earnshaw, W., Harrison, S. & Lenk, E. (1976). Structure and assembly of the capsid of bacteriophage P22. *Philos. Trans. R. Soc. London, Ser. B*, **276**, 37–49.
14. Prevelige, P. E., Jr & King, J. (1993). Assembly of bacteriophage P22: a model for ds-DNA virus assembly. *Prog. Med. Virol.* **40**, 206–221.
15. Prasad, B. V., Prevelige, P. E., Marietta, E., Chen, R. O., Thomas, D., King, J. & Chiu, W. (1993). Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. *J. Mol. Biol.* **231**, 65–74.
16. Jiang, W., Li, Z., Zhang, Z., Baker, M. L., Prevelige, P. E., Jr & Chiu, W. (2003). Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat. Struct. Biol.* **10**, 131–135.
17. Prevelige, P. E., Jr, Thomas, D. & King, J. (1988). Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells *in vitro*. *J. Mol. Biol.* **202**, 743–757.
18. Prevelige, P. E., Jr, Thomas, D. & King, J. (1993). Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* **64**, 824–835.
19. Parker, M. H., Casjens, S. & Prevelige, P. E., Jr (1998). Functional domains of bacteriophage P22 scaffolding protein. *J. Mol. Biol.* **281**, 69–79.
20. Parker, M. H., Stafford, W. F., 3rd & Prevelige, P. E., Jr (1997). Bacteriophage P22 scaffolding protein forms oligomers in solution. *J. Mol. Biol.* **268**, 655–665.
21. Thuman-Commike, P. A., Greene, B., Malinski, J. A., King, J. & Chiu, W. (1998). Role of the scaffolding protein in P22 procapsid size determination suggested by $T=4$ and $T=7$ procapsid structures. *Biophys. J.* **74**, 559–568.
22. Thuman-Commike, P. A., Greene, B., Jakana, J., Prasad, B. V., King, J., Prevelige, P. E., Jr & Chiu, W. (1996). Three-dimensional structure of scaffolding-containing phage p22 procapsids by electron cryomicroscopy. *J. Mol. Biol.* **260**, 85–98.
23. Thuman-Commike, P. A., Greene, B., Malinski, J. A., Burbea, M., McGough, A., Chiu, W. & Prevelige, P. E., Jr (1999). Mechanism of scaffolding-directed virus assembly suggested by comparison of scaffolding-containing and scaffolding-lacking P22 procapsids. *Biophys. J.* **76**, 3267–3277.
24. Parker, M. H. & Prevelige, P. E., Jr (1998). Electrostatic interactions drive scaffolding/coat protein binding and procapsid maturation in bacteriophage P22. *Virology*, **250**, 337–349.
25. Galisteo, M. L. & King, J. (1993). Conformational transformations in the protein lattice of phage P22 procapsids. *Biophys. J.* **65**, 227–235.
26. Fowler, A. G., Foote, A. M., Moody, M. F., Vachette, P., Provencher, S. W., Gabriel, A. *et al.* (1983). Stopped-flow solution scattering using synchrotron radiation: apparatus, data collection and data analysis. *J. Biochem. Biophys. Methods*, **7**, 317–329.
27. Golub, G. H. & Van Loan, C. F. (1996). Matrix Computations, 3rd edit. The Johns Hopkins University Press, Baltimore, MD.
28. Lanman, J., Tuma, R. & Prevelige, P. E., Jr (1999). Identification and characterization of the domain structure of bacteriophage P22 coat protein. *Biochemistry*, **38**, 14614–14623.
29. Svergun, D. I., Semenyuk, A. V. & Feigin, L. A. (1988). Small-angle scattering data treatment by the regularization method. *Acta Crystallogr. A*, **44**, 244–250.
30. Guinier, A. (1939). Diffraction of X-rays of very small angles—application to the study of ultramicroscopic phenomena. *Ann. Phys.* **12**, 161.
31. Endres, D. & Zlotnick, A. (2002). Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* **83**, 1217–1230.
32. Feigin, L. A. & Svergun, D. I. (1987). Structure Analysis by Small-Angle X-Ray and Neutron Scattering. Plenum Press, New York.
33. Kihara, H. (1994). Stopped-Flow Apparatus for X-ray Scattering and XAFS. *J. Synchrotron Rad.* **1**, 74–77.
34. Sun, Y., Parker, M. H., Weigele, P., Casjens, S., Prevelige, P. E., Jr & Krishna, N. R. (2000). Structure of the coat protein-binding domain of the scaffolding protein from a double-stranded DNA virus. *J. Mol. Biol.* **297**, 1195–1202.
35. Parker, M. H., Jablonsky, M., Casjens, S., Sampson, L., Krishna, N. R. & Prevelige, P. E., Jr (1997). Cloning, purification, and preliminary characterization by circular dichroism and NMR of a carboxyl-terminal domain of the bacteriophage P22 scaffolding protein. *Protein Sci.* **6**, 1583–1586.
36. Parent, K. N., Zlotnick, A. & Teschke, C. M. (2006). Quantitative analysis of multi-component spherical virus assembly: scaffolding protein contributes to the

global stability of phage P22 procapsids. *J. Mol. Biol.* **359**, 1097–1106.
37. Parent, K. N., Suhanovsky, M. M. & Teschke, C. M. (2007). Phage P22 procapsids equilibrate with free coat protein subunits. *J. Mol. Biol.* **365**, 513–522.
38. Parent, K. N., Doyle, S. M., Anderson, E. & Teschke, C. M. (2005). Electrostatic interactions govern both nucleation and elongation during phage P22 procapsid assembly. *Virology*, **340**, 33–45.
39. Tuma, R., Tsuruta, H., Benevides, J. M., Prevelige, P. E., Jr & Thomas, G. J., Jr (2001). Characterization of subunit structural changes accompanying assembly of the bacteriophage P22 procapsid. *Biochemistry*, **40**, 665–674.
40. Kang, S. & Prevelige, P. E., Jr (2005). Domain study of bacteriophage p22 coat protein and characterization of the capsid lattice transformation by hydrogen/deuterium exchange. *J. Mol. Biol.* **347**, 935–948.
41. Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J. *et al.* (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a $T=3$ viral capsid. *J. Mol. Biol.* **369**, 541–552.
42. Parker, M. H., Brouillette, C. G. & Prevelige, P. E., Jr (2001). Kinetic and calorimetric evidence for two distinct scaffolding protein binding populations within the bacteriophage P22 procapsid. *Biochemistry*, **40**, 8962–8970.
43. Zlotnick, A. (2003). Are weak protein–protein interactions the general rule in capsid assembly? *Virology*, **315**, 269–274.
44. Wyckoff, E. & Casjens, S. (1985). Autoregulation of the bacteriophage P22 scaffolding protein gene. *J. Virol.* **53**, 192–197.
45. Zhang, T. & Schwartz, R. (2006). Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics. *Biophys. J.* **90**, 57–64.
46. Ceres, P. & Zlotnick, A. (2002). Weak protein–protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry*, **41**, 11525–11531.
47. Canady, M. A., Tsuruta, H. & Johnson, J. E. (2001). Analysis of rapid, large-scale protein quaternary structural changes: time-resolved X-ray solution scattering of *Nudaurelia capensis* o virus (NoV) maturation. *J. Mol. Biol.* **311**, 803–814.
48. Quinn, G. P. & Keough, M. J. (2002). Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge, UK.
49. Paatero, P. & Tapper, U. (1993). Analysis of different modes of factor analysis as least squares fit problems. *Chemom. Intell. Lab. Syst.* **18**, 183–194.
50. Paatero, P. & Tapper, U. (1994). Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, **5**, 111–126.

Molecular Cell, Vol. 7, 845–854, April, 2001, Copyright ©2001 by Cell Press

Self-Assembly of a Viral Molecular Machine from Purified Protein and RNA Constituents

Minna M. Poranen, Anja O. Paatero,† Roman Tuma, and Dennis H. Bamford*
Department of Biosciences and
Institute of Biotechnology
University of Helsinki
00014 Helsinki
Finland

Summary

We present the assembly of the polymerase complex (procapsid) of a dsRNA virus from purified recombinant proteins. This molecular machine packages and replicates viral ssRNA genomic precursors in vitro. After addition of an external protein shell, these in vitro self-assembled viral core particles can penetrate the host plasma membrane and initiate a productive infection. Thus, a viral procapsid has been assembled and rendered infectious using purified components. Using this system, we have studied the mechanism of assembly of the common dsRNA virus shell and the incorporation of a symmetry mismatch within an icosahedral capsid. Our work demonstrates that this molecular machine, self-assembled under defined conditions in vitro, can function in its natural environment, the cell cytoplasm.

Introduction

Crucial cellular functions are performed by macromolecular complexes composed of several components. How such structures are assembled is one of the intriguing questions of modern biology. The now classic in vitro studies on reconstitution of ribosomes have provided detailed mechanistic insight into the assembly of this vital cytoplasmic machinery (reviewed by Nomura, 1973). Another cytoplasmic molecular machine is the polymerase complex of dsRNA viruses (Bamford, 2000). This particle performs genome packaging, RNA replication, messenger synthesis, and capping activities and is carried, as part of the virion, from one infected cell to another. Although, dsRNA viruses infect a diversity of host organisms from bacteria (ϕ6; *Cystoviridae*) to vertebrates (rotavirus, reovirus, bluetongue virus; *Reoviridae*), the architecture of the polymerase complex is strikingly similar (Cheng et al., 1994; Butcher et al., 1997; Grimes et al., 1998; Reinisch et al., 2000).

Detailed understanding of any assembly processes requires defined in vitro systems where the conditions and the assembly components can be manipulated. On the other hand, an assay for the functionality of the assembly product is required to monitor the quality of the end product. One of the advantages of virus models is that the functionality of the assembly product (viral particle) can be monitored in vivo using a sensitive plaque assay.

The assembly of a virus particle is a complex process that involves a large number of protein–protein, protein–nucleic acid, and protein–lipid interactions. Assembly of infectious virions of simple viruses, like tobacco mosaic virus (TMV) and cowpea chlorotic mottle virus (CCMV), have been achieved in vitro by condensation of the genomic ssRNA and coat protein constituents (Fraenkel-Conrat and Williams, 1955; Bancroft and Hiebert, 1967). A more complex assembly cascade is utilized by many animal and bacterial viruses, which first assemble an empty precursor capsid, the procapsid, into which the genome is subsequently packaged. For several dsDNA viruses, procapsid-like structures have been successfully assembled in vitro from purified components (Prevelige et al., 1988; Cerritelli and Studier, 1996; Newcomb et al., 1999; Wang et al., 2000), without genome packaging. In contrast, efficient genome packaging systems have been established for the in vivo produced procapsids of several other dsDNA viruses (Guo et al., 1986; Hwang and Feiss, 1995) and the dsRNA virus ϕ6 (Gottlieb et al., 1990, 1991). Interestingly, ϕ6 packaging shares some similarities with the dsDNA viruses, like the symmetry mismatch between the packaging machinery and the rest of the capsid (Dube et al., 1993; de Haas et al., 1999; Simpson et al., 2000). Such a symmetry mismatch has been proposed to assist the rotation of the packaging machinery during nucleic acid transport into the capsid (Hendrix, 1978). To date, no system has been reported for which a combination of procapsid assembly and packaging has been achieved in vitro using purified components. Such a system would allow one to address the role of symmetry mismatch in both procapsid assembly and genome packaging. This study describes the in vitro assembly of packaging competent procapsids (polymerase complexes) of bacteriophage ϕ6 and their subsequent maturation to infectious nucleocapsids (NCs).

Bacteriophage ϕ6 has a genome composed of three segments of dsRNA (S, M, and L) packaged inside a polymerase complex, which is covered by a protein shell and a lipid–protein envelope (reviewed by Mindich and Bamford, 1988). The first assembly intermediate detected in bacteriophage ϕ6 infection is an empty polymerase complex, the procapsid (PC), consisting of a shell-forming protein P1 (Olkkonen and Bamford, 1987), an RNA-dependent RNA polymerase P2 (Makeyev and Bamford, 2000), a packaging NTPase P4 (Gottlieb et al., 1992), and a packaging factor P7 (Juuti and Bamford, 1995, 1997). The major structural protein P1 appears to be in two sets of distinct environments in a 120-subunit, icosahedral shell structure (Butcher et al., 1997), equivalent to the arrangement found in bluetongue and reovirus cores (Grimes et al., 1998; Reinisch et al., 2000). The packaging NTPase (P4) of ϕ6 is a ring-shaped hexamer located at the five-fold symmetry positions of the 47 nm size PC, thus creating a symmetry mismatch (Juuti et al., 1998; de Haas et al., 1999).

*To whom correspondence should be addressed (e-mail: dennis.bamford@helsinki.fi).
†Present address: Department of Microbiology, University of Alabama, Birmingham, Alabama 35294.

Table 1. Properties of ϕ6 NC Components

Protein	Mass (Da)[a]	Copies/ Virion[b]	Multimeric Status	Hydrodynamic Radius (nm)	Shape and Dimensions (nm)[g]	Genome Segment	Base Pairs[h]
P1	84,986	120	Monomer[c]	3.8 ± 0.4[c]	Ellipsoid, 15 × 5	S	2948
P2	74,791	12	Monomer[d]	3.3 ± 0.2[c]	Sphere, r = 3.3	M	4063
P4	35,032	72	Hexamer[e]	5.9[e]	Ring, r = 5.9	L	6374
P7	17,169	60	Dimer[f]	3.7[f]	Rod, 19 × 2.1		
P8	15,873	720	Trimer[c]	2.8 ± 0.2[c]	Sphere, r = 2.8		

[a] Calculated from the cDNA sequence (excluding N-terminal Met that is removed).
[b] Estimated copy numbers, based on Day and Mindich (1980) and the symmetry of the virion.
[c] This study.
[d] Makeyev and Bamford, 2000.
[e] Juuti et al., 1998.
[f] Juuti and Bamford, 1997.
[g] Hydrodynamic dimensions were obtained using ellipsoid, sphere, or cylinder models, and approximate formulas for hydrodynamic radius (Schmitz, 1990). Dimensions are shown by length x width, or by radius (r).
[h] GenBank accession numbers: M12921, M17462, and M17461.

Empty PC of ϕ6 can be produced in *Escherichia coli* cells by coexpressing the component proteins (Gottlieb et al., 1988). Purified recombinant PC specifically recognize the three viral plus-sense genomic precursors (s, m, and l) by their unique packaging signals at the 5′ ends (Gottlieb et al., 1994; Pirttimaa and Bamford, 2000) and package a single copy of each (Day and Mindich, 1980). The translocation of the segments into a preformed capsid is an NTP hydrolysis-dependent process (Gottlieb et al., 1991; Frilander and Bamford, 1995). The polymerase (P2) replicates (minus strand synthesis) the packaged ssRNAs into dsRNA genome segments (S, M, and L) inside the particle. These can then be used as templates for transcription (plus strand synthesis; Gottlieb et al., 1990).

The virion-derived, dsRNA-containing polymerase particle (viral core) has a more expanded structure than the empty PC particle (Butcher et al., 1997), and it has been suggested that conformational changes occur during RNA packaging and/or replication. The dsRNA-containing particles can be coated in vitro with purified NC surface protein P8, which forms a T = 13 surface lattice (Olkkonen et al., 1990, 1991; Butcher et al., 1997). NC purified from virions are infectious for host cells with exposed plasma membrane (spheroplasts; Ojala et al., 1990). The NC penetrates the host plasma membrane by an endocytic-like mechanism, which is dependent on membrane voltage and protein P8 (Romantschuk et al., 1988; Poranen et al., 1999). The viral core is then delivered into the cytoplasm, where it produces ssRNA copies from the dsRNA genome segments. The newly synthesized RNAs are exported from the polymerase complex particle and are translated on ribosomes and/or packaged into the progeny PC.

We report here that infectious ϕ6 NC can be assembled in a test tube from constituent purified proteins and nucleic acid segments. We have set up an efficient self-assembly system for the ϕ6 PC and used kinetic studies to analyze the process. The in vitro assembled PC (PC_a) are indistinguishable from the recombinant particles purified from bacteria (control particles, PC_c) as determined by sedimentation analysis, electron microscopy (EM), and enzymatic assays. Isolated, self-assembled PC_a, which have packaged and replicated the viral genome segments in vitro, are converted into infectious particles by coating them with purified NC surface shell protein P8, and the biological functionality is measured using an infectivity assay. Infectious NC are formed from PC_a in a system with no background activity. The present study demonstrates that in vitro assemblies can be introduced into the cell and can then function as cytoplasmic molecular machines, leading to the synthesis of infectious enveloped virions.

Results and Discussion

Building Blocks for the ϕ6 NC Assembly

The components of the ϕ6 NC are listed in Table 1, and some relevant properties of the proteins and the RNA molecules are described. Three of the four PC proteins, the polymerase P2, the packaging NTPase P4, and the packaging factor P7, have been purified as recombinant proteins (Juuti and Bamford, 1997; Juuti et al., 1998; Makeyev and Bamford, 2000). A purification protocol for the NC surface protein P8 has also been described (Bamford et al., 1995). In this study a method to purify the shell-forming protein P1 was developed. The previous attempts to isolate soluble P1 either from virions or by expressing cloned *gene1* have not been successful (Olkkonen and Bamford, 1987; Gottlieb et al., 1988; Paatero et al., 1998). We set up a purification protocol starting from incomplete recombinant polymerase complex particles containing proteins P1 and P4. We found that purified P1P4 can be disrupted by repeated cycles of freezing and thawing. Disruption and subsequent gel filtration to separate P1 and P4 were performed in the presence of 0.5M NaCl. The P1 concentration in the peak fraction was ~0.3 mg/ml, yielding 4.5 mg P1 per liter of culture. The apparent molecular mass of 81 ± 4 kDa, obtained by gel filtration and light scattering, was as expected for P1 monomers (85 kDa). Dynamic light scattering yielded a hydrodynamic radius (R_h) of 3.8 ± 0.4 nm, indicating that P1 is an asymmetric molecule (a prolate ellipsoid with 3:1 axial ratio). The steps in P1 purification, the other purified NC proteins (P2, P4, P7, and P8), and the purified singl-stranded genomic precursors (s, m, and l) are shown in Figure 1.

In Vitro Assembly of ϕ6 Polymerase Complex

The ϕ6 polymerase complex was assembled from the purified recombinant proteins. The assembly was initiated by addition of 6% polyethyleneglycol (PEG) 4000,

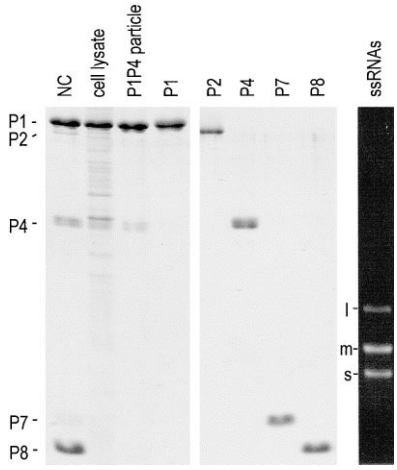


Figure 1. Purified Components for the In Vitro ϕ6 NC Assembly

Steps in protein P1 purification (*E. coli* JM109 [pLM358] lysate, purified P1P4 particle, and purified P1), purified recombinant proteins P2, P4, P7, and virion-derived P8 analyzed in SDS–PAGE (left and center). Proteins of virion-derived NC are shown on the left. An agarose gel analysis of the in vitro produced ssRNA segments s, m, and l used as genomic precursors (right).

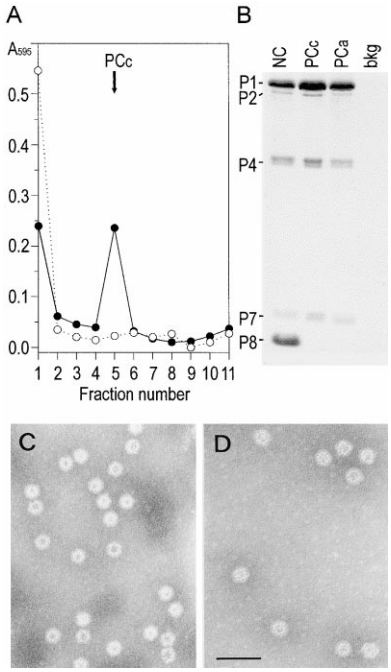


Figure 2. Sedimentation, Protein Composition, and Negative Staining EM of the In Vitro Assembled Polymerase Complex

(A) Purified recombinant proteins P1, P2, P4, and P7 were incubated with (assembly reaction; closed circles) or without (control reaction; open circles) 6% PEG 4000, 50 mM Tris-HCl (pH 8.9) prior to sedimentation. Protein contents of the gradient fractions were analyzed using the Bradford assay. The increase of the absorbance by PEG 4000 is corrected. The fractions are numbered from top to bottom, and the sedimentation position of the recombinant in vivo assembled polymerase complex is indicated by an arrow (PC_c).

(B) The protein composition of the in vitro assembled polymerase complex (PC_a) (fraction 5 from sedimentation analysis of the assembly reaction) and the in vivo assembled virion derived NC and recombinant PC (PC_c). Bkg is a background control (fraction 5 from the sedimentation analysis of the control reaction).

(C and D) Negative staining electron micrographs of the PC_c (C) and PC_a (D) particles. Bar, 100 nm.

50 mM Tris-HCl (pH 8.9) (final concentrations) to the desired mixture of proteins P1, P2, P4, and P7. Protein P2 was added as a stable monomer, and P7 was added as a stable dimer (Table 1). Protein P4 was supplemented with 5 mM $MgCl_2$, 2 mM $CaCl_2$, and 5 mM ATP during purification to maintain it as a hexamer (Juuti et al., 1998). When a polydisperse form of P4 was used (no nucleotide or divalent cations), the formation of particles was detected only after addition of ADP (or ATP) and $MgCl_2$ (data not shown), indicating that the hexameric form of P4 is crucial for in vitro assembly.

Products of the in vitro assembly reactions were analyzed by sedimentation in a sucrose gradient, and the protein content of the collected fractions was assayed (Figure 2A). After a standard assembly reaction, proteins were distributed into two fractions: the fast-sedimenting assembly product (fraction 5) and the slowly sedimenting nonassembled precursors (fraction 1, top). When the mixture of the four proteins was incubated in the absence of 6% PEG 4000, 50 mM Tris-HCl (pH 8.9) (control reaction), proteins were only found in the top fraction (Figure 2A). In either case, no aggregation was detected. The sedimentation velocity of the PC_a was similar to the PC_c, and the light scattering zone containing the PC_a was very sharp, implying high homogeneity. The particle-containing zone was collected and analyzed by SDS polyacrylamide gel electrophoresis (SDS–PAGE) and EM. SDS–PAGE analysis confirmed that the protein composition of the isolated PC_a was similar to that of PC_c (Figure 2B); the four proteins P1, P2, P4, and P7 were present in the same relative proportions. Finally, the EM experiments demonstrated that the isolated PC_a particles were indistinguishable from the PC_c particles (Figures 2C and 2D).

In Vitro Assembly Pathway for ϕ6 Polymerase Complex

Previously, a set of incomplete recombinant polymerase particles (missing protein P2, P4, or P7, or both proteins P2 and P7) has been produced by expression in *E. coli* (Gottlieb et al., 1988, 1990; Paatero et al., 1998). Although this has enabled the purification and structural and functional characterization of the assembly products (Juuti and Bamford, 1995; de Haas et al., 1999), it provided limited insight into the initiation and control of the assembly process. The in vitro assembly system presented here allowed us to study the role of the individual proteins in the assembly process.

The formation of particles was monitored by following the increase in turbidity in the protein mixture after initiation of the assembly reaction (Figure 3A). All the tested combinations contained the shell-forming protein P1, as proteins P2, P4, and P7 did not self-assemble without P1 (data not shown). Reactions containing only P1, or P1 with P2 or P7, or all three proteins did not show a detectable increase in light scattering (Figure 3A). When the P4 hexamer was included, a substantial increase was observed (Figure 3A; P1+P4, P1+P2+P4, P1+P4+P7, and P1+P2+P4+P7). Thus, P4 is required to initiate the assembly in vitro. Although the requirement of P4 in the initiation can be bypassed by overexpressing P1, P2, and P7, or P1 alone in *E. coli*, the yields

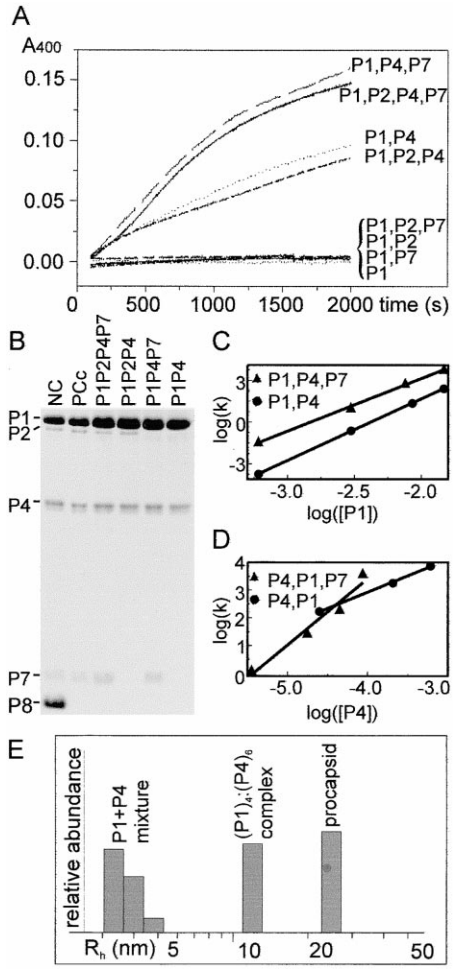


Figure 3. In Vitro Assembly of Incomplete Polymerase Complexes

(A) Increase of light scattering in the solution of polymerase complex proteins after initiation of the assembly reaction. The protein components in each reaction are indicated. The mixing time was 100 s.

(B) Protein composition of the in vitro assembled incomplete polymerase complex particles, P1P2P4, P1P4P7, and P1P4, after separation by rate zonal centrifugation. The complete in vitro assembled PC_a (P1P2P4P7) and the in vivo assembled NC and PC_c are included as controls.

(C) P1 concentration dependence of the initial assembly rate. Logarithms of the steepest slope of the kinetic curves ([A] and additional data not shown) were plotted against logarithm of P1 concentration. Concentrations of P4 and P7 were 0.2 and 0.1 mg/ml, respectively. The excess of the latter constituents over P1 was large enough to assume their concentration was approximately constant during the initial kinetics.

(D) P4 concentration dependence of the initial rate. Logarithms of the steepest slope of the kinetic curves ([A] and additional data not shown) were plotted against logarithm of P4 concentration. Concentrations of P1 and P7 were 0.18 and 0.15 mg/ml, respectively.

(E) Dynamic light scattering of P1 (0.4 mg/ml) -P4 (0.2 mg/ml) mixture was obtained 4 min after assembly initiation with 4% PEG 4000 at 20°C. The graph shows the apparent distribution of hydrodynamic radii (R_h). The R_h values were corrected for PEG viscosity.

of stable, soluble, particles are low (Paatero et al., 1998). As protein P4 initiates assembly in vitro at low protein concentrations, it probably also does so during normal infection.

The effect of minor protein constituents, P2 and P7, on the assembly kinetics was also examined. Both proteins were stoichiometrically incorporated into the assemblies (Figure 3B), but their effect on the assembly kinetics was different. The polymerase P2 did not have a significant effect on kinetics or yield (Figure 3A, compare P1+P4 vs. P1+P2+P4 and P1+P4+P7 vs. P1+P2+P4+P7). On the other hand, the presence of P7 doubled the assembly rate (Figure 3A; P1+P4+P7 and P1+P2+P4+P7). This suggested that P7 is not only a packaging factor (Juuti and Bamford, 1995, 1997) but acts also as an assembly cofactor, albeit a nonessential one.

Analysis of the assembly reaction products by sedimentation and SDS–PAGE (Figure 3B) confirmed that the turbidity of the reaction mixture was proportional to the fraction of assembled particles and that proteins were incorporated into the assembled particles in native-like ratios. Despite the difference in kinetics, the final amount of assembled particles (after 90 min) was approximately the same in all particle-producing reactions (Figure 3B).

The overall kinetics of the assembly reactions (Figure 3A) suggest that the assembly reaction proceeds relatively rapidly after passage of the slow nucleation step. Thus, the initial assembly rates are indicative of the kinetic order of the nucleation step and provide information about the composition of the rate-limiting nucleation complex (Prevelige et al., 1993). We analyzed this rate-limiting step using kinetic methods combined with systematic variation of protein concentrations (Figures 3C and 3D).

Variation of P1 concentration in the presence of excess of P4 revealed an apparent fourth-order P1 concentration dependence of the initial rate of assembly (Figure 3C). The slope of the log-log plot varied between 3.7 and 4.3 in different experiments, ruling out both the third and fifth reaction order. An excess of P7 accelerated the assembly, but the kinetic order of P1 remained the same. The fourth order P1 concentration dependence suggested that the P1 tetramer, $(P1)_4$, is a critical part of the nucleation complex. Because the rate increased in the presence of P7, the formation of this complex may be stabilized by the dimeric P7 as depicted in the proposed model (Figure 6).

As P4 is required for the in vitro assembly, the nucleation complex must also contain P4. Similar analyses of the P4 concentration dependence in the presence of an excess of P1 and P7 (Figure 3D) revealed that P7 has a significant effect on the apparent kinetic order of P4. In the absence of P7, the initial rate was proportional to the P4 concentration (slope 1.17 $\pm$ 0.05, Figure 3D), which is consistent with a nucleation complex $(P1)_4$:$(P4)_6$. On the other hand, in the presence of P7, the reaction was apparently second order in P4 (slope 2.4 $\pm$ 0.3, Figure 3D), indicating a $(P1)_4$:$(P7)_2$:$(P4)_{12}$ nucleation complex. The most likely explanation is that the $(P1)_4$ forms more readily when stabilized by $(P7)_2$ and the subsequent association with the second $(P4)_6$ becomes the rate-limiting step.

Further insight into the nature of the putative nucleation complex was gained from time-resolved dynamic light scattering. Proteins P1, P4, and P7 do not seem to interact in the absence of PEG. The addition of 3%–4% PEG to the P1-P4 mixture promoted the forma-

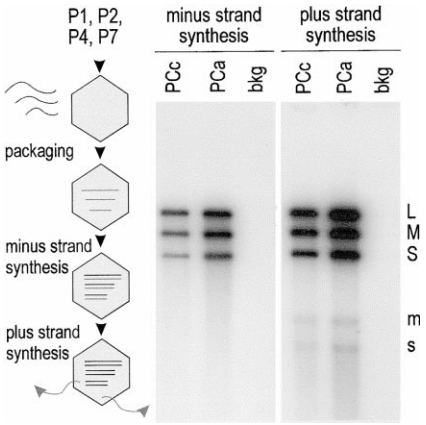


Figure 4. RNA Synthesis Activity of the In Vitro Assembled Polymerase Complex

An autoradiogram of an agarose gel showing the RNA synthesis activity of the in vitro (PC_a) and the in vivo (PC_c) assembled polymerase complexes. Bkg is a background control. The mobility of the ϕ6 dsRNA segments (capital letters) and ssRNA transcripts (lower case letters) are indicated on the right and the flow of reactions in the in vitro RNA replication cycle is illustrated schematically on the left: packaging of the three single-stranded genomic precursors into preassembled isolated PC, replication of the packaged segments into dsRNA genome segments, and transcription of segments S and M. The relatively low plus strand synthesis activity of PC_a compared to the amount of dsRNA segments originate from the rather high carry-over of sucrose and NaCl (11% (w/v) and 84 mM, respectively) from the sucrose gradient peak of PC_a to the RNA synthesis reaction. The concentrations of these components were compensated in the control reactions carried out with PC_c.

tion of larger complexes with apparent R_h ~12 nm (Figure 3E). Similarly, P1 and P7 formed a complex with R_h ~9 nm (data not shown). These structures probably correspond to the kinetically detected nucleation complexes, which contain the $(P1)_4$. The relatively large hydrodynamic size of these assemblies apparently results from the highly elongated and flat shape (as shown schematically in Figure 6). In the case of the P1-P4 mixture, the observed complexes were converted to the final product (R_h 25 to 30 nm) without any detectable larger intermediates (Figure 3E). Therefore, assembly proceeds via rapid addition of P1 and P4 after the nucleation step.

The kinetic analyses described above provide molecular insights into the assembly pathway of the ϕ6 polymerase complex. Interestingly, the $(P4)_6$ is an example of a ring-shaped genome packaging machinery (portal protein) that introduces a symmetry mismatch into the five-fold vertex of the icosahedral capsid. Genetic studies on several dsDNA bacterial viruses have postulated that such portal assemblies direct the correct assembly of the procapsid and are involved in assembly initiation (reviewed by Valpuesta and Carrascosa, 1994). The current in vitro assembly systems of dsDNA viruses (Prevelige et al., 1988, 1993; Cerritelli and Studier, 1996; Newcomb et al., 1999; Wang et al., 2000) have not, however, yielded any further insight into the mechanism of how the packaging machinery and the symmetry mismatch are incorporated into the highly symmetrical viral capsid. Our results show that P4 is needed for the nucleation of the assembly, and we propose that the role of $(P4)_6$ in nucleation is most likely the stabilization of the P1 dimer (Figure 6). In the nucleation complex, P1 and P4 oligomers are held together via multiple weak interactions, and the complex is stabilized by the elongated $(P7)_2$ acting presumably at a site of two-fold symmetry. We propose that the critical step for the assembly is the formation of the two-fold symmetry axis between two P1 dimers (Figure 6), not the formation of complete P1 decamer around the $(P4)_6$. The latter has been proposed for bluetongue virus (Grimes et al., 1998) but seems unlikely to apply to ϕ6. Because P1 does not dimerize readily the formation of five P1 dimers around the $(P4)_6$ would be less favorable than formation of only two assisted by favorable interactions with P4 and P7. The symmetry mismatch at the five fold is built up as the capsid grows and most likely results from a steric hindrance during completion of the P1 decamer around each $(P4)_6$. As P2 does not have an effect on the kinetics, it is most likely incorporated into the capsid after the initial nucleation step (Figure 6).

RNA Synthesis Activity of the In Vitro Assembled Particles

After successful in vitro assembly of ϕ6 polymerase complex particles, we analyzed whether the in vitro formed particles are enzymatically active using the minus (combined ssRNA packaging and replication) and plus strand synthesis (combined ssRNA packaging, replication, and transcription) reactions. These reactions were carried out using PC_a isolated by centrifugation in a sucrose gradient. As a background control (bkg) we used the same fraction from a gradient of the control reaction (loaded with P1, P2, P4, and P7 incubated without PEG 4000). The minus strand synthesis reactions were treated with the dsRNA-specific V1 nuclease to confirm that the replicated RNA was actually packaged inside the particle and was thus protected. Synthesized radioactively labeled RNase-resistant dsRNA segments from the minus strand synthesis reaction and dsRNA segments and transcripts from the plus strand synthesis reactions were analyzed in an agarose gel (Figure 4). The isolated in vitro assembled particles were enzymatically active in the in vitro RNA synthesis reactions. Thus, they were able to bind and package the ssRNA genomic precursors (s, m, and l), replicate the packaged segments into dsRNA form (S, M, and L; Figure 6), and produce new transcripts from S and M segments. (Transcripts are not produced from L in standard in vitro conditions; van Dijk et al., 1995). Interestingly, the replication and transcription activities of PC_a particles were somewhat higher than those of PC_c in reactions containing equal amounts of P1 (Figure 4).

In Vitro Assembly of Infectious ϕ6 NC

The most sensitive parameter for correct assembly is the production of infectious virions, since the intermediates might be abortive or aberrant products of the reaction. For this purpose, the dsRNA-containing PC_a particles were coated with purified NC surface protein P8 (Olkkonen et al., 1990), as only P8-coated particles are infectious for host cell spheroplasts (Olkkonen et al., 1991). The P8 shell assembly reaction mixture was used to

IC/ml			
NC(virion)	NC(PC_c)	NC(PC_a)	bkg
$(1.1\pm0.4)\times10^9$	$(8.3\pm3.2)\times10^6$	$(2.6\pm0.8)\times10^6$	–

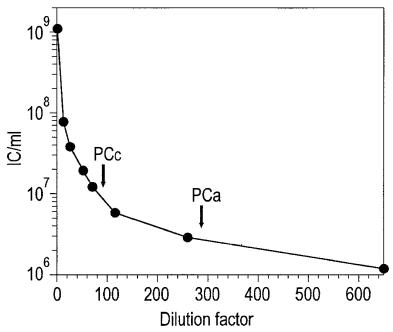


Figure 5. Infectivity of Virion-Derived NC, NC(virion), and the In Vitro Assembled NC Originating Either from In Vitro Assembled Polymerase Complex, NC(PC_a), or from Recombinant Polymerase Complex, NC(PC_c)

Spheroplasts were infected with P8 assembly reaction mixture, which originated from 0.2 μg of empty PC or with approximately the same particle amount of NC(virion). The table shows results from several repeated experiments. The IC counts presented are per milliliter of particle preparation. Bkg is a background control originating from sucrose gradient loaded with P1, P2, P4, and P7. The curve below shows IC yields obtained with different NC dilutions, and the arrows point where the IC yields are the same as the IC counts obtained with NC(PC_a) or NC(PC_c).

infect spheroplasts, and infective centers (IC) were counted (Figure 5). A maximum of 2.6×10^6 IC/ml of P8 assembly reaction was obtained. As a control, we used NC isolated from purified virions as well as PC_c, which were packaged and coated in vitro (Figure 5). The number of particles was the same in infections carried out with NC of different origins. To confirm that the obtained IC were not contaminating viral particles, all steps preceding the P8 assembly reaction, as well as the P8 assembly reactions originating from the background control, were analyzed in the infection assay. None of these controls produced any IC.

Assembly Efficiency

In any assembly system, only a portion of the proteins or nucleic acid substrates can be converted to assembly intermediates or final structures due to the reduction of the reactants as the assembly proceeds. The efficiency of assembly refers to the number of functional assemblies that are produced from the original pool of reactants. From the in vitro reconstituted TMV virions, approximately one-tenth appeared as normal TMV in EM, and about 3% of the correct size virions had the infectivity of the control TMV particles (Fraenkel-Conrat and Williams, 1955), leading to an efficiency of ~0.3%.

Based on previously presented data and that presented here, we estimated the efficiencies of the steps in the in vitro assembly pathway for ϕ6 NC; formation of $(P4)_6$, assembly of empty PC, ssRNA packaging into empty PC, production of complementary minus strands to the packaged ssRNAs, and P8 shell assembly on the dsRNA-filled particles. The assembly pathway is illus-

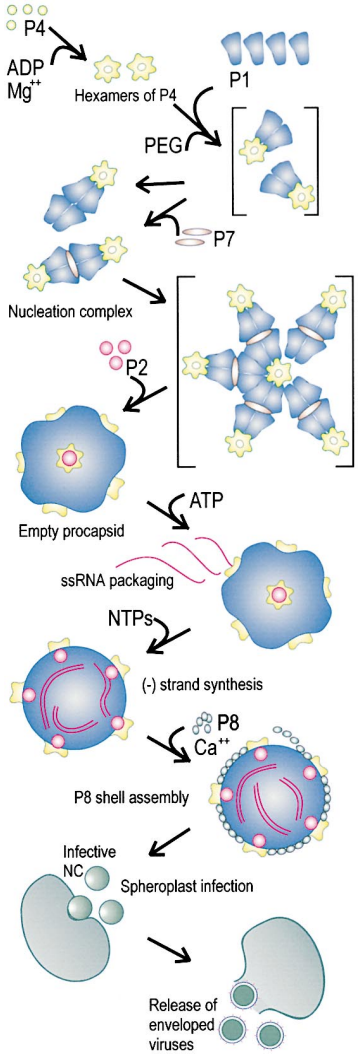


Figure 6. Steps in the ϕ6 NC In Vitro Assembly Pathway

$P4$ hexamers are assembled in the presence of Mg^{2+} ions and ADP (or ATP). Empty PC are formed from purified recombinant proteins P1, P2, P4, and P7 when incubated with PEG 4000. P4 and P1 nucleate the PC assembly. One $(P4)_6$ probably first induces the dimerization of P1 (shown in brackets). The actual nucleation complex is composed of four P1 molecules and either one or two $(P4)_6$. The number of $(P4)_6$ in the nucleation complex depends on the presence of $(P7)_2$ in the reaction. In the absence of P7, $(P1)_4$:$(P4)_6$ complex nucleates the assembly, but when P7 is present the rate-limiting step is the formation of the $(P1)_4$:$(P7)_2$:$(P4)_{12}$ structure. Thus, P7 probably accelerates the formation of the $(P1)_4$:$(P4)_6$ assemblies, and thus the incorporation of the second $(P4)_6$ into this complex becomes rate limiting. After the nucleation, the reaction proceeds rapidly to a complete capsid. A hypothetical assembly intermediate, depicting how the symmetry mismatch could be built up, is shown in brackets. P2 is incorporated into the capsid during the capsid growth.

The complete in vitro assembled PC packages the three plus-sense genomic precursors, s, m, and l, if energy is provided in a form of a NTP (for example, ATP). The polymerase is activated when all four NTPs are present and complementary minus strands are synthesized to the packaged plus strands. The RNA (S, M, and L segments) filled particles can be coated with NC surface shell protein P8 in the presence of Ca^{2+} ions. The in vitro assembled NC productively infects host cell spheroplasts, and enveloped infectious virions are released.

trated in Figure 6, and the specific requirements for each step are indicated.

The assembly of $(P4)_6$, as well as complete PC, is rather efficient. Purified protein P4 is isolated as a stable hexamer after addition of $MgCl_2$, $CaCl_2$, and ATP (Juuti et al., 1998). When $(P1)_1$, $(P2)_1$, $(P4)_6$, and $(P7)_2$ are incubated in the in vitro assembly conditions, approximately 50% of the proteins are assembled into particles, as was concluded from the sedimentation analysis shown in Figure 2.

The overall efficiency of the in vitro assembly pathway from PC to NC was measured in this study by infectivity. The in vitro reaction products were compared to the different dilutions of virion-derived NC (Figure 5). The IC counts originating from PC_a corresponded to IC obtained with a $^1/_{280}$ dilution of NC; IC originating from PC_c corresponded to a $^1/_{90}$ dilution of NC. Thus, the overall efficiency of the assembly steps from empty PC to infective NC can be judged to be $\sim$0.35% for PC_a and $\sim$1.1% for PC_c.

The overall efficiency depends on packaging, minus strand synthesis, and P8 shell assembly reactions. Based on the data presented previously by Frilander and Bamford (1995), it was calculated that about 2%–5% of PC particles packaged all the three genomic ssRNA precursors in vitro. By comparing the dsRNA produced by PC (in the combined in vitro RNA packaging, minus, and plus strand synthesis reaction) to the dsRNA amounts in the virion-derived NC samples, we estimated that approximately 5% of the PC particles were able to both package and replicate the three genomic precursors into double-stranded form in vitro (data not shown). This is in good agreement with the calculated proportion of packaging competent particles (2%–5%) and indicates that the ssRNA packaging, rather than the minus strand synthesis, is the limiting step in the formation of dsRNA genome-containing particle. The P8 shell is required to make the dsRNA-filled polymerase complex infectious. Earlier studies of in vitro assembly of P8 on virion-derived viral cores (polymerase complexes containing the dsRNA genome) showed that 5%–10% of the original infectivity of these particles could be restored (Olkkonen et al., 1991). Thus, the critical steps in the in vitro assembly pathway (from empty PC to infectious NC) are ssRNA packaging (2%–5%) and P8 shell assembly (5%–10%), leading to a theoretical estimation for overall efficiency (from PC to NC) of 0.1% to 0.5%. The overall efficiency obtained for PC_a fits this range (0.35%), thus indicating that the produced assemblies are functionally close to the in vivo controls (PC_c- and virion-derived core).

Conclusions

Infectious viral particles were assembled in defined conditions from five protein species and three segments of ssRNA. Four of these proteins were first assembled into empty PC. The absence of aberrant products and the protein composition, sedimentation, and morphology identical to the in vivo control all suggest high specificity of this in vitro assembly reaction. The assembled PC were active molecular machines capable of ssRNA packaging, replication, and transcription, producing particles with three viral dsRNA genome segments. These particles were coated with a shell of the fifth viral protein, the NC surface protein, and the in vitro generated particles were shown to productively infect their host.

The ϕ6 in vitro assembly system described here is significantly more complex than any of the previously studied viral assembly systems using purified single-protein starting material. Simple viruses containing one protein species and a single genome segment have been successfully assembled in vitro (TMV and CCMV), but the established assembly systems for more complex viruses either start from preformed assembly intermediates or do not lead to biologically functional particles. The ϕ6 NC contains about 1000 protein molecules and three segments of dsRNA (Table 1) and is a functional molecular machine capable of ssRNA packaging, replication, and transcription (Figure 4).

The infectivity (Figure 5) of the in vitro assembled ϕ6 NC showed that the structure is not only functional in vitro but also in vivo. The surface shell of the NC (protein P8) can interact with the host plasma membrane and trigger an endocytic-like uptake of the particle. This process is probably driven by the viral components, as bacterial cells do not possess endocytic mechanisms in general. The polymerase complex subsequently released to the cytoplasm is capable of producing new viral mRNAs from the dsRNA genome segments residing inside the particle. This demonstrates that complex multienzyme assemblies formed in the absence of cellular components, in defined in vitro conditions, can act as functional cytoplasmic molecular machines.

The established in vitro assembly system for ϕ6 PC allows us to delineate the role of individual proteins in the assembly process (Figure 6). The nucleation of the PC assembly by a P4 hexamer reconstitutes an active packaging vertex into a virus and thus provides a system to study the assembly and packaging of a viral procapsid in defined conditions. The kinetic analyses allowed us to identify P7 as an assembly cofactor, which accelerates the assembly by stabilizing the nucleation complex. On the other hand, the polymerase protein P2 did not have any effect on the kinetics, indicating a postnucleation incorporation into the particle. This work also gave interesting insight into the assembly of the 120-subunit capsid lattice through the formation of a tetrameric P1 complex and delineated how a symmetry mismatch might be incorporated into a five-fold vertex of a viral capsid via a two-fold centered nucleation complex.

Experimental Procedures

Bacterial Strains and Plasmids

P. syringae pathovar phaseolicola HB10Y (Vidaver et al., 1973) was used for ϕ6 (Vidaver et al., 1973) propagation. Plasmids pLM687 (Mindich et al., 1994) and pLM358 (Gottlieb et al., 1990) were used for production of empty PC and P1P4 particles, respectively. Plasmids pEM2 (Makeyev and Bamford, 2000) and pJTJ7 (Ojala et al., 1993) encode ϕ6 proteins P2 and P4, respectively.

A new plasmid, pEM7, was constructed for P7 production. The coding sequence for ϕ6 *gene7* was amplified by PCR from the plasmid pLM687 using Pfu DNA polymerase (Stratagene) and primers 5′-CTTTTATAAGGACGGTGTCATATGACTTTGTACCT-3′ and 5′-CTC TCGAATTCATCAGGCGCTTACCT-3′. The PCR fragment was digested with NdeI and EcoRI and ligated with NdeI-EcoRI cut vector pT7–7 (Tabor, 1990).

Isolation of Particles

For NC production and protein P8 purification, bacteriophage ϕ6 was grown on HB10Y and purified as described previously (Bamford et al., 1995). For infectivity assays, NC were further purified in 5%–20% (w/v) sucrose gradient in 20 mM Tris-HCl (pH 7.4), 150 mM NaCl (Sorvall SW41 rotor, 30,000 rpm, for 50 min, +15°C). For ssRNA production, NC were purified as described by Pirttimaa and Bamford (2000).

Recombinant empty PC were produced and purified using Triton X-114 extraction and sedimentation through a 5%–20% (w/v) sucrose gradient in 20 mM Tris-HCl (pH 8), 150 mM NaCl (Frilander and Bamford, 1995). The particle-containing zone was collected and stored at −80°C.

Purification of ϕ6 NC Proteins P1, P2, P4, P7, and P8

P1P4 particles were purified as PC, except that the centrifugation was for 2 hr 10 min. The NaCl concentration in the sucrose gradient zone containing the particles was increased from 0.15 to 0.5 M. The particles were then disrupted by three repeated cycles of freezing (−80°C) and thawing (room temperature, RT), and the material was loaded onto a HiLoad 26/60 Superdex 200 column (Pharmacia) equilibrated with 20 mM Tris-HCl (pH 8), 0.5 M NaCl, 0.5 mM EDTA. Gel filtration fractions containing purified P1 were stored at −80°C.

The purification protocols for P2, P4, and P7 have been described in detail previously (Juuti and Bamford, 1997; Juuti et al., 1998; Makeyev and Bamford, 2000). To obtain better yields of P4 (10 mg protein/l culture), the following buffer was used through out the purification: 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 0.1 mM ATP, 5 mM $MgCl_2$, 2 mM $CaCl_2$. Protein P8 was purified from ϕ6 virions (Bamford et al., 1995). Stocks of P2 and P7 were stored on ice, and P4 and P8 were stored at −80°C.

Production and Purification of Viral Transcripts

The single-stranded genomic precursors for RNA synthesis reactions were made by activating isolated NC to produce transcripts according to Pirttimaa and Bamford (2000). The resulting RNAs were extracted with phenol:chloroform:isoamyl alcohol (25:24:1, v/v) and chloroform:isoamyl alcohol (24:1, v/v). Subsequently, the genomic dsRNA was precipitated with 2.8 M LiCl, and the ssRNA transcripts were precipitated from the resulting supernatant with ammonium acetate and ethanol.

Polymerase Complex Assembly

The four purified recombinant proteins P1 (in 20 mM Tris-HCl [pH 8] 0.5 M NaCl, 0.5 mM EDTA), P2 (in 50 mM Tris-HCl [pH 8], 90 mM NaCl, 0.1 mM EDTA), P4 (in 20 mM Tris-HCl [pH 7.5], 175 mM NaCl, 5 mM ATP, 5 mM $MgCl_2$, 2 mM $CaCl_2$), and P7 (in 20 mM Tris-HCl [pH 7.5], 250 mM NaCl) were mixed in the proportions found in the ϕ6 virion (70:6:17:7) so that the final P1 concentration was about 0.2 mg/ml. The in vitro assembly reaction was initiated by the addition of 50 mM Tris-HCl (pH 8.9), 6% PEG 4000 to the protein mixture. The sedimentation analysis was carried out after 90 min incubation at RT in 10%–30% (w/v) sucrose gradient made in 20 mM Tris-HCl (pH 8), 150 mM NaCl (Sorvall SW41 rotor, 35,000 rpm, for 80 min, +10°C). The gradient was fractionated using a BioComp gradient fractionator. The PC_a fraction was used for enzymatic analyses, infectivity assays, and negative staining EM.

Light scattering experiments were performed using a V-560 UV/VIS Spectrophotometer (JASCO) set at 22°C. Light scattering was recorded at 350 nm (2.0 nm band width) in 2 or 5 s time increments for 2000–8000 s. The P1 concentration in the assembly reaction was varied between 16 and 160 μg/ml. P4 concentration dependence was measured in the range 4 to 18 μg/ml in the presence of an approximately 4-fold molar excess of P1 (180 μg/ml). The total NaCl concentration in the in vitro assembly reaction varied from 250 to 400 mM.

RNA Polymerization Reaction Conditions

Minus and plus strand synthesis assays were performed under optimal conditions determined earlier (van Dijk et al., 1995). Each reaction contained approximately 0.5 μg of isolated PC in a 25 μl reaction volume. After 90 min reaction time, the minus strand synthesis reactions were treated with a dsRNA-specific cobra venom V1 nuclease (0.2 U/reaction) for 15 min at 30°C. The same amount of nuclease digests dsRNA isolated from 2.5 μg of ϕ6 NC.

In Vitro Assembly of ϕ6 NC

The in vivo and in vitro assembled empty PC were filled with the genomic dsRNA segments using the standard in vitro plus strand synthesis conditions. Subsequently the excess unpackaged RNA was removed with RNase A (0.1 μg/μl final concentration) treatment for 15 min at RT. The NC surface protein P8 was then assembled on the RNA-filled particles in a 21 μl reaction volume containing 0.3 μg of PC from RNase A treated RNA synthesis reaction and 1 μg of protein P8 (in 10 mM Tris-HCl [pH 8], 300 mM NaCl, 0.1 mM EDTA). The reactions were supplemented with 1 mM $CaCl_2$, 110 mM NaCl, 0.5 mM EDTA (final concentrations), and were allowed to proceed for 0.5 to 1 hr at RT and immediately used for the infectivity assay.

Preparation of Host Cell Spheroplasts and Spheroplast Infection

Host cell spheroplasts were prepared as previously described (Ojala et al., 1990; Poranen et al., 1999). Spheroplasts were infected with virion-derived NC (in 1 mM $CaCl_2$, 130 mM NaCl, 0.5 mM EDTA, 2% sucrose) or with the in vitro assembled particles for 1 hr at RT. The number of IC was determined on HB10Y lawns. The infection mixture contained 40 μl spheroplasts and 0.47 μg of virion-derived NC, or P8 assembly reaction mixture originating from 0.2 μg of empty PC. As 3/7 of the NC mass is made up by the PC proteins and the rest by the P8 shell, the amount of particles was approximately the same in all infections.

Analytical Methods

Protein concentrations were determinated with Coomassie brilliant blue using bovine serum albumin (BSA) as a standard (Bradford, 1976). EM was performed using a JOEL 1200 EX electron microscope (at the EM unit, Institute of Biotechnology, University of Helsinki) operating at 60 kV. Particles, dialysed against 20 mM Tris-HCl (pH 8), 150 mM NaCl for 90 min at RT, were negatively stained with 1% (w/v) ammonium molybdate (pH 7.2). Standards used to estimate protein mass in light scattering and gel filtration experiments were BSA (67 kDa) and aldolase (158 kDa). Hydrodynamic properties and interactions of individual proteins as well as characterization of assembly intermediates were performed using PDDLS Batch dynamic light scattering instrument equipped with deconvolution software (Precision Detectors) for correlation function analysis.

Quantitation of the In Vitro Assembly Process

The efficiency of the in vitro PC assembly reaction was determined from the rate zonal gradient fractions by measuring the protein content of fractions containing the assembly product as well as those containing the nonassembled material. To estimate the in vitro RNA packaging and replication efficiency, plus strand synthesis reaction products and different dilutions of NC were loaded onto an agarose gel. After electrophoresis, the dsRNA amount produced in plus strand synthesis reaction was compared to the dsRNA amounts in the NC samples. The amount of polymerase complex particles and NC particles in the samples was estimated by SDS–PAGE by comparing the amounts of P1 in the samples.

Acknowledgments

M.Sc. Eugene Makeyev is thanked for critical reading of the manuscript, for construction of plasmid pEM7, and for donating purified proteins P2 and P4. We are grateful to Dr. Sarah Butcher for providing purified P7, and we wish to thank M.Sc. Markus Pirttimaa for the NC preparation used in ssRNA production. The technical assistance of Marja-Leena Perälä is greatly appreciated. This investigation was supported by research grants 62993, 68694, and 64298 (Finnish Centre of Excellence Program [2000-2005]) from the Academy of Finland.

Received February 14, 2000; revised February 5, 2001.

References

Bamford, D.H. (2000). Virus structures: those magnificent molecular machines. Curr. Biol. *10*, R558–R561.

Bamford, D.H., Ojala, P.M., Frilander, M., Walin, L., and Bamford, J.K.H. (1995). Isolation, purification, and function of assembly intermediates and subviral particles of bacteriophages PRD1 and ϕ6. In Methods in Molecular Genetics, K.W. Adolph, ed. (San Diego: Academic Press), pp. 455–74.

Bancroft, J.B., and Hiebert, E. (1967). Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus. Virology *32*, 354–356.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye-binding. Anal. Biochem. *72*, 248–254.

Butcher, S.J., Dokland, T., Ojala, P.M., Bamford, D.H., and Fuller, S.D. (1997). Intermediates in the assembly pathway of the double-stranded RNA virus ϕ6. EMBO J. *16*, 4477–4487.

Cerritelli, M.E., and Studier, F.W. (1996). Assembly of T7 capsids from independently expressed and purified head protein and scaffolding protein. J. Mol. Biol. *258*, 286–298.

Cheng, R.H., Caston, J.R., Wang, G.J., Gu, F., Smith, T.J., Baker, T.S., Bozarth, R.F., Trus, B.L., Cheng, N., Wickner, R.B., et al. (1994). Fungal virus capsids, cytoplasmic compartments for the replication of double-stranded RNA, formed as icosahedral shells of asymmetric Gag dimers. J. Mol. Biol. *244*, 255–258.

Day, L.A., and Mindich, L. (1980). The molecular weight of bacteriophage ϕ6 and its nucleocapsid. Virology *103*, 376–385.

de Haas, F., Paatero, A.O., Mindich, L., Bamford, D.H., and Fuller, S.D. (1999). A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage ϕ6. J. Mol. Biol. *294*, 357–372.

Dube, P., Tavares, P., Lurz, R., and van Heel, M. (1993). The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. EMBO J. *12*, 1303–1309.

Fraenkel-Conrat, H., and Williams, R.C. (1955). Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. Proc. Natl. Acad. Sci. USA *41*, 690–698.

Frilander, M., and Bamford, D.H. (1995). In vitro packaging of the single-stranded RNA genomic precursors of the segmented double-stranded RNA bacteriophage ϕ6: the three segments modulate each other's packaging efficiency. J. Mol. Biol. *246*, 418–428.

Gottlieb, P., Strassman, J., Bamford, D.H., and Mindich, L. (1988). Production of a polyhedral particle in *Escherichia coli* from a cDNA copy of the large genomic segment of bacteriophage ϕ6. J. Virol. *62*, 181–187.

Gottlieb, P., Strassman, J., Qiao, X.Y., Frucht, A., and Mindich, L. (1990). In vitro replication, packaging, and transcription of the segmented double-stranded RNA genome of bacteriophage ϕ6: studies with procapsids assembled from plasmid-encoded proteins. J. Bacteriol. *172*, 5774–5782.

Gottlieb, P., Strassman, J., Frucht, A., Qiao, X.Y., and Mindich, L. (1991). In vitro packaging of the bacteriophage ϕ6 ssRNA genomic precursors. Virology *181*, 589–594.

Gottlieb, P., Strassman, J., and Mindich, L. (1992). Protein P4 of the bacteriophage ϕ6 procapsid has a nucleoside triphosphate-binding site with associated nucleoside triphosphate phosphohydrolase activity. J. Virol. *66*, 6220–6222.

Gottlieb, P., Qiao, X., Strassman, J., Frilander, M., and Mindich, L. (1994). Identification of the packaging regions within the genomic RNA segments of bacteriophage ϕ6. Virology *200*, 42–47.

Grimes, J.M., Burroughs, J.N., Gouet, P., Diprose, J.M., Malby, R., Zientara, S., Mertens, P.P., and Stuart, D.I. (1998). The atomic structure of the bluetongue virus core. Nature *395*, 470–478.

Guo, P., Grimes, S., and Anderson, D. (1986). A defined system for in vitro packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage ϕ29. Proc. Natl. Acad. Sci. USA *83*, 3505–3509.

Hendrix, R.W. (1978). Symmetry mismatch and DNA packaging in large bacteriophages. Proc. Natl. Acad. Sci. USA *75*, 4779–4783.

Hwang, Y., and Feiss, M. (1995). A defined system for in vitro λ DNA packaging. Virology *211*, 367–376.

Juuti, J.T., and Bamford, D.H. (1995). RNA binding, packaging and polymerase activities of the different incomplete polymerase complex particles of dsRNA bacteriophage ϕ6. J. Mol. Biol. *249*, 545–554.

Juuti, J.T., and Bamford, D.H. (1997). Protein P7 of phage ϕ6 RNA polymerase complex, acquiring of RNA packaging activity by in vitro assembly of the purified protein onto deficient particles. J. Mol. Biol. *266*, 891–900.

Juuti, J.T., Bamford, D.H., Tuma, R., and Thomas, G.J., Jr. (1998). Structure and NTPase activity of the RNA-translocating protein (P4) of bacteriophage ϕ6. J. Mol. Biol. *279*, 347–359.

Makeyev, E.V., and Bamford, D.H. (2000). Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage ϕ6. EMBO J. *19*, 124–133.

Mindich, L., and Bamford, D.H. (1988). Lipid-containing bacteriophages. In The Bacteriophages, R. Calendar, ed. (New York: Plenum Publishing Corporation), pp. 475–519.

Mindich, L., Qiao, X., Onodera, S., Gottlieb, P., and Frilander, M. (1994). RNA structural requirements for stability and minus-strand synthesis in the dsRNA bacteriophage ϕ6. Virology *202*, 258–263.

Newcomb, W.W., Homa, F.L., Thomsen, D.R., Trus, B.L., Cheng, N., Steven, A., Booy, F., and Brown, J.C. (1999). Assembly of the herpes simplex virus procapsid from purified components and identification of small complexes containing the major capsid and scaffolding proteins. J. Virol. *73*, 4239–4250.

Nomura, M. (1973). Assembly of bacterial ribosomes. Science *179*, 864–873.

Ojala, P.M., Romantschuk, M., and Bamford, D.H. (1990). Purified ϕ6 nucleocapsids are capable of productive infection of host cells with partially disrupted outer membranes. Virology *178*, 364–372.

Ojala, P.M., Juuti, J.T., and Bamford, D.H. (1993). Protein P4 of double-stranded RNA bacteriophage ϕ6 is accessible on the nucleocapsid surface: epitope mapping and orientation of the protein. J. Virol. *67*, 2879–2886.

Olkkonen, V.M., and Bamford, D.H. (1987). The nucleocapsid of the lipid-containing double-stranded RNA bacteriophage ϕ6 contains a protein skeleton consisting of a single polypeptide species. J. Virol. *61*, 2362–2367.

Olkkonen, V.M., Gottlieb, P., Strassman, J., Qiao, X.Y., Bamford, D.H., and Mindich, L. (1990). In vitro assembly of infectious nucleocapsids of bacteriophage ϕ6: formation of a recombinant double-stranded RNA virus. Proc. Natl. Acad. Sci. USA *87*, 9173–9177.

Olkkonen, V.M., Ojala, P.M., and Bamford, D.H. (1991). Generation of infectious nucleocapsids by in vitro assembly of the shell protein on to the polymerase complex of the dsRNA bacteriophage ϕ6. J. Mol. Biol. *218*, 569–581.

Paatero, A.O., Mindich, L., and Bamford, D.H. (1998). Mutational analysis of the role of nucleoside triphosphatase P4 in the assembly of the RNA polymerase complex of bacteriophage ϕ6. J. Virol. *72*, 10058–10065.

Pirttimaa, M.J., and Bamford, D.H. (2000). RNA secondary structures of the bacteriophage ϕ6 packaging regions. RNA *6*, 880–889.

Poranen, M.M., Daugelavicius, R., Ojala, P.M., Hess, M.W., and Bamford, D.H. (1999). A novel virus-host cell membrane interaction. Membrane voltage -dependent endocytic-like entry of bacteriophage ϕ6 nucleocapsid. J. Cell Biol. *147*, 671–682.

Prevelige, P.E., Jr., Thomas, D., and King, J. (1988). Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells in vitro. J. Mol. Biol. *202*, 743–757.

Prevelige, P.E., Jr., Thomas, D., and King, J. (1993). Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. Biophys. J. *64*, 824–835.

Reinisch, K.M., Nibert, M.L., and Harrison, S.C. (2000). Structure of the reovirus core at 3.6 Å resolution. Nature *404*, 960–967.

Romantschuk, M., Olkkonen, V.M., and Bamford, D.H. (1988). The nucleocapsid of bacteriophage ϕ6 penetrates the host cytoplasmic membrane. EMBO J. *7*, 1821–1829.

Schmitz, K.S. (1990). An Introduction to Dynamic Light Scattering by Macromolecules. (Boston: Academic Press), pp 58–62.

Simpson, A.A., Tao, Y., Leiman, P.G., Badasso, M.O., He, Y., Jardine, P.J., Olson, N.H., Morais, M.C., Grimes, S., Anderson, D.L., et al. (2000). Structure of the bacteriophage ϕ29 DNA packaging motor. Nature *408*, 745–750.

Tabor, S. (1990). Expression using the T7 RNA polymerase/promoter system. In Current Protocols in Molecular Biology, F.A. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. (New York: Greene Publishing and Wiley-Interscience), pp. 16.2.1–16.2.11.

Valpuesta, J.M., and Carrascosa, J.L. (1994). Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. Q. Rev. Biophys. *27*, 107–155.

van Dijk, A.A., Frilander, M., and Bamford, D.H. (1995). Differentiation between minus- and plus-strand synthesis: polymerase activity of dsRNA bacteriophage ϕ6 in an in vitro packaging and replication system. Virology *211*, 320–323.

Vidaver, A.K., Koski, R.K., and Van Etten, J.L. (1973). Bacteriophage ϕ6: a lipid-containing virus of *Pseudomonas phaseolicola*. J. Virol. *11*, 799–805.

Wang, S., Palasingam, P., Nokling, R.H., Lindqvist, B.H., and Dokland, T. (2000). In vitro assembly of bacteriophage P4 procapsids from purified capsid and scaffolding proteins. Virology *275*, 133–144.

doi:10.1016/S0022-2836(03)00322-X J. Mol. Biol. (2003) 328, 791–804

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Conserved Intermediates on the Assembly Pathway of Double-stranded RNA Bacteriophages

Denis E. Kainov, Sarah J. Butcher, Dennis H. Bamford and Roman Tuma*

Department of Biosciences
Institute of Biotechnology
University of Helsinki
P.O. Box 65 Viikinkaari
1 FIN-00014 Helsinki, Finland

*Corresponding author

Double-stranded RNA (dsRNA) viruses are complex RNA processing machines that sequentially perform packaging, replication and transcription of their genomes. In order to characterize the assembly intermediates of such a machine we have developed an efficient *in vitro* assembly system for the procapsid of bacteriophage ϕ8. The major structural protein P1 is a stable and soluble tetramer. Three tetramers associate with a P2 monomer (RNA-dependent RNA polymerase) to form the nucleation complex. This complex is further stabilized by a P4 hexamer (packaging motor). Further assembly proceeds *via* rapid addition of individual building blocks. The incorporation of the packaging and replication machinery is under kinetic control. The *in vitro* assembled procapsids perform packaging, replication and transcription of viral RNA. Comparison with another dsRNA phage, ϕ6, indicates conservation of key assembly intermediates in the absence of sequence homology and suggests that a general assembly mechanism for the dsRNA virus lineage may exist.

Keywords: assembly *in vitro*; double-stranded RNA virus; molecular machine; packaging; replication

Introduction

Large nucleoprotein assemblies such as ribosomes, nucleosomes and viruses perform complicated tasks, like biopolymer synthesis and transport. Such assemblies can be viewed as molecular machines that are programmed by the genes coding for their subunits.[1,2] Additionally, the subunits are programmed to self-assemble but this process is not well understood.[3] Historically, simple viruses with their symmetry, amenable genetics and biochemistry became the early models for studying self-assembly.[4] The more complex viruses, in particular the double-stranded RNA (dsRNA) viruses that contain packaging and polymerase enzymes, can be viewed as RNA processing machines.[2,5] Insights into the mechanisms of assembly, RNA packaging, and replication were obtained using the *in vitro* packaging system of ϕ6, a dsRNA bacteriophage belonging to the Cystoviridae family (Figure 1).[6–10]

The cystoviruses (ϕ6, ϕ7, ϕ8, ϕ9, ϕ10, ϕ11, ϕ12 and ϕ13[11]) all have a genome composed of three segments of dsRNA packaged inside a polymerase complex, which is coated by a lipid-protein envelope.[12,13] Proteins P1 (major structural protein[14]), P2 (RNA-dependent RNA polymerase[15]), P4 (packaging enzyme[16]) and P7 (packaging factor[17]) form a stable assembly intermediate, the procapsid (PC), which in turn performs RNA packaging and replication (Figure 1).[10,18–21] The dodecahedral procapsid framework is composed of 60 P1 dimers forming a $T = 1$ icosahedral lattice (also referred to as $T = 2$ architecture[22]), an arrangement similar to that of other dsRNA viruses.[2,21,23] The 5-fold vertices of the procapsid are occupied by P4 hexamers.[24] The polymerase P2, a compact monomer,[25] has been localized at the 5-fold vertices under the P4 hexamers.[26] The position of the P7 dimers is less certain, most likely forming stabilizing clamps across 2-fold axes.[9,26,27]

Abbreviations used: dsRNA and ssRNA, double-stranded and single-stranded RNA; DLS, dynamic light-scattering; EM, electron microscopy; NTPase, nucleotide triphosphatase; ORF, open reading frame; P12, subviral particle or assembly reaction containing proteins P1 and P2; P124, subviral particle or assembly reaction containing proteins P1, P2 and P4; PC, procapsid; R_h, hydrodynamic radius.

E-mail address of the corresponding author: roman.tuma@helsinki.fi

0022-2836/03/$ - see front matter

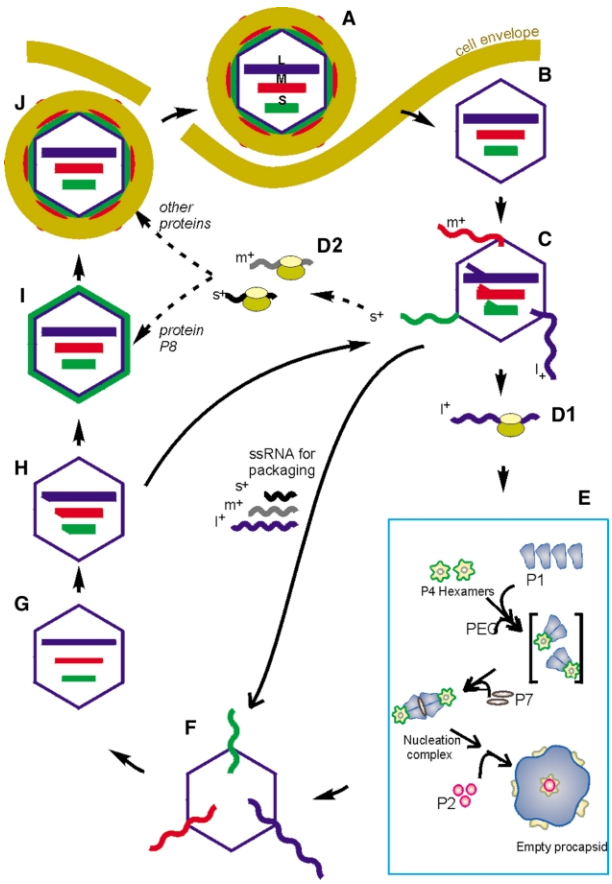


Figure 1. Life-cycle of bacteriophage ϕ6.[10] The virion, containing three dsRNA segments, S (small), M (medium) and L (large), (A) attaches to the host cell and (B) the polymerase complex penetrates the host membranes, losing both the envelope and the nucleocapsid coat). (C) The polymerase complex is activated and (D1 and D2) produces mRNAs that are translated into viral proteins. (E) An empty polymerase complex (also called procapsid, PC) is self-assembled from proteins P1, P2, P4 and P7; procapsid assembly intermediates are highlighted in the inserted box.[9] (F) The procapsid specifically packages the three ssRNA molecules s^+, m^+, and l^+. (G) The viral RNA polymerase P2 replicates the packaged ssRNA (H) into the double-stranded form inside the polymerase complex. (I) The polymerase complex is coated by a shell of protein P8 to form a nucleocapsid, which is (J) subsequently enveloped and mature virions leave the cell by lysis.

Notably, a similar arrangement of RNA processing enzymes has been observed in the viral cores of reovirus,[28] cypovirus[29] and rotavirus.[30]

High-resolution structures of dsRNA virus cores revealed the intricate structural framework of the machinery and pointed to significant similarities shared among all dsRNA viruses.[22,23,28] Despite this extensive structural information, relatively little is known about the process of polymerase complex assembly. A decamer of the bluetongue virus protein VP3, an α-helical protein analogous to ϕ6 P1, has been proposed as an assembly intermediate solely on the basis of the three-dimensional structure of the mature core.[22] More recently, the ϕ6 *in vitro* reconstitution system was used to characterize ϕ6 procapsid assembly intermediates (Figure 1).[9] A nucleation complex, composed of one P1 tetramer and one or two P4 hexamers, was proposed. Furthermore, the stabilizing role of P7 in the formation of the nucleation complex was established. On the other hand, no further assembly intermediate was detected and the identity of the stage during which P2 was incorporated into the growing procapsids remained elusive.

In order to judge the relevance of ϕ6 intermediates for the assembly of other dsRNA viruses and to characterize assembly intermediates further, we have selected another cystovirus, ϕ8, for *in vitro* assembly studies. ϕ8 shares less than 15% overall sequence homology with ϕ6 and the homologous regions are limited to the conserved polymerase (P2) and nucleotide triphosphatase (NTPase, P4) motifs.[31] Thus, observation of assembly intermediates common to both viruses would indicate conservation of the assembly pathway within the Cystoviridae family. Furthermore, given the absence of sequence homology, this would suggest that the conservation of the assembly mechanism reflects structural similarity and may extend to other dsRNA virus families with related morphology, e.g. the Reoviridae.

In the present work, we refine our understanding of dsRNA virus assembly by developing an *in vitro* system for bacteriophage ϕ8. *In vitro* assembly shows unequivocally that the major procapsid protein P1 assembles *via* a stable tetrameric intermediate. The nucleation complex is formed upon interaction of several P1 tetramers with the polymerase (P2) and is stabilized further by the packaging enzyme (P4 hexamer). Comparison with ϕ6 reveals a remarkable conservation of key assembly intermediates and suggests that assembly pathways of viruses with similar architecture may be conserved irrespective of viral sequences.

Results

Identity of the P7 gene

P7 associated with the isolated virus core had an apparent molecular mass of 19.2 kDa (Figure 2(A)) consistently lower than the 25.4 kDa expected on the basis of the published gene assignment.[31] Furthermore, mass spectrometric analysis of core proteins revealed a 19.2 kDa polypeptide but no 25 kDa moiety. A tryptic digestion of the core-associated 19.2 kDa protein produced fragments with masses matching a set predicted from the last open reading frame (ORF) of the viral L segment (Table 1; Figure 2(B)). Thus, the P7 protein is encoded by the last ORF of the L-segment, while the ORF preceding the P2 gene is designated as *H* (hypothetical) in accordance with the data bank information (Figure 2(B)).

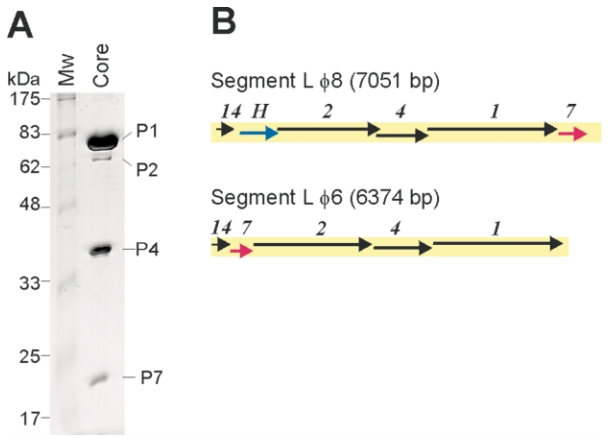


Figure 2. Identification of the P7 protein and gene 7 localization. (A) SDS-PAGE analysis of core-associated proteins showing migration of P7. (B) Map of the L segments from ϕ8 and ϕ6 indicating the gene order and the position of gene 7 (red).

Table 1. Tryptic peptides of ϕ8 P7 determined by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and predicted from sequence

Tryptic peptides, amino acid sequence	Experimental mass (Da)	Predicted mass (Da)
QVVVGYIR	933.7	933.2
TVMVDGYR	940.6	940.2
VGELTLSAR	945.7	945.2
TDPITAVAINR[a]	1170.8	1170.4
AAWASMEDEEK	1266.6	1266.5
VAHDVGGDEYTR	1318.6	1318.5
YVSDADDLGISPR	1407.7	1407.6
ITIGGDESSLEDDEQADEA[b]	1994.0	1994.2

[a] This peptide represents the N-terminal sequence, the terminal Met was removed during post-translational processing.
[b] This peptide represents the very C-terminal sequence, thus it does not end with K or R.

Expression and characterization of procapsid constituents

P1 and P4 expression (for plasmids and strains see Table 2) at 37 °C yielded large amounts of insoluble protein (not shown). The expression temperature was reduced to 18°C and sufficient quantities of soluble proteins were obtained and purified to homogeneity (Figure 3). Similarly, an improved purification protocol for P2 yielded pure and monodisperse protein (Figure 3).

All our attempts to obtain soluble protein from recombinant strains producing P7 alone have failed (not shown). Therefore, we have resorted to P7 co-expression with the other PC proteins and purification from the resulting recombinant procapsids (Table 2). P7 was released from the purified procapsids by repeated cycles of freezing and thawing and separated from other PC proteins by size-exclusion chromatography yielding sufficient amounts of pure P7 (Figure 3(A)).

Table 2. List of plasmids and strains used to produce proteins and assemblies

Plasmid	Host	Relevant property	Reference
pDK5	BL21(DE3)	P1 expression	This study
pSJ1b	BL21(DE3)	P4 expression	This study
pHY1	BL21(DE3)	P2 expression	[39]
pLM2424	JM109	PC expression and P7 purification	[31]

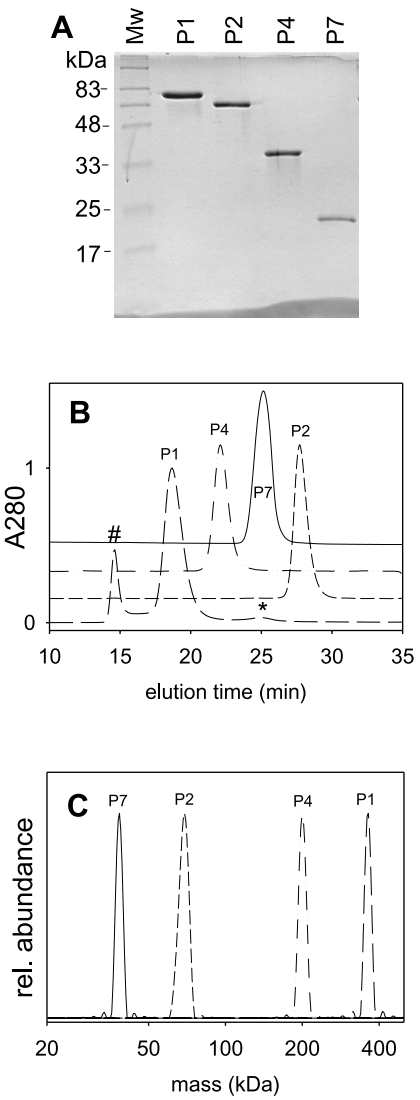


Figure 3. (A) Purity of the constituents used for the *in vitro* ϕ8 PC assembly: SDS-PAGE of proteins. (B) Oligomeric status of PC proteins. Purified proteins were separated on a size-exclusion column (Superdex 200, Amersham Biosciences) and detected by monitoring absorption at 280 nm (note that the traces were shifted for clarity). The peak labeled with # corresponds to P1 aggregate, which appeared after storage at 4 °C. A low-intensity peak (*) in the P1 trace corresponds to the residual P1 monomer. (C) Distribution of molecular masses that were calculated from calibrated light-scattering intensity and absorbance traces.

Table 3. Properties of ϕ8 PC components

Protein	Mass (kDa)[a]	Copies/ virion[b]	Multimeric status[c,d]	Hydrodynamic radius[c,e] (nm)	Secondary structure composition[c,f] (%)	Proposed function[g]
P1	86.768	120	Monomer Tetramer	3.7 ± 0.2 5.5 ± 0.2	Helix: 81.0 Coil: 13.5 Sheet: 0.0 Turn: 5.5	Major structural protein of PC
P2	71.495	12	Monomer	2.6 ± 0.2		RNA-dependent RNA polymerase
P4	34.057	72	Hexamer	5.1 ± 0.2	Helix: 50.0 Coil: 12.0 Sheet: 38.0	NTPase, genome packaging, helicase[h]
P7	19.039	60	Dimer	2.8 ± 0.2		Enhances genome packaging and replication

[a] Calculated from the cDNA sequence (N-terminal Met is removed).
[b] Estimated copy numbers, based on reports by Day & Mindich[54] and Hoogstraten *et al*.[31]
[c] This study.
[d] Multimeric status of proteins, determined by analytical gel-filtration (Superdex-200).
[e] Determined by dynamic light-scattering.
[f] Determined by CD as described.[55]
[g] Functions were assigned in analogy to the ϕ6 system.[10]
[h] D.E.K., D.H.B. & E. Makeyev, unpublished results.

The oligomeric status, shape and secondary structure of the procapsid building blocks were characterized using various methods (Table 3). P1 was soluble and highly α-helical. The molecular mass of P1, obtained by gel-filtration and light-scattering (Figure 3(B)), was 350(±10) kDa and matched that of a tetramer (346.8 kDa). A small amount of P1 monomer was also detected (<1% of total P1, Figure 3(B)). The hydrodynamic radius of the P1 monomer, $R_h = 3.7(\pm 0.2)$ nm, indicated significant asymmetry (asymmetry ratio $R_h/R_s = 1.15$, where $R_s = 3.2$ nm is the radius of an equivalent hydrated sphere). On the other hand, the P1 tetramer was more compact, with $R_h = 5.5(\pm 0.2)$ nm and $R_h/R_s = 1.06$. This is in contrast to the P1 from ϕ6, which could be purified only as a relatively insoluble monomer in small amounts from disrupted procapsids.[9]

The packaging enzyme P4, a hexamer with an α/β fold (apparent mass 206(±10) kDa, $R_h = 5.1(\pm 0.2)$ nm), also exhibited RNA helicase activity (D.E.K., D.H.B. & E. Makeyev, unpublished results). The hydrodynamic data and analogy to ϕ6 suggests that ϕ8 P4 forms a hexameric ring.[32]

The P2 monomer (apparent mass = 70(±5) kDa) appeared spherical ($R_h = 2.6(\pm 0.2)$ nm, $R_h/R_s = 0.97$) as expected from comparison with the ϕ6 polymerase.[25]

The virion-derived P7 was a dimer (apparent mass = 41(±5) kDa) with a slightly asymmetric shape ($R_h = 2.8(\pm 0.2)$ nm, $R_h/R_s = 1.12$), significantly less asymmetric than the ϕ6 P7 dimer.[9]

In vitro assembly

In order to determine which protein combinations (designated PXY according to protein composition X + Y, e.g. P12 reaction contained P1 and P2) could assemble *in vitro*, we first employed conditions similar to those used for ϕ6 (i.e. one hour incubation in the presence of 4% (w/v) PEG 6000[9]). The reaction products were negatively stained immediately and analyzed by electron microscopy (EM) (Figure 4(A)). Aliquots were separated by sucrose-gradient centrifugation and light-scattering fractions were analyzed by SDS-PAGE (Figure 4(B)). In all cases, the protein ratios in the assembled particles were similar to those found in the virus core control (Figure 4(C)). The results demonstrated clearly that in the presence of PEG all P1-containing reactions yielded procapsid-like particles.

Electron micrographs demonstrated that the final products were identical in shape and size (50 nm diameter) with the control particles (Figure 4(A) and (B)). In addition to the procapsid-like particles, EM revealed other assemblies: 10–15 nm L-shaped and linear structures (mainly in the P12 assembly reaction) and 20 nm diameter "O-rings" (in all reactions) were observed. These assemblies may represent an equilibrium population of intermediates at the time of staining or aberrant aggregates.

Nucleation mechanism of procapsid assembly

Because P1 alone could assemble in the presence of PEG, it was likely that P1 would assemble at sufficiently high concentrations in the absence of PEG. Indeed, P1 solutions contained assembled and aggregated material after being concentrated to about 5 mg/ml. After removing the assembled material, the concentration of the soluble P1, c_{P1}, was 4.6 mg/ml, which is a good approximation to the critical concentration for P1 assembly. Hence, all further assembly experiments excluded PEG

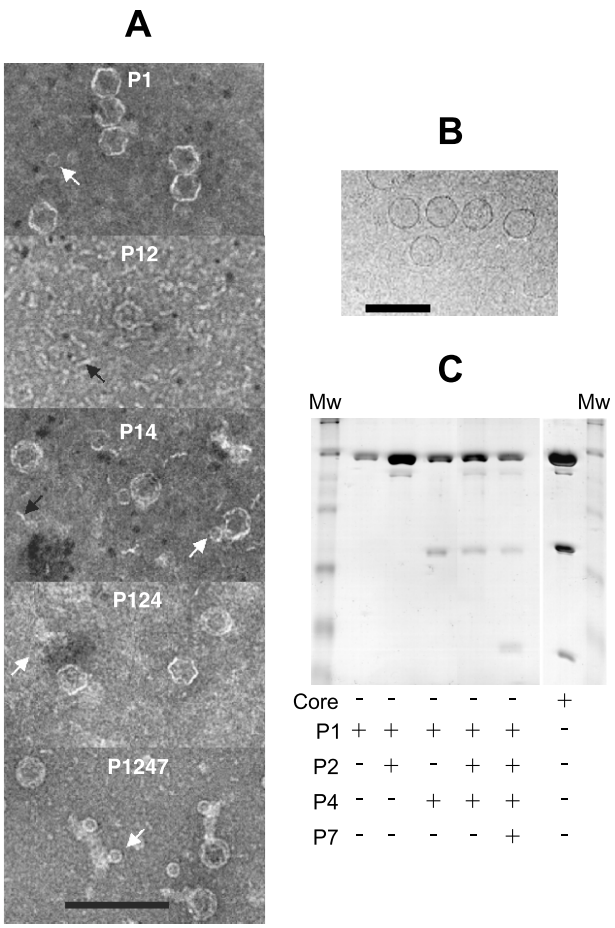


Figure 4. *In vitro* assembly in the presence of 4% (w/v) PEG 6000. (A) Negative stain electron microscopy of assembly products. Assemblies were named after their protein compositions, e.g. an assembly containing proteins P1, P2 and P4 was named P124. The scale bar represents 200 nm. Distinct assembly intermediates 10–12 nm in size such as L-shaped substructures (P12 assembly reaction) and O-ring substructures (in all cases) are indicated by arrows. (B) Cryo-electron micrograph of the recombinant *in vivo* assembled PC, purified using sedimentation through a 5%–20% (w/v) sucrose gradient. The scale bar represents 100 nm. (C) Protein composition of the *in vitro* assembled particles, P1, P12, P14, P124 and PC after separation by rate zonal centrifugation. The *in vivo* assembled PC is included as a control.

simplifying the reactions and light-scattering experiments.

The critical concentration behavior of P1 indicated that procapsid proteins might assemble in a nucleation-limited fashion. Furthermore, specific procapsid proteins might lower the critical concentration *via* stabilization of the nucleation complex. To explore the nucleation mechanism, a turbidity assay was adopted to follow the assembly time-course (kinetics) for different combinations and concentrations of procapsid proteins (Figure 5). The turbidity increase was found to be proportional to procapsid formation (rate-zonal sucrose-gradient analysis, not shown) and thus constitutes a good quantitative measure of assembly progress.

Stoichiometric amounts of procapsid proteins were mixed employing concentrations of P1 below c_{P1} (Figure 5(A)). Both combinations P1 + P7 (not shown) and P1 + P4 resulted in no turbidity increase. On the other hand, P1 could assemble in the presence of P2. In all cases the turbidity kinetics showed a distinct lag phase, indicative of nucleation-limited assembly.[33] Although P4 alone failed to initiate assembly, it shortened the lag phase substantially and accelerated the assembly in the presence of P2. Thus, P4 augmented the nucleation effect of P2. P7 had only a small additional effect on the assembly kinetics. Thus, P1 and P2 form the minimal nucleation complex, which is stabilized substantially by P4.

In order to estimate the size of the minimal nucleation complex, the concentration-dependence of the assembly kinetics was examined for the P12 assembly (Figure 5(B)).[9,33] The assembly rate dependence on the P1 concentration in an excess of P2 showed a non-linear increase and the apparent molecularity $n = 2.9(\pm 0.2)$. These results suggest that about three P1 tetramers associate to form the nucleation complex.

Similar analyses for P2 yielded an approximately linear rate increase up to a concentration of P2 of 2 mg/ml (Figure 5(C)). This suggests that one P2 monomer participates in nucleation.

Detection of assembly intermediates by dynamic light-scattering

Dynamic light-scattering (DLS) measures the fluctuation of scattered light intensity and typically gives distributions of hydrodynamic radii, $I(R_h)$. In comparison with turbidity, which is proportional to the weight average molecular mass of the assemblies, DLS provides additional information about the polydispersity of the sample.[34]

Because the distribution intensity $I(R_h)$ is proportional to the number density times the scattering cross-section of the species with R_h (scattering cross-section is roughly proportional to the square of the species mass), the method is particularly sensitive to the presence of larger assemblies. This property could be used to detect procapsid formation even at low number density (e.g. less than 0.1% (w/v) of assembled protein). Therefore, comparison of time-resolved R_h distributions ($I(t, R_h)$; a few snapshots are shown in Figure 5(F) and (G)) with the turbidity increase should be instrumental in distinguishing the fast growth of procapsids from the accumulation of intermediates. In the former case, the distribution should be bimodal, encompassing two sets of peaks. The first set at low R_h (below 6 nm) would correspond to the mixture of precursor subunits, while the second set should be observed at R_h around 25, corresponding to procapsids (R_h 25 nm) or larger assemblies ($R_h > 35$ nm). Alternatively, the early distribution would contain broader

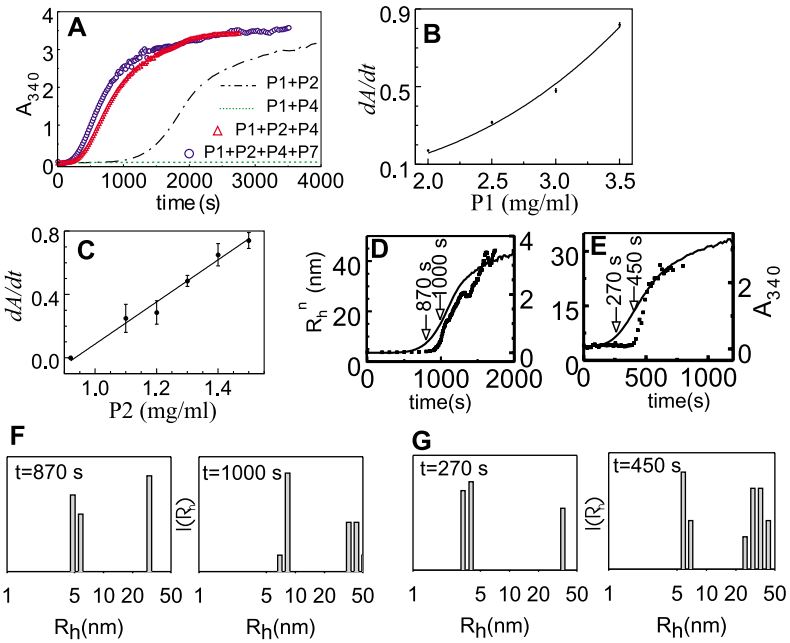


Figure 5. Kinetics of *in vitro* assembly: (A) Kinetics of PC assembly followed by turbidity at 340 nm, 20 °C. Assembly of P1 in the presence of P2, P4 and P7. Concentrations: P1, 1.2 mg/ml; P2, 0.2 mg/ml; P4, 0.17 mg/ml. (B) Concentration-dependence of P12 assembly kinetics on P1:P2 concentration, 5 mg/ml. The slope of the steepest turbidity increase $k = dA/dt$ was plotted as a function of P1 concentration and fitted to an exponential curve $k = 0.021[P1]^{2.91}$ according to equation (1). (C) Dependence of P12 assembly kinetics on the concentration of P2. The concentration of P1 was held constant at 2 mg/ml. The assembly rate was fitted by a linear function of P2 concentration. (D) Kinetics of P12 assembly followed by turbidity (continuous line) and time-resolved DLS (squares, number average hydrodynamic radius R_h^n computed according to equation (2) from data in (F) and additional data not shown). Concentrations: P1, 2 mg/ml; P2, 4.3 mg/ml; temperature, 23 °C. Arrows show two time-points at which the representative R_h distributions shown in (F) were obtained. (E) Kinetics of P124 assembly followed simultaneously by turbidity and DLS. Other conditions as in (D), except the concentration of P4 was 0.5 mg/ml. (F) P12 assembly $I(t, R_h)$ distributions for the two representative time-points. Due to the time-resolved nature of the data (15 second scans) and due to the logarithmic distribution of the correlator channels, the positions of peaks within $I(t, R_h)$ are reported to ± 2 nm for $R_h < 10$ nm, ± 5 nm for the 20–30 nm range and ± 10 nm above R_h 30 nm. (G) The same as in (F) but for P124 assembly (conditions as described for (E)).

peak(s) at R_h values between 8 nm and 20 nm, indicating significant accumulation of intermediates.

$I(t, R_h)$ measured for the P12 assembly reaction at the end of the lag phase ($t = 870$ seconds; Figure 5(D) and (F)) showed a bimodal distribution with a cluster of peaks at 5 nm and a well-separated peak at 25 nm corresponding to procapsid-sized assemblies. Similar results were obtained for the P124 assembly (Figure 5(E) and (G); $t = 270$ seconds). The absence of any detectable intermediates and the presence of procapsid-sized assemblies at early stages suggests that nucleation is followed by rapid growth and completion of procapsids. Furthermore, the results imply that no significant amount of intermediates accumulated during the lag phase, confirming a nucleation-limited assembly mechanism (i.e. the nucleation complex constitutes the species of lowest abundance[33]).

$I(t, R_h)$ distributions remained bimodal well into the growth phase (1000 seconds for the P12 and 450 seconds for the P124 assembly, respectively, Figure 5(F) and (G)). The only significant difference from the earlier time-points was the shift of the precursor peak from $R_h \approx 5$ nm to values between 8 nm and 10 nm. The peak shifted further to higher R_h with time and eventually merged with the peak of procapsid-sized assemblies (not shown). In order to rule out the possibility of deconvolution artifacts the significance of these shifts was confirmed directly by comparing the initial decays of correlation function (not shown). Thus, this peak represents a discrete species that is experimentally well resolved from the mixture of precursors and the products, i.e. it most likely corresponds to assembly intermediates.

The observed increase of the intermediate size is summarized in the time-resolved number-average hydrodynamic radius ($R_h^n(t)$ plotted in Figure 5(D) and (E)). R_h^n was computed from R_h distributions that were weighted by the inverse of the scattering cross-section (equation (2)) and therefore it is representative of the stoichiometrically most abundant species. However, the large scattering cross-section of procapsids and the instrumentation limits would make it difficult to detect intermediates with $R_h \approx 10$ nm if they represented less than 15% of the total protein mass. Given the scattering intensities in Figure 5(F) and (G), the observed intermediates were significantly abundant (more than 40% of total protein mass as judged from $I(1000, R_h)$ for P12 and $I(450, R_h)$ for P124, respectively) in the middle of the growth phase. Apart from the undetectable 15% referred to above, these intermediates were gradually (step-wise) converted to procapsids during the later part of the growth phase (Figure 5(D) and (E)). This suggests that the observed intermediates were on the procapsid assembly pathway.

The continuous increase of the apparent Rh after about 1200 seconds indicates that the rise in turbidity resulted from the assembly of aberrant

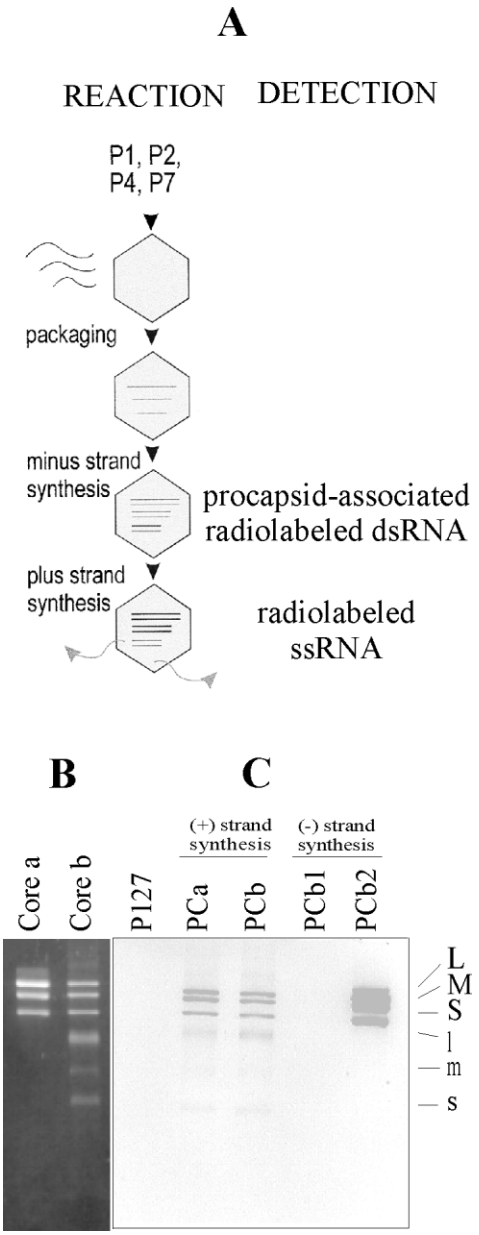


Figure 6. RNA packaging and RNA synthesis activity of the *in vitro* assembled procapsids. (A) A flow-chart depicting the sequence of packaging, minus (replication) and plus (transcription) strand syntheses reactions (left) and product detection scheme (right). (B) Ethidium bromide-stained gel showing products of minus strand synthesis (lane core a) and plus strand synthesis (lane core b) performed by virus-derived cores. (C) An autoradiograph of a native agarose gel that separated products of the *in vitro* replication reactions: lane P127, PC lacking P4; lane PCa, *in vivo* PC; lane PCb, *in vitro* assembled PC; lane PCb1, pellet of SDS-disrupted procapsids; lane PCb2, procapsid pellet after sedimentation through a 12% sucrose cushion. L, M and S denote dsRNA segments while l, m, and s correspond to ssRNA transcripts.

P12 products (Figure 5(D)). Because the rise of the $R_h^n(t)$ is steeper than the turbidity slope, it is likely that these assemblies were filamentous.[35] Long filaments were also seen in electron micrographs of P12 assemblies (data not shown). Figure 5(E) shows a similar analysis for the P124 assembly. The apparent number-average Rh reached a plateau at 25 nm (600–800 seconds) with only a slight increase at longer times. Thus, the P124 assembly seemed to be less prone to further aggregation.

In vitro assembled procapsids can package, replicate and transcribe RNA

Packaging of genomic single-stranded RNA (ssRNA) segments and their subsequent replication within the viral core are the hallmarks of polymerase complex activity (Figure 1 (F)–(H)). If the *in vitro* assembled procapsids represent the biologically relevant form they should package and replicate the viral genome. Given that packaging is a prerequisite for minus strand synthesis, a simplified assay was designed that detected RNA replication associated with *in vitro* assembled particles (e.g. combined packaging and minus strand synthesis, Figure 6(A)).

Unlabeled ϕ8 genomic ssRNA, which was transcribed *in vitro* by viral cores, was used as a precursor for the reactions (Figure 6(B)). *In vitro* assembled and purified procapsids were mixed with ssRNA substrate and radiolabeled nucleotides and the product of the reaction was sedimented through a 12% (w/v) sucrose cushion to remove any RNA that was not associated with procapsids. A comparable amount of radiolabeled dsRNA was found in the pellets of the *in vivo* (Figure 6(C), lane PCa) and the *in vitro* (Figure 6(C), lane PCb2) assembled procapsids. The labeled dsRNA was also tightly associated with the procapsid and was released only upon procapsid disruption with SDS (Figure 6(C), lane PCb1). Thus, *in vitro* assembled procapsids package and replicate viral RNA with activity similar to that of the *in vivo* assembled control.

A negative control was provided by *in vitro* assembled procapsids lacking P4, which did not synthesize dsRNA (Figure 6(C), lane P127). This shows that procapsid-associated P2 cannot replicate viral RNA without prior ssRNA packaging (P4-dependent).

Plus strand synthesis activity (combined ssRNA packaging, replication, and transcription) was measured by the production of radiolabeled ssRNA (Figure 6(A)). The activity of *in vitro* assembled procapsids was found to be comparable to that of recombinant procapsids (cf. lanes PCb and PCa in Figure 6(C)). In summary, *in vitro* assembled procapsids were able to package, replicate and transcribe viral RNA.

Discussion

Gene order of ϕ8 L-segment

In the course of cloning and purifying the ϕ8 procapsid genes we noticed a discrepancy between

the apparent molecular mass of the recombinant gp7 protein (25 kDa according to the published gene assignment[11,31]) and the core-associated P7 (19 kDa). The latter has been assigned to the last ORF of the viral L-segment using tryptic fragmentation and mass spectrometry. During the progress of this study, a corrected version of the ϕ8 sequence (accession number NC_003299) was deposited in the data bank that made the same P7 gene assignment. Our findings confirmed the correction.

In contrast to the L-segment of other cystoviruses, ϕ8 contains an extra ORF (designated H) encoding a protein of unknown function and the gene order is H2417. Interestingly, the highly basic 25 kDa protein could bind to the procapsids *in vitro* (data not shown), although it was not found in virions produced during normal infection. Additionally, gpH and P7 have limited sequence similarity (25%). One possibility is that these two genes have arisen by a gene duplication event. Such events are common in the evolution of dsDNA viruses[36] but have been detected in only a few cases of dsRNA viruses.[37,38]

Assembly pathway for dsRNA bacteriophages

Figure 7 summarizes the assembly pathway as proposed for bacteriophage ϕ8. Assembly proceeds *via* nucleation-limited polymerization, and proteins P1 and P2 are required and sufficient to nucleate procapsid assembly in the absence of PEG. In the presence of PEG or at high concentrations, P1 alone assembles into procapsid-like particles. The nucleation complex is composed of several P1 tetramers associated with a P2 monomer (estimated stoichiometry 3:1). This complex is further stabilized by the P4 hexamer. The size of the stabilized nucleation complex is approximated by the smallest detectable intermediate with R_h 8–10 nm, which was identified by hydrodynamic analysis.

The growth phase is rather fast in the presence of a sufficient concentration of building blocks. After the reaction mixture was depleted of the building blocks (essentially P1) certain assembly intermediates are left unfinished, as seen in the electron micrographs in Figure 4(A) and by DLS (R_h 8–10 nm, Figure 5(D) and (E)).

The P4 hexamer seems to increase the fidelity of assembly. In the absence of P4 (e.g. P12 assembly) the product is polydisperse (Figure 5(D)) and may contain filamentous material (Figure 5(D) and EM data not shown). P4-containing reactions lead to a product with the correct size (R_h 25 nm Figure 5(E)). P4 hexamer presumably introduces curvature into the growing intermediates and steers the assembly towards closed shells *via* stabilization of the vertices and competition with the aberrant pathways. Thus, P4 most likely provides key interactions for the correct formation of the 5-fold vertices.

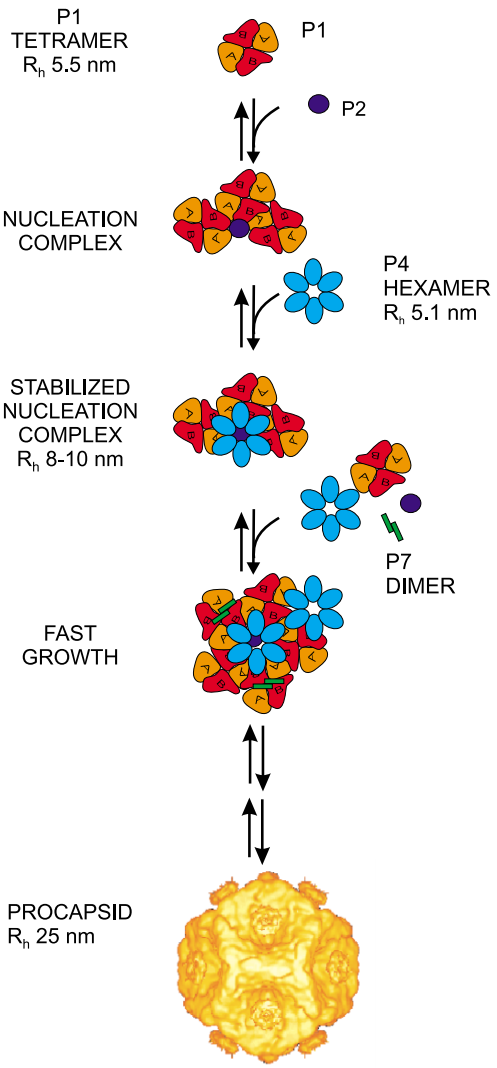


Figure 7. *In vitro* assembly pathway model for ϕ8 procapsid. The procapsid building block is a compact P1 tetramer. Three tetramers associate with one P2 monomer to form a nucleation complex. This complex is stabilized further by a P4 hexamer and assists the formation of neighboring 5-fold vertices. Further assembly proceeds rapidly by addition of individual building blocks. When building blocks are depleted, an intermediate with R_h 8–10 nm is observed transiently and most likely corresponds to the stabilized nucleation complex. Procapsid representation is according to de Haas *et al.*[24]

The negligible effect of P7 on the ϕ8 assembly kinetics confirmed its role in stabilizing the P1 tetramer during ϕ6 procapsid assembly.[9] Such a function is redundant in ϕ8 assembly due to the stability of the P1 tetramer and no kinetic effect would thus be expected.

Further comparison of the ϕ8 assembly with the ϕ6 system[9] identifies several common features and intermediates. A ϕ6 P1 tetramer has been proposed solely on the kinetic basis to constitute an essential assembly intermediate. In the ϕ8 system, P1 forms a stable tetrameric building block. Thus, the

tetrameric intermediate is common for both assembly pathways. As proposed previously, each tetramer can bridge together two P4 hexamers and thus facilitates association of the neighboring 5-fold vertices during assembly.

In the ϕ6 system, P4 is the necessary component for *in vitro* nucleation of assembly. In ϕ8 assembly, P2 plays a similar role, while P4 is involved in the stabilization of the nucleation complex. These results suggest that incorporation of both P2 and P4 is under kinetic control. The incorporation of P2 into the ϕ6 procapsid had no apparent effect on the assembly kinetics in the low concentration regime (0.3 mg/ml of P1 monomer in 6% (w/v) PEG 4000).[9] It is possible that the kinetic effect of P2 at low concentrations was diminished by the key role of P4 and P7 in stabilizing the P1 tetramer. It is conceivable that the ϕ6 P2 associates with an assembly intermediate similar to that of the ϕ8 nucleation complex, in which the P1 tetramer is already formed.

Assembly intermediates with an apparent R_h of 10–12 nm have been observed for ϕ6 assembly at the end of the lag phase.[9] Similar intermediates (smallest species with R_h 8–10 nm) were observed for ϕ8 in the middle of the growth phase. These intermediates most likely represent metastable species that are formed by stabilization of the nucleation complex (Figure 7) although the determination of the exact stoichiometry of these intermediates is beyond the scope of the current analysis. The absence of the early detectable intermediates in the ϕ8 system reflects the higher concentrations of protein employed for ϕ8 assembly, leading to faster growth rates. This is consistent with the proposed nucleation-limited assembly mechanism.

We show that the *in vitro* assembled procapsids package, replicate and transcribe viral RNA. The present results show that the packaging NTPase P4 is necessary for subsequent replication. Because ϕ8 P2 can replicate ssRNA in the absence of procapsids,[39] this suggests that the active site of P2 is not accessible on the procapsid surface. Such an internal location is consistent with the recently determined position of P2 within the ϕ6 procapsid.[26] Thus, successful packaging of ssRNA is a prerequisite for ϕ8 RNA replication, which takes place inside the procapsid. This mode of replication is identical with that of ϕ6.[40–42] The ϕ8 polymerase complex (resulting from procapsid packaging and replication) is able to produce mRNA *de novo* as seen for the ϕ6 system.[41,43]

In conclusion, the ϕ6 and ϕ8 systems exhibit similar building blocks, share key procapsid assembly intermediates, package and replicate RNA in a similar fashion despite their very low level of sequence homology.[31] Such conservation of the assembly pathway and function may reflect common ancestry as well as physico-chemical constraints imposed on the procapsid assembly and may extend to other members of the Cystoviridae. It remains to be seen whether the tight regulation of genome replication and transcription observed for ϕ6 is applicable to ϕ8 and other cystoviruses.[8,41,44–46]

Implications for other dsRNA viruses

The unique polymerase core architecture (so-called $^{22}T = 2$) and function seem conserved between different families of prokaryotic and eukaryotic dsRNA viruses.[2] Despite extensive structural knowledge, very little is known about the assembly intermediates of the eukaryotic viruses of the Reoviridae family. Given the structural and functional similarity, the proposed mechanism of Cystoviridae assembly may have implications for the assembly of other dsRNA viruses.

The polymerase core framework is built from 120 subunits of the same protein that adopt two distinct conformations (i.e. $T = 1$ shell of 60 asymmetric dimers). At which stage of the assembly are the two non-equivalent conformations attained? The significance of the tetrameric intermediate suggests that the conformational switching may happen during the formation of the P1 tetramer, which could be envisioned as a dimer of dimers.

A decamer of the bluetongue virus protein VP3 (P1 analog) has been proposed as an assembly intermediate.[22] Such a large assembly would pose two problems for efficient assembly: (1) its formation would be kinetically unfavorable, i.e. such an intermediate would have a low concentration; (2) subsequent growth of shells by addition of these large building blocks would be impeded by both slow diffusion and low concentration. On the other hand, the tetramer is sufficiently small to assemble efficiently. Additionally, the tetramer would also tie together neighboring 5-fold vertices thus facilitating formation of closed shells. Interestingly, a dimeric asymmetric unit of VP2 that connects two neighboring vertices has been identified in the rotavirus core.[47] The asymmetric dimer may play the same role in rotavirus assembly, as the P1 tetramer does in the bacteriophage.

The replication and transcription machinery of many dsRNA viruses resides inside the polymerase complex.[26,30,48,49] Our present results suggest that incorporation of the transcription complex (including the RNA polymerase) might be under kinetic control and may happen during the early phase of assembly. This would assure internalization of the transcription complex before closure of the polymerase complex shell. If the P2 monomer and the P4 hexamer are incorporated early during assembly when symmetry is non-existent or low their incorporation does not depend on the final icosahedral symmetry. This may explain why the transcription complex does not follow icosahedral symmetry and explains the mechanism for incorporating the symmetry mismatch at the 5-fold vertices of dsRNA viruses (one transcription complex associated with a decamer of the core protein[24,30,49]).

In summary, we have established an *in vitro* assembly system for the dsRNA bacteriophage ϕ8, which allowed detection of intermediates and delineation of the nucleation mechanism. In addition, comparison with the assembly pathway of another dsRNA bacteriophage ϕ6 suggested conservation of assembly pathways in the absence of significant sequence homology. Such conservation may extend to eukaryotic dsRNA viruses. The assembly mechanism also explains how asymmetry may be incorporated into the icosahedral capsids of dsRNA viruses.

Materials and Methods

Bacterial strains and plasmids

Escherichia coli DH5 (Gibco-BRL) was the host for plasmid propagation and molecular cloning. Plasmid pLM2424 was used for PC production.[31] Plasmid pHY1 encoded ϕ8 protein P2.[39]

To construct a plasmid for P1 protein expression, the P1 gene was PCR-amplified from pLM2424 template with recombinant *Pfu* DNA polymerase (Stratagene) and the oligonucleotides

5′-GGAGTTGA*CATATG*AGTAAGCTTGATCT-3′
and 5′-TAGGATCCGTCATGTCACATACCTT-3′

(restriction sites underlined) served as upstream and downstream primers, respectively. The PCR fragment was inserted into pT7-7 plasmid[50] between the *Nde*I-*Bam*HI. *E. coli* BL21(DE3) (Novagen) was transformed with the resultant plasmid pDK5 to give the P1 producing strain BL21(DE3/pDK5).

To construct the ϕ8 P4 expression strain, the *p4* gene was amplified from pLM2424 using the primers

5′-TCGTCAA*CATATG*GCTAGAAAAACGAAAGT-3′
and 5′-CCCAGATC*AAGCTT*ACTCATTTGTCA-3′.

The PCR fragment was inserted into pET32b(+) (Novagen) at *Nde*I-*Hin*dIII sites to produce plasmid pSJ1b. The pSJ1b was introduced into BL21(DE3) (Novagen), yielding BL21(DE3/pSJ1b).

Preparation of viral cores and recombinant PCs

For core production bacteriophage ϕ8 was grown on *Pseudomonas syringae* LM2489 and purified as described.[31] For ssRNA production, cores were purified as described.[51] Recombinant empty PCs were expressed in *E. coli* JM109 harboring plasmid pLM2424 and purified using Triton X-114 extraction and sedimentation through a 5%–20% (w/v) sucrose gradient in 20 mM Tris–HCl (pH 8), 50 mM NaCl, 7.5 mM $MgCl_2$. The light-scattering zone was collected and stored at −80 °C. For cryo-EM, a fresh sucrose band was pelleted in a Beckman airfuge at 95,000 rpm (A-95 rotor) for 15 minutes at 15 °C, immediately resuspended in buffer A (20 mM Tris–HCl (pH 8.0), 50 mM NaCl, 7.5 mM $MgCl_2$) and vitrified as described.[51]

P7 gene identification

The purified PC and core proteins were separated on SDS-PAGE (15% polyacrylamide gel) and stained with Coomassie brilliant blue. The putative P7 band was cut out of the gel and digested with trypsin as described.[52] The resulting tryptic peptide mixture was analyzed by automated matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (Voyager DE-STR, Applied Biosystems). Peptides were assigned using an in-house built sequence database of cystovirus genes.

Expression and purification of recombinant P1, P2, P4 and P7 proteins

To achieve expression of soluble P1 and P4, starter cultures of BL21(DE3/pDK5) and BL21(DE3/pSJ1b) in Luria–Bertani medium containing 150 μg/ml of ampicillin were grown at 37 °C until the absorbance at 540 nm (A_{540}) reached 0.5. These were then diluted 50-fold into 1 l of fresh medium each. The diluted cultures were further grown at 37 °C to an A_{540} of 0.6. The cultures were chilled on ice and induced with 1 mM isopropyl-β,D-thiogalactopyranoside (IPTG). IPTG-induced cells were incubated for an additional 12 hours at 18 °C. All the following steps for P4 purification were performed at 4 °C. P1 purification was performed at room temperature unless otherwise stated.

Bacteria were collected by centrifugation, resuspended in 10 ml of buffer and lysed using a French pressure cell at ≈105 MPa. Phenylmethylsulfonyl fluoride (PMSF) was added to a concentration of 1 mM after lysis. Cells debris were removed by centrifugation at 120,000*g* for 1.5 hours.

P1

The supernatant fraction was brought to 10% saturation with ammonium sulfate and kept on ice for ten minutes. Precipitated P1 was collected by centrifugation (Sorvall SS-34 rotor, 12,000 rpm for 20 minutes at 4 °C), dissolved in 2–3 ml of buffer A, filtered and loaded onto an affinity column (heparin-agarose, Amersham Biosciences). Proteins were eluted with a linear 0.1–1 M NaCl gradient buffered with 20 mM Tris–HCl (pH 8.0), 7.5 mM $MgCl_2$. Fractions containing P1 were pooled and diluted fivefold with ice-cold distilled water and applied onto an anion-exchange column (HiTrap HP Q-Sepharose, Amersham Biosciences). Proteins were eluted with a linear 0.1–1 M NaCl gradient buffered by 20 mM Tris–HCl (pH 8.0), 7.5 mM $MgCl_2$. Fractions containing P1 were pooled, filtered and resolved by gel-filtration (Superdex-200, Amersham Biosciences) equilibrated with buffer A. Purified P1 was stored at −80 °C.

P4

P4 was purified following a protocol similar to that used for the purification of P1 but without the precipitation in ammonium sulfate step. The purified protein was stored at +4 °C.

P2

P2 was expressed in *E. coli* BL21 from plasmid pHY1 and purified using a modification of the published protocol.[39] Briefly, the bacterial pellet from a 2 l culture was resuspended in 30 ml of 50 mM Tris–HCl (pH 7.4),

100 mM NaCl, 1 mM EDTA, 1 mg/ml of hen egg-white lysozyme (Boehringer Mannheim), 1 mM Pefablock SC (Roche), 0.05 mg/ml of DNase I (Sigma), and lysed by passage through a French press at ∽105 MPa. After clearance of the lysate at 11,5000*g* for 1.5 hours, it was applied to two 5 ml Hitrap Q Sepharose HP columns (Amersham Biosciences) connected in series, and the flow-through was loaded directly onto a Reactive Brown 10 agarose (Sigma) column in 50 mM Tris–HCl (pH 7.4), 100 mM NaCl, 1 mM EDTA. The polymerase was eluted in one step with 50 mM Tris–HCl (pH 7.4), 500 mM NaCl, 1 mM EDTA. A greater degree of homogeneity was achieved when required by the following two additional steps. After dilution to 50 mM Tris–HCl (pH 7.4), 30 mM NaCl, 1 mM EDTA, the protein was applied to a heparin-Sepharose HP column (Amersham Biosciences) and eluted in a 0–700 mM NaCl gradient in 50 mM Tris–HCl (pH 7.4), 1 mM EDTA. Gel-filtration was then carried out on a HiLoad 16/60 Superdex 200 column (Amersham Biosciences). The protein was stored at +4 °C.

P7

P7 was purified from recombinant PC particles. The particles were expressed in *E. coli* BL21(DE3/pLM24224)[31] then disrupted by three repeated cycles of freezing (−80 °C) and thawing (room temperature). Dissociated PC proteins were separated on a HiLoad 10/30 Superdex 200 column (Pharmacia) equilibrated with 20 mM Tris–HCl (pH 8), 50 mM NaCl. Gel-filtration fractions containing purified P7 were stored at −80 °C.

Procapsid assembly

The purified recombinant proteins P1, P2, P4, and P7 in buffer A were mixed in the proportions found in the ϕ8 virion. Proteins were assembled in the presence of 4% (w/v) PEG 6000. The *in vitro* assembly reaction was initiated by the addition of P1 so that its final concentration was 2 mg/ml; other proteins were added in stoichiometric amounts. The samples for negative staining were prepared after 60 minutes incubation at room temperature to examine products of reactions including any remaining intermediates. Assembled particles were separated by sedimentation (Beckman SW41 rotor, 35,000 rpm for 80 minutes at 10 °C) on a 5%–20% (w/v) sucrose gradient made in buffer A. The gradients were fractionated using a BioComp gradient fractionator and the fractions corresponding to the light-scattering zone were analyzed by SDS-PAGE and detected by staining with Coomassie brilliant blue. The fractions containing PC assemblies were further assayed for enzymatic activities (e.g. packaging and RNA synthesis, see below).

The efficiency of the *in vitro* PC assembly reactions was determined from the rate zonal gradient fractions by measuring the protein content of fractions containing the assembled products as well as those containing the non-assembled material (top of the gradient).

Critical concentration for P1 assembly

The P1 tetramer was concentrated to about 5 mg/ml using a Schleicher–Schuell microconcentrator (100 kDa cut-off) and the assembled material was removed by ultracentrifugation (Beckman Airfuge, rotor A-95, 29.5 psi (1 psi ≈6.9 kPa) for 30 minutes at 15 °C). The critical concentration, defined as the concentration of soluble P1 in the supernatant, was determined as the ratio between absorbance at 280 nm and the P1 extinction coefficient ($\epsilon_{280} = 0.66\ cm^2\ mg^{-1}$).

Assembly kinetics

The concentration dependent assembly kinetics were performed in buffer A at 20 °C. Light-scattering (turbidity) experiments were performed using a V-560 UV/VIS spectrophotometer (JASCO). Light-scattering was recorded as the increase of absorbance at 340 nm (2.0 nm bandwidth) in two or five second time increments up to 8000 seconds. The relative rates of assembly were determined as the slope of the steepest part of the turbidity curve before a significant change in the reactant concentration.[33]

The concentration dependence of the assembly rate (k) was approximated by the following relationship:

$$k = C[\mathrm{P1}]^n \tag{1}$$

where C is a normalization constant and n is the molecularity of the nucleation reaction.

Analysis of assembly intermediates by dynamic light-scattering

Buffer solutions used for DLS were filtered through a 20 nm pore syringe filter (Millipore). Proteins were centrifuged in a Beckman Airfuge (rotor A-95, 29.5 psi for 30 minutes at 15 °C) and examined by DLS to assure no aggregates or assemblies were present ($R_h < 6$ nm).

A PDDLS/Batch dynamic light-scattering instrument (Precision Detectors) was used to collect time-resolved correlation functions during assembly. Each correlation function was accumulated for 15 seconds. The correlator data collection window (10 ms) was spread logarithmically over 190 physical channels. PrecisionDeconvolve software was used to transform the correlation functions into time-resolved R_h distributions $I(t, R_h)$.[35] Number average R_h values, designated R_h^n were computed from the measured R_h distributions by dividing the scattering intensity with the function expected for the relative scattering power of spherical assemblies:

$$R_h^n = \sum_{Rh} [I(R_h)/R_h^5]/(\sum_{Rh} [I(R_h)/R_h^6] \tag{2}$$

Production and purification of viral transcripts

The single-stranded genomic precursors for RNA synthesis reactions were made by activating isolated cores to produce transcripts under the conditions determined earlier for ϕ6.[43] The resulting unlabeled RNAs were extracted with phenol/chloroform/isoamyl alcohol (25:24:1, by vol.) and chloroform/isoamyl alcohol (24:1, v/v). Subsequently, the genomic dsRNA was precipitated with 2.8 M LiCl, and the ssRNA transcripts were precipitated from the resulting supernatant with ammonium acetate and ethanol. Unlabeled genomic ssRNA was subsequently used for coupled *in vitro* packaging, minus strand synthesis and transcription reactions (see the next section).

RNA packaging and polymerase activity of procapsids

A coupled RNA translocation and synthesis activity reaction was assayed in a 50 μl reaction mixture

containing 2.5 mM each ATP, GTP and CTP, 1.5 mM UTP, 50 mM ammonium acetate, 100 mM KCl, 50 mM Tris–HCl (pH 9.0), 1 mM $MgCl_2$, 2.5 mM $MnCl_2$, 5% (w/v) PEG 6000, 5 mM DTT, 0.2 mg/ml of bovine serum albumin, 0.4 unit/μl of RNAsin (Promega). Unlabeled, *in vitro* transcribed ϕ8 ssRNA was used as a precursor for the reactions. The final concentration of RNA was 200 μg/ml. Reactions were started by the addition of 0.2 mg/ml of *in vitro* assembled PC, isolated by centrifugation in a sucrose gradient, and incubated at 30 °C for one hour. *In vitro* assembled PC lacking P4 and *in vivo* assembled PC were used as negative and positive controls. Both plus and minus strand syntheses were detected by incorporation of [α-^{32}P]UTP into the nascent RNA. The mixture was supplemented with 0.1 mCi of [α-^{32}P]UTP (3000 Ci/mmol; Amersham).

To confirm that RNA synthesis occurred within the PC, the packaged particles after plus strand synthesis were sedimented through a 12% (w/v) sucrose cushion and the pellet and top fractions were analyzed for the presence of labeled dsRNA. Top and bottom fractions as well as the products of the initial reaction were treated with 3% (v/v) glycerol, 0.025% (w/v) bromphenol blue, 0.025% (w/v) xylene cyanol, 5 mM EDTA, 3.0 M urea, 0.1% (w/v) SDS and analyzed by agarose gel electrophoresis followed by autoradiography (Fuji Super RX film). In a control experiment, the same packaged particles were disrupted with 0.25% SDS and analyzed as described above.

Analytical methods

Protein concentrations were determined by measuring UV absorption at 280 nm using extinction coefficients based on amino acid compositions, and with Coomassie brilliant blue using bovine serum albumin as a standard.[53]

Particles were negatively stained with 1% (w/v) potassium phosphotungstate (pH 7.4). EM was performed using a Tecnai F20 transmission electron microscope operating at 200 kV. Micrographs of negatively stained particles were recorded on a Gatan CCD camera. Cryo samples were prepared as described[51] and imaged in an Oxford CT 3500 specimen holder at the temperature of liquid nitrogen. Images were recorded under low-dose conditions on S0163 film and developed in full strength D19 for 12 minutes.

The molecular mass of proteins and their complexes was estimated using analytical gel-filtration coupled to a light-scattering flow-cell (PDI2020, Precision detectors) calibrated with bovine serum albumin (67 kDa) and aldolase (158 kDa).

The hydrodynamic properties of individual proteins were determined using a PDDLS/Batch dynamic light-scattering instrument (Precision Detectors). Precision-Deconvolve software was used for correlation function analysis and for computation of R_h distributions.[35] The R_h values (*Z*-average derived from the measured R_h distribution[34]) reported for individual proteins represent averages of at least three independent measurements.

CD spectra were recorded on a Jasco (Tokyo, Japan) J-720 spectropolarimeter using a 2 mm path-length sample cell. Each spectrum was an average of five scans corrected for the contribution of the solvent baseline.

Acknowledgements

Sanna Jaakkola and Dr Leonard Mindich are thanked for providing plasmids pSJ1b and pLM2424, Rabah Soliymani and Dr Marc Baumann for help with mass spectrometry and Pasi Laurinmäki for performing the electron microscopy. The technical assistance of Riitta Tarkiainen is greatly appreciated. Support was provided by the Academy of Finland (172623 to R.T.; 178778 to S.J.B.) and the Finnish Centres of Excellence program (168694, 164298). D.K. was supported by the National Graduate School in Informational and Structural Biology.

References

1. Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, **92**, 291–294.
2. Bamford, D. H. (2000). Virus structures: those magnificent molecular machines. *Curr. Biol.* **10**, R558–R561.
3. Woolford, J. (2002). Chaperoning ribosome assembly. *Mol. Cell.* **10**, 8–10.
4. Bancroft, J. B., Hills, G. J. & Markham, R. (1967). A study of the self-assembly process in a small spherical virus. Formation of organized structures from protein subunits *in vitro*. *Virology*, **31**, 354–379.
5. Bamford, D. H., Gilbert, R. J., Grimes, J. M. & Stuart, D. I. (2001). Macromolecular assemblies: greater than their parts. *Curr. Opin. Struct. Biol.* **11**, 107–113.
6. Gottlieb, P., Strassman, J., Frucht, A., Qiao, X. Y. & Mindich, L. (1991). *In vitro* packaging of the bacteriophage ϕ6 ssRNA genomic precursors. *Virology*, **181**, 589–594.
7. Frilander, M. & Bamford, D. H. (1995). *In vitro* packaging of the single-stranded RNA genomic precursors of the segmented double-stranded RNA bacteriophage ϕ6: the three segments modulate each other's packaging efficiency. *J. Mol. Biol.* **246**, 418–428.
8. Poranen, M. M. & Bamford, D. H. (1999). Packaging and replication regulation revealed by chimeric genome segments of double-stranded RNA bacteriophage ϕ6. *RNA*, **5**, 446–454.
9. Poranen, M. M., Paatero, A. O., Tuma, R. & Bamford, D. H. (2001). Self-assembly of a viral molecular machine from purified protein and RNA constituents. *Mol. Cell.* **7**, 845–854.
10. Poranen, M. M., Pirttimaa, M. J. & Bamford, D. H. (2002). Bacteriophage f6. In *Viral Genome Packaging* (Catalano, C., ed.), Landes Bioscience, Georgetown, TX.
11. Mindich, L., Qiao, X., Qiao, J., Onodera, S., Romantschuk, M. & Hoogstraten, D. (1999). Isolation of additional bacteriophages with genomes of segmented double-stranded RNA. *J. Bacteriol.* **181**, 4505–4508.
12. Vidaver, A. K., Koski, R. K. & Van Etten, J. L. (1973). Bacteriophage ϕ6: a lipid-containing virus of *Pseudomonas phaseolicola*. *J. Virol.* **11**, 799–805.
13. Semancik, J. S., Vidaver, A. K. & Van Etten, J. L. (1973). Characterization of segmented double-helical RNA from bacteriophage ϕ6. *J. Mol. Biol.* **78**, 617–625.
14. Gottlieb, P., Strassman, J., Bamford, D. H. & Mindich, L. (1988). Production of a polyhedral particle in *Escherichia coli* from a cDNA copy of the large

genomic segment of bacteriophage ϕ6. *J. Virol.* **62**, 181–187.
15. Makeyev, E. V. & Bamford, D. H. (2000). Replicase activity of purified recombinant protein P2 of double- stranded RNA bacteriophage ϕ6. *EMBO J.* **19**, 124–133.
16. Paatero, A. O., Mindich, L. & Bamford, D. H. (1998). Mutational analysis of the role of nucleoside triphosphatase P4 in the assembly of the RNA polymerase complex of bacteriophage ϕ6. *J. Virol.* **72**, 10058–10065.
17. Juuti, J. T. & Bamford, D. H. (1997). Protein P7 of phage ϕ6 RNA polymerase complex, acquiring of RNA packaging activity by *in vitro* assembly of the purified protein onto deficient particles. *J. Mol. Biol.* **266**, 891–900.
18. Qiao, X., Qiao, J. & Mindich, L. (1997). Stoichiometric packaging of the three genomic segments of double-stranded RNA bacteriophage ϕ6. *Proc. Natl Acad. Sci. USA*, **94**, 4074–4079.
19. Mindich, L. (1999). Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage ϕ6. *Microbiol. Mol. Biol. Rev.* **63**, 149–160.
20. Bamford, D. H. & Mindich, L. (1980). Electron microscopy of cells infected with nonsense mutants of bacteriophage ϕ6. *Virology*, **107**, 222–228.
21. Butcher, S. J., Dokland, T., Ojala, P. M., Bamford, D. H. & Fuller, S. D. (1997). Intermediates in the assembly pathway of the double-stranded RNA virus ϕ6. *EMBO J.* **16**, 4477–4487.
22. Grimes, J. M., Burroughs, J. N., Gouet, P., Diprose, J. M., Malby, R., Zientara, S. *et al.* (1998). The atomic structure of the bluetongue virus core. *Nature*, **395**, 470–478.
23. Bamford, D., Burnett, R. & Stuart, D. (2002). Evolution of viral structure. *Theor. Popul. Biol.* **61**, 461–470.
24. de Haas, F., Paatero, A. O., Mindich, L., Bamford, D. H. & Fuller, S. D. (1999). A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage ϕ6. *J. Mol. Biol.* **294**, 357–372.
25. Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. (2001). A mechanism for initiating RNA-dependent RNA polymerization. *Nature*, **410**, 235–240.
26. Ikonen, T., Kainov, D. M., Serimaa, R. E., Timmins, P. & Tuma, R. (2003). Localization of replication machinery within the polymerase complex of dsRNA bacteriophage ϕ6. *J. Appl. Crystallog.* In Press.
27. Benevides, J. M., Juuti, J. T., Tuma, R., Bamford, D. H. & Thomas, G. J., Jr (2002). Characterization of subunit-specific interactions in a double-stranded RNA virus: Raman difference spectroscopy of the ϕ6 procapsid. *Biochemistry*, **41**, 11946–11953.
28. Reinisch, K. M., Nibert, M. L. & Harrison, S. C. (2000). Structure of the reovirus core at 3.6 Å resolution. *Nature*, **404**, 960–967.
29. Hill, C. L., Booth, T. F., Prasad, B. V., Grimes, J. M., Mertens, P. P., Sutton, G. C. & Stuart, D. I. (1999). The structure of a cypovirus and the functional organization of dsRNA viruses. *Nature Struct. Biol.* **6**, 565–568.
30. Prasad, B. V., Rothnagel, R., Zeng, C. Q., Jakana, J., Lawton, J. A., Chiu, W. & Estes, M. K. (1996). Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. *Nature*, **382**, 471–473.
31. Hoogstraten, D., Qiao, X., Sun, Y., Hu, A., Onodera, S. & Mindich, L. (2000). Characterization of ϕ8, a bacteriophage containing three double-stranded RNA genomic segments and distantly related to Φ6. *Virology*, **272**, 218–224.
32. Juuti, J. T., Bamford, D. H., Tuma, R. & Thomas, G. J., Jr (1998). Structure and NTPase activity of the RNA-translocating protein (P4) of bacteriophage ϕ6. *J. Mol. Biol.* **279**, 347–359.
33. Prevelige, P. E., Jr, Thomas, D. & King, J. (1993). Nucleation and growth phases in the polymerization of coat protein and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* **64**, 824–835.
34. Johnson, C. S. Jr & Gabriel, D. A. (1994). *Laser Light Scattering*, Dover, New York.
35. Lomakin, A., Benedek, G. B. & Teplow, D. B. (1999). *Monitoring Protein Assembly Using Quasielastic Light Scattering Spectroscopy* Methods in Enzymology, vol. 309, Academic Press, New York pp. 429–459.
36. Ward, C. W. (1993). Progress towards a higher taxonomy of viruses. *Res. Virol.* **144**, 419–453.
37. Mattion, N. M., Bellinzoni, R. C., Blackhall, J. O., Estes, M. K., Gonzalez, S., La Torre, J. L. & Scodeller, E. A. (1990). Genome rearrangements in porcine rotaviruses: biochemical and biological comparisons between a supershort strain and its standard counterpart. *J. Gen. Virol.* **71**, 355–362.
38. Gorziglia, M., Nishikawa, K. & Fukuhara, N. (1989). Evidence of duplication and deletion in super short segment 11 of rabbit rotavirus Alabama strain. *Virology*, **170**, 587–590.
39. Yang, H., Makeyev, E. V. & Bamford, D. H. (2001). Comparison of polymerase subunits from double-stranded RNA bacteriophages. *J. Virol.* **75**, 11088–11095.
40. Frilander, M., Gottlieb, P., Strassman, J., Bamford, D. H. & Mindich, L. (1992). Dependence of minus-strand synthesis on complete genomic packaging in the double-stranded RNA bacteriophage ϕ6. *J. Virol.* **66**, 5013–5017.
41. Gottlieb, P., Strassman, J., Qiao, X. Y., Frucht, A. & Mindich, L. (1990). *In vitro* replication, packaging, and transcription of the segmented double-stranded RNA genome of bacteriophage ϕ6: studies with procapsids assembled from plasmid-encoded proteins. *J. Bacteriol.* **172**, 5774–5782.
42. van Dijk, A. A., Frilander, M. & Bamford, D. H. (1995). Differentiation between minus- and plus-strand synthesis: polymerase activity of dsRNA bacteriophage ϕ6 in an *in vitro* packaging and replication system. *Virology*, **211**, 320–323.
43. Ojala, P. M. & Bamford, D. H. (1995). *In vitro* transcription of the double-stranded RNA bacteriophage ϕ6 is influenced by purine NTPs and calcium. *Virology*, **207**, 400–408.
44. Frilander, M., Poranen, M. & Bamford, D. H. (1995). The large genome segment of dsRNA bacteriophage ϕ6 is the key regulator in the *in vitro* minus and plus strand synthesis. *RNA*, **1**, 510–518.
45. Mindich, L., Qiao, X. & Qiao, J. (1995). Packaging of multiple copies of reduced-size genomic segments by bacteriophage ϕ6. *Virology*, **212**, 213–217.
46. Gottlieb, P., Strassman, J., Qiao, X., Frilander, M., Frucht, A. & Mindich, L. (1992). *In vitro* packaging and replication of individual genomic segments of bacteriophage ϕ6 RNA. *J. Virol.* **66**, 2611–2616.
47. Lawton, J. A., Zeng, C. Q., Mukherjee, S. K., Cohen, J., Estes, M. K. & Prasad, B. V. (1997). Three-dimensional structural analysis of recombinant rotavirus-like particles with intact and amino-terminal-deleted

VP2: implications for the architecture of the VP2 capsid layer. *J. Virol.* **71**, 7353–7360.
48. Lawton, J. A., Estes, M. K. & Prasad, B. V. (1997). Three-dimensional visualization of mRNA release from actively transcribing rotavirus particles. *Nature Struct. Biol.* **4**, 118–121.
49. Gouet, P., Diprose, J. M., Grimes, J. M., Malby, R., Burroughs, J. N., Zientara, S. *et al.* (1999). The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. *Cell*, **97**, 481–490.
50. Tabor, S. (1990). Expression using the T7 RNA polymerase/promoter system. In *Current Protocols in Molecular Biology* (Ausubel, F. A., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds.), Greene Publishing and Wiley-Interscience, New York.
51. Yang, H., Makeyev, E. V., Butcher, S. J., Gaidelyte, A. & Bamford, D. H. (2003). Two distinct mechanisms ensure transcriptional polarity in dsRNA bacteriophages. *J. Virol.* **77**, 1195–1203.
52. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858.
53. Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.
54. Day, L. A. & Mindich, L. (1980). The molecular weight of bacteriophage ϕ6 and its nucleocapsid. *Virology*, **103**, 376–385.
55. Yang, J. T., Wu, C.-S. C. & Martinez, H. M. (1986). *Calculation of Protein Conformation From Circular Dichroism* Methods in Enzymology, vol. 130, Academic Press, New York pp. 208–269.

Edited by M. F. Summers

(Received 12 November 2002; received in revised form 27 February 2003; accepted 5 March 2003)

RNA Biology 10:4, 1–9; April 2013; © 2013 Landes Bioscience

A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence

Alexander Borodavka, Roman Tuma* and Peter G. Stockley*
Astbury Centre for Structural Molecular Biology; Faculty of Biological Sciences; University of Leeds; Leeds, UK

Long RNAs often exist as multiple conformers in equilibrium. For the genomes of single-stranded RNA viruses, one of these conformers must include a compacted state allowing the RNA to be confined within the virion. We have used single molecule fluorescence correlation spectroscopy to monitor the conformations of viral genomes and sub-fragments in the absence and presence of coat proteins. Cognate RNA-coat protein interactions in two model viruses cause a rapid collapse in the hydrodynamic radii of their respective RNAs. This is caused by protein binding at multiple sites on the RNA that facilitate additional protein-protein contacts. The collapsed species recruit further coat proteins to complete capsid assembly with great efficiency and fidelity. The specificity in RNA-coat protein interactions seen at single-molecule concentrations reflects the packaging selectivity seen for such viruses in vivo. This contrasts with many in vitro reassembly measurements performed at much higher concentrations. RNA compaction by coat protein or polycation binding are distinct processes, implying that defined RNA-coat protein contacts are required for assembly.

Keywords: RNA folding, viral genomes, single molecule fluorescence

Submitted: 12/08/12

Revised: 01/30/13

Accepted: 01/31/13

http://dx.doi.org/10.4161/rna.23838

*Correspondence to: Roman Tuma and Peter G. Stockley; Email: r.tuma@leeds.ac.uk and p.g.stockley@leeds.ac.uk

Introduction

Single-stranded (ss) RNA viruses are major pathogens in every kingdom of life. They have mono- or multipartite genomes ranging in length up to > 10 kb, encoding multiple open reading frames. These genomes often serve multiple functions acting as both mRNAs for viral protein expression and templates for RNA-dependent RNA polymerase during replication. During these processes, the RNAs are not in compact conformations. Formation of progeny virions, however, requires that the RNAs be confined in the restricted space formed by assembly of a protective protein shell. The precise mechanism that controls both the conformational changes accompanying virion assembly and the selective packaging of cognate genomes, rather than potential competitor cellular RNAs, has remained vague. The current paradigm, extrapolating from the fact that many viral coat proteins contain positively charged domains and will self-assemble in vitro around non-viral RNAs, assumes that genome packaging is non-sequence specific.[1-3] Our recent results with two simple viruses that infect bacteria and plants overturn this view.[4] Using single molecule fluorescence assembly assays that avoid artifacts due to high protein concentration, we show that packaging is both sequence-specific and two-stage. The first stage is a rapid compaction of the RNA that is required to allow it to fit into the capsid, driven by multiple coat protein-RNA and coat protein-coat protein interactions. The second stage is recruitment of the remaining complement of coat proteins to these partially formed compact complexes. This mechanism mirrors aspects of other RNA folding reactions, such as ribosome assembly, and provides novel insights into the biology of RNA viruses that could be exploited therapeutically.

RNA compaction. Folding into a compact state plays an important role in the function of many RNAs. This has been demonstrated for both short RNAs, such

as ribozymes or riboswitches (< 500 nt long), and also for longer rRNA (16S bacterial rRNA, 1,530 nt). Ribozymes, riboswitches and shorter fragments of rRNAs (e.g., the 5' domain of 16S rRNA)[5] fold in the presence of multivalent cations. Folding of these RNAs usually proceeds through rapid formation of compact structures, which may contain non-native tertiary contacts that are resolved during a subsequent slower phase.[6]

Longer rRNAs also undergo rapid formation of secondary structure and collapse into a compact ensemble but require ribosomal proteins for stabilization of the native fold under physiological salt concentrations.[7] Ribosomal protein assembly onto the largely pre-folded RNA core is a co-operative, step-wise process,[8,9] characterized by a gradual decrease in RNA flexibility upon addition of proteins.[10] Some ribosomal proteins, such as S4 and S7, are largely disordered prior to interaction with the rRNA and may also fold upon assembly. Such co-folding further increases the co-operativity and specificity of assembly.[7,11] Formation of secondary structure and initial compaction of long rRNAs is thought to occur co-transcriptionally in cells and in the later stages is assisted by assembly of ribosomal proteins.[7] Parallel assembly pathways may be utilized[12] to bypass potential roadblocks due to mutations or protein deficiency.[13]

Considerably less is known about the structures of other long RNAs, such as mRNAs, long non-coding (lnc) RNAs and viral genomic RNAs. While local secondary and tertiary structures are important in mRNAs for control of gene expression, processing and stability,[14-16] their overall structures are less well defined. Due to their limited lifetimes and coding functions, they are not required to adopt a uniquely folded structure.

Likewise, viral RNAs were thought not to harbor extensive regions of tertiary structure, especially those which also act as mRNAs (e.g., viruses with positive sense ssRNA genomes). Medium resolution cryo-EM reconstructions have demonstrated, however, that substantial portions of genomic RNA are ordered inside the capsids of many viruses.[17-21] This order includes regions in contact with the coat protein layer as well as regions free of protein. Specific packaging signals (PSs), short conserved sequences/motifs that are recognized by viral coat proteins during assembly, have been identified in many viral RNAs.[22-29] These sequences are thought to facilitate selective packaging of viral genomic RNA. However, many viruses do not exhibit such clearly defined packaging signals and most of those which do can also package non-specific RNA in vitro or cellular mRNA when their proteins are overexpressed from a DNA vector in cells, or even during normal infection cycles.[30]

Since RNA is a polyanion, buffer salt concentration and composition are crucial for folding and condensation, e.g., for ribosomal and other structured RNAs. Many viral coat proteins possess positively charged polypeptide arms or domains which are in direct contact with RNA in the virion and help to neutralize RNA charge. Some viruses also encapsidate polyvalent cations, such as spermidine, to aid charge neutralization.[31-33] As a result of these features, charge neutralization has been thought to be the essential step in viral genome packaging,[1,3,34-36] although such a mechanism cannot easily explain the observed preference for encapsidation of cognate genomes in vivo.

How can a virus selectively package its own RNA inside an infected cell cytoplasm full of heterologous RNAs? One strategy, adopted by dsRNA viruses and some ssRNA viruses, is to sequester the sites of replication and assembly away from the cytoplasm into "virus factories," inclusion bodies or onto membrane surfaces.[37-39] This process requires a level of sophistication and specialized gene products that organize these assemblies, leading to co-localization and direct coupling between RNA replication and packaging.[40] Simpler viruses, with limited coding capacity, including major pathogens, do not have such luxury and replicate in the cytoplasm and, therefore, have to select their own RNAs based on coat protein affinity. Given the apparent paucity of high affinity packaging signals and the relatively high concentration of cellular competitors, affinity alone may be insufficient to achieve specificity. Nature's solution to this problem seems to be the evolution of co-operativity in the packaging process, based on multiple weak RNA-coat protein interactions rather than on recognition of a single high affinity site.[4] A key to observing such co-operativity during co-assembly of RNA and viral coat proteins was to mimic early stages in virus assembly when the coat protein concentration is naturally low. This was achieved using single molecule detection of assembling intermediates, based on fluorescence correlation spectroscopy (FCS).

RNA virus assembly. Coat proteins of RNA viruses, such as bacteriophage MS2[41] and the small plant virus satellite tobacco necrosis virus (STNV) (**Fig. 1**),[42,43] interact with genomic RNA primarily via two mechanisms. The MS2 coat protein recognizes a high affinity, RNA stem-loop within the cognate genome (TR),[44-47] which interacts with both subunits of a coat protein dimer.[41,48-50] This complex initially represses translation of the phage replicase and is thought to nucleate assembly (**Fig. 1**). STNV is an example of a large group of viruses in which the coat proteins interact with RNA, at least in part, using positively charged extended polypeptide arms. In different viral proteins, these can be at the N or C terminus and have been thought to neutralize RNA charge with no sequence specificity. However, in vitro SELEX against the STNV coat protein allowed us to identify aptamers with sequence/structure matches to multiple degenerate potential stem-loop structures positioned throughout the known STNV genomes, suggesting sequence selectivity.[22] The latter has been demonstrated in in vitro reassembly assays.[51] Multiple RNA-coat protein contacts are effective at promoting efficient assembly.[52] Similar mechanisms seem to be adopted by other RNA viruses, including some plant (TCV), animal and human viruses.[23-29]

Many viruses can, however, package non-cognate RNAs in vitro[53-55] or assemble without any RNA,[56] leading to the dominance of a protein-centric assembly model in which coat protein binding neutralizes RNA charge, gradually condensing it to fit into the capsid. This concept cannot account for the highly specific packaging observed for RNA viruses in vivo, perhaps reflecting the difference from in vitro conditions, which usually employ high protein and nucleic acid concentrations.

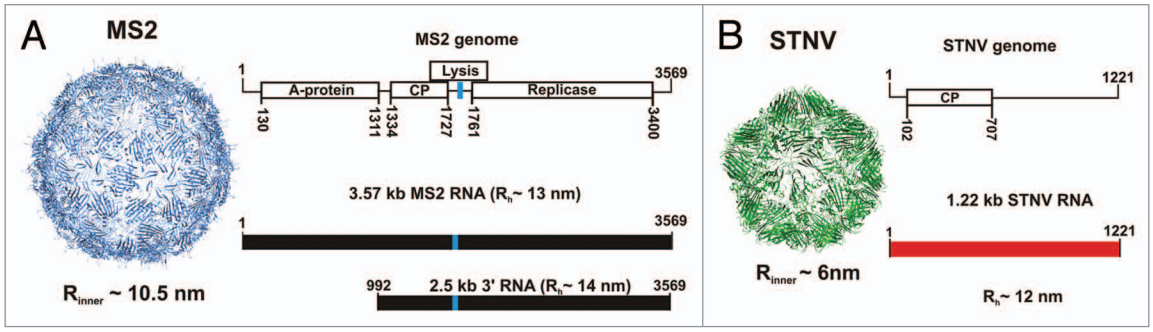


Figure 1. The bacteriophage MS2 and satellite tobacco necrosis virus (STNV) components. (**A**) Structure of the MS2 phage capsid (inner capsid radius, R_{inner}~10.5 nm), its genomic map and a cartoon showing the RNA fragments used for the smFCS experiments described here (below). The location of TR is indicated by a blue stripe. (**B**) Structure of the STNV capsid (inner capsid radius, R_{inner}~6.0 nm), together with its genomic map and corresponding transcript employed in assembly experiments (1.22 kb, shown in red below). Average hydrodynamic radii of the protein-free RNAs in a polyvalent ion-free buffer (see text) are also shown.

In vivo, protein and RNA concentrations build up from scratch during viral infection. In many cases, replication is completed during the early stages of infection and precedes coat protein production. A plausible scenario in the infected cell is that assembly is initiated on viral RNA at coat protein concentrations that are much lower than those used in vitro. Consequently, most of the newly synthesized coat would be incorporated into the growing capsids leaving little free coat protein for non-specific interactions. This scenario begs a number of questions: Can assembly be triggered at low concentrations? How is the viral RNA compacted at such low protein concentrations? Is there any difference between the packaging of cognate vs. non-cognate RNA? Answers to these came from single-molecule assays of RNA packaging during capsid assembly, which are described in the following section.

Single molecule techniques in RNA folding. FCS is a correlation-based method which exploits spontaneous fluctuations in fluorescent signals in order to obtain characteristic time scales (e.g., relaxation times or rates) for molecular processes.[57,58] The fluctuations are due to changes in the number density of fluorescent species within the measurement volume and are related to rates of chemical (e.g., the conversion of a fluorescent to non-fluorescent substrate by quenching) or photo-physical reactions (e.g., the dynamics of triplet state formation, on the ~1 μs time scale). When performed at the single molecule level using a small confocal volume ~1 fL, the number density changes due to molecules diffusing in and out of the detection volume (**Fig. 2**). The resulting fluctuations are then transformed in real time, usually by a specialized hardware correlator, to an autocorrelation function (CF) (**Fig. 2**), and the characteristic times (τ_d) are obtained from fitting the correlation function to a model.[57] At single molecule concentrations (~1 nM), the diffusion represents the longest correlated process and the corresponding diffusion time, which is related to the average time the molecule takes to transverse the confocal volume, is readily identifiable from the CF profile (**Fig. 2**). A small molecule (e.g., fluorescent dye), which has a large diffusion coefficient, will spend a relatively short time in the measured volume and, thus, the fluorescent signal will only be correlated during this short period (~80 μs for an AlexaFluor 488 dye molecule, **Fig. 2**). Viral coat proteins, assembly intermediates, genomic RNAs and capsids diffuse progressively more slowly producing considerably longer correlations in the fluorescent signal (~0.4–1 ms, **Fig. 2**). This provides a way to size selectively labeled molecules in the presence of unlabeled species. Their hydrodynamic radii (R_h) are then computed using calibration of the confocal volume with free dye and Stokes-Einstein equation.[57] For the viral systems described here, these R_h values agree very well with separate estimations based on X-ray structures or mass spectrometry.[41,43,59-61]

FCS assays of viral assembly rely on selective and efficient labeling of RNA

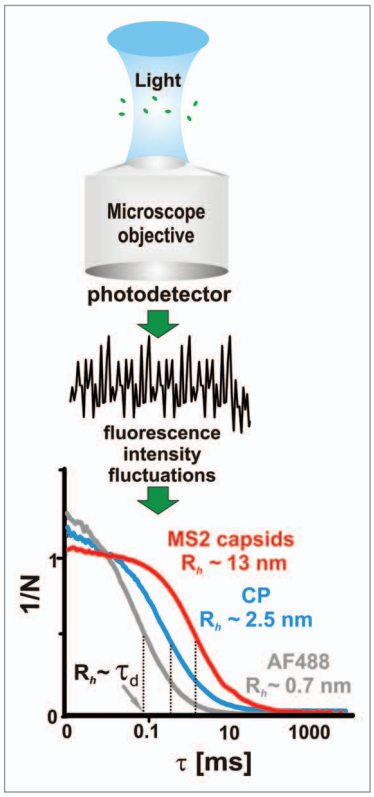


Figure 2. Single-molecule fluorescence correlation spectroscopy. The fluorescent signal is collected from single molecules passing through the confocal volume of the microscope objective. Rapid fluctuations in the fluorescent signal (black trace) are collected by the detector and correlated to obtain an autocorrelation function (CF). Typical CF values for Alexa Fluor 488 dye (gray), MS2 bacteriophage coat protein dimer (blue) and capsid (red) are shown below. The τ_d values are indicated by dotted vertical lines and the corresponding hydrodynamic radii (R_h) are shown. The CF amplitude scales inversely with the average number of molecules in the confocal volume (N) as indicated on the Y axis.

and their cognate interacting proteins. Many strategies have been developed for covalent labeling of proteins, ranging from relatively non-specific lysine-reactive dye conjugates to more selective cysteine-reactive maleimide derivatives and CLICK chemistry.[62-66] Specific covalent labeling of RNA can be achieved at either 5' or 3' termini.[67-71] Two 3'-end strategies were employed in the work described here: (1) end ligation of a dye labeled dA_{10} oligomer; (2) incorporation of 3'-amino adenosine by poly-A-polymerase followed by amine-specific labeling.[4] The advantage of the former is the elimination of chemical labeling steps, while the latter is more effective in situations where the 3' end of the RNA is partially occluded, preventing the incorporation of the bulky oligomer.

Other single molecule techniques, such as Förster resonance energy transfer (smFRET), provide a more detailed view of conformational changes, as elegantly demonstrated for the folding pathway of the *S*-adenosyl methionine regulated riboswitch[72,73] and for protein-assisted folding of telomerase RNA.[74] Unlike FCS, smFRET interrogates local, specific structural details and employs detailed structural information to position the donor and acceptor dyes. In principle, such an approach is also applicable to large RNAs but positioning of the dyes needs to be carefully guided (e.g., by secondary structure probing) in order to avoid interfering with RNA folding and function (e.g., virus assembly).

The single molecule methods discussed here may be useful in many other situations involving large RNA molecules. For instance, it would be an excellent way to investigate the structures and interactions of long non-coding RNAs (e.g., Xist, HOTAIR, NRON).[75-78]

Sizes of protein-free viral RNAs. Given the need for compaction during assembly, it is important to consider the structure of the protein-free viral RNA. All long ssRNAs, with the exception of synthetic homopolymers, will form short stretches of secondary structure by intramolecular base pairing.[79] Raman spectra of MS2 and other viral RNAs suggest that at least 85% of their nucleotides are base-paired,[80-83] and are consistent with structure probing studies.[84-86] This contrasts with early electron microscopy (EM) images of bacteriophage RNAs, which revealed extended conformations with few loops and low secondary structure content. It is now presumed that this appearance was due to disruption of the native structure during sample preparation.[87] Estimation of RNA size and shape in solution has also proved difficult. Analysis of MS2 RNA by SAXS was complicated by self-association due to the high concentrations required.[88,89] Sedimentation velocity measurements of MS2 genomic and subgenomic RNAs under native conditions, however, indicates that these viral RNAs have compact conformations.[60] Attempts have been made to estimate the overall shape and compactness of long RNA molecules.[90] The extensive branching of viral RNAs and the requirement that they fit into the cramped space of a capsid led to the proposal that they are more compact than RNAs with the same base composition but randomized sequences. A recent SAXS study and direct visualization of a few examples of long RNA molecules in frozen solution by cryoEM demonstrated that they are indeed folded into highly branched, relatively compact but elongated structures.[91] However, the limited number of RNA types examined by this method thus far precludes a definitive conclusion on the overall compactness of viral vs. non-viral RNA.

SmFCS, coupled with non-disruptive end-labeling with dye, is an important development. It provides a convenient tool for sizing RNA molecules in extremely dilute solutions in which intermolecular RNA interactions would be expected to be negligible. Hydrodynamic radii estimated from such measurements demonstrate that the solution sizes of STNV and MS2 genomic RNAs, as well as several subgenomic MS2 RNA fragments (**Fig. 1**), are larger than the volume available inside their cognate capsids (R_h values of the genome and a fragment lacking roughly a third of the 5' end-3' RNA, are ~13 and ~14 nm, respectively, compared with R_{inner} ~10.5 nm, for the RNA volume in a capsid). This confirms the need for compaction during assembly. Similar assays with a range of non-viral RNAs have demonstrated that they may be as compact as viral RNA.[4] Hence, the initial size of viral RNA cannot be the sole factor by which the coat protein selects the genome. Are there other properties which might distinguish the cognate RNA from others during packaging?

Cooperative collapse. The answer to this question came from following viral assembly reactions by FCS in real time.[4] The results of the most revealing experiment are summarized in **Figure 3**. When the hydrodynamic radius of labeled MS2 genomic RNA is measured before and after addition of sufficient MS2 coat protein to complete capsid assembly on every RNA (**Fig. 3A**, black line), a large (~30%) and rapid (faster than the experimental dead time, 60 s) collapse in the R_h value is seen. This is followed by an increase in R_h in a second stage of assembly that appears to reflect the completion of capsid assembly from pre-compacted, cognate coat protein-RNA complexes. In the EM, these have the appearance of partially formed shells. Under the same conditions, no collapse occurs with STNV genomic RNA mixed with MS2 coat protein (**Fig. 3A**, red line). In contrast, when STNV coat protein was used, STNV RNA undergoes a collapse (**Fig. 3B**, red line). Hydrodynamic collapse is therefore specific to viral RNAs making cognate interactions with their respective coat proteins. For MS2, a mutant coat protein that binds its RNA normally but is deficient in protein-protein interactions fails to elicit the collapse, showing that both interactions are required for this effect and that the collapsed state is not simply a random complex of protein and RNA. This result can only be explained if there are multiple, correctly placed protein binding sites, i.e., packaging signals (PSs), along the viral RNAs. No collapse was seen with non-viral RNAs, or for non-cognate viral RNAs.[4] Interestingly, sub-genomic RNA fragments retain the property of CP-induced collapse, indicating that the co-operativity extends throughout the RNA. Perhaps surprisingly, given the above, all the RNAs tested in these assays trigger assembly of capsid-like aggregates. However, only cognate interactions yield capsids of the correct size and symmetry (T = 1 for STNV and T = 3 for MS2), whereas non-cognate assembly reactions are relatively

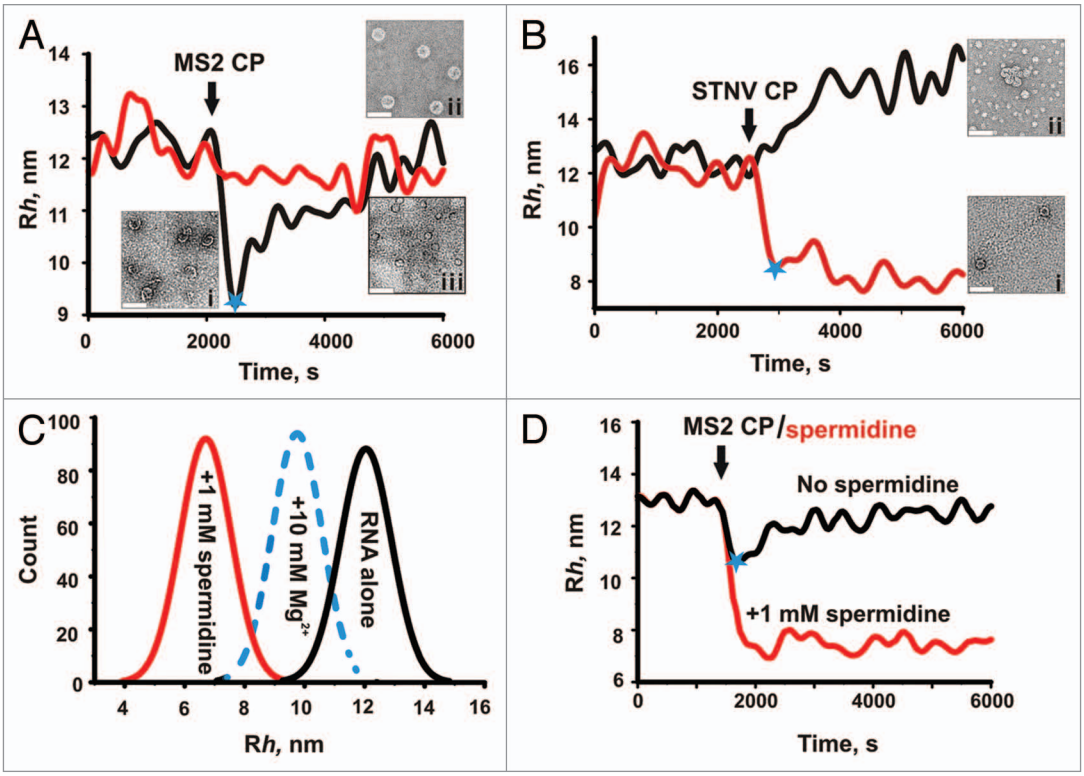


Figure 3. Two-stage assembly of cognate viral RNAs. Time-resolved changes in the apparent R_h of MS2 (black) and STNV (red) genomic RNAs are shown before and after addition (black arrow) of stoichiometric amounts of MS2 (**A**) or STNV (**B**) coat proteins, i.e., sufficient protein to allow each RNA to form a complete capsid. Blue stars denote the end of the respective compaction stages for interactions between cognate molecules. For MS2, this is followed by a slower increase in R_h correlated with formation of the capsids with high yield and fidelity. Capsid formation may already be complete for STNV following stage 1. Non-cognate interactions do not lead to collapse but do lead to inefficient formation of mostly aberrant aggregates. Electron micrographs of negatively stained assembly reactions at defined points in the pathway are shown. (**A**) i, assembly intermediates observed ~1 min after addition of MS2 CP to MS2 RNA; ii, $T = 3$ MS2 capsids present at the end of stage 2, assembled with MS2 RNA; iii, aberrant assembly products and aggregates formed by co-assembly of MS2 CP and STNV RNAs (scale bars 50 nm). (**B**) i, $T = 1$ STNV capsids at the end of the assembly reaction and ii, aberrant aggregates formed by co-assembly of STNV CP and MS2 genomic RNA. (**C**) Effects of multivalent cations on the apparent R_h of the MS2 3' RNA sub-genomic fragment (**Fig. 1**) shown as size (R_h) distributions. RNA in a monovalent buffer is shown in black, followed by separate reactions in which divalent (Mg^{2+}, 10 mM, blue) or trivalent (spermidine, 1 mM, red) ions were added. (**D**) Condensation by counterions or coat protein subunits are not equivalent processes. The MS2 3' RNA undergoes condensation upon addition of 1 mM spermidine. Addition of 200 nM MS2 CP_2 to this sample (red) or one in the monovalent buffer (black) has no effect or results in compaction and then recovery, respectively. The latter sample produces capsids, whereas no assembly occurs under these conditions in the presence of spermidine unless the coat protein concentration is raised significantly (not shown).

low yield and produce a high proportion of misassembled species.

Given the roles of multivalent cations in RNA folding, and the paradigm of electrostatically controlled assembly, a question arises of whether simple charge neutralization, e.g., by magnesium ions or spermidine, could also trigger RNA collapse? The data shown in **Figure 3C** suggest that these cations do indeed collapse RNA conformations to approximately the same (Mg^{2+}) or even greater (spermidine) degrees than binding of coat proteins. Interestingly, once the RNA is compacted by spermidine, it is no longer a good substrate for capsid assembly, although increasing the CP concentrations significantly can reverse this effect (**Fig. 3D**). Thus, it seems that while both coat protein binding and electrostatic neutralization are sufficient to compact the viral RNA, they are not equivalent processes on the pathway to capsid assembly. The role(s) of electrostatics in virus assembly may therefore be less prominent than previously thought.[1,34]

The requirement for protein-protein interactions to drive the RNA collapse implies that assembly is in fact a cooperative process. Sub-stoichiometric (with respect to capsid) amounts of coat protein can also cause the full collapse, implying a nucleated effect governed by coat protein-RNA packaging signal affinities. Putative packaging signals for both MS2 and STNV have now been identified and consist of stem-loops dispersed throughout their genomes. These have only a minimal consensus recognition motif (ref. 22; Dykeman, submitted), explaining why they have not been identified previously. The collapse is rapid, occurring within the dead time of the FCS instrument (60 s per CF). This is consistent with previous solution structure probing of the MS2 genome that suggests that many of the predicted PSs should be present in the RNA and, thus, able to bind coat proteins immediately.[92,93] The first stage of assembly resembles the millisecond compaction of smaller RNAs by Mg^{2+}, previously observed using

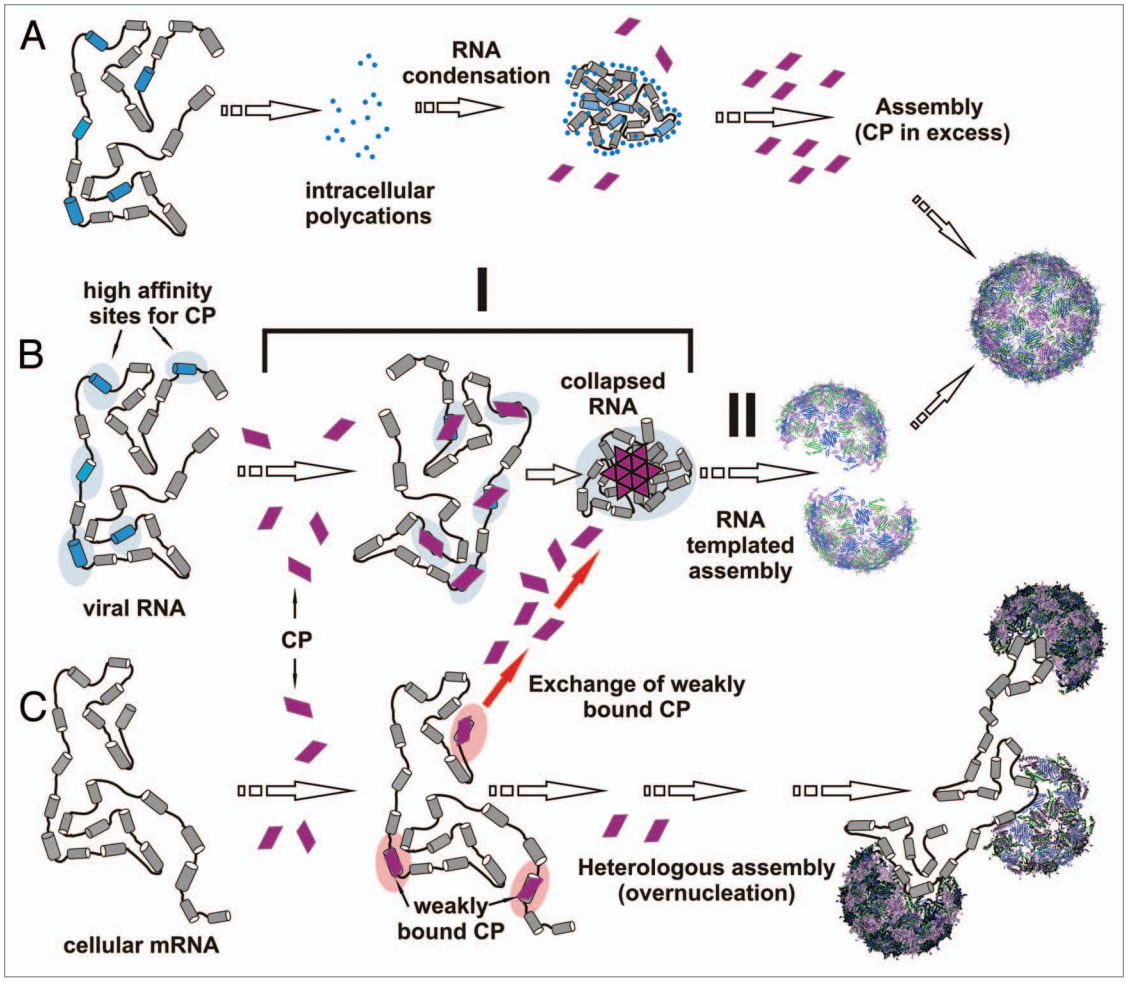


Figure 4. The two-stage assembly mechanism for viral RNA. Our results suggest that there are three forms of RNA condensation/compaction in the context of virus assembly. (**A**) Non-specific condensation of viral RNA by multivalent ions (Mg^{2+}, spermidine) inhibits assembly by coat proteins (middle). Only at very high CP concentrations (> 5 μM) is this block overcome (right), showing that simple electrostatic condensation is not on pathway to capsid formation. (**B**) A two-stage mechanism of assembly in which CPs first bind to cognate RNAs displaying multiple packaging sites (blue segments), distributed throughout the viral genome to facilitate protein-protein interactions, thus mediating a rapid RNA collapse (stage I). The collapse is followed by cooperative recruitment of additional CP subunits (stage II), even at low concentration (< 1 μM), to complete capsid assembly. (**C**) Assembly with non-cognate (cellular) RNAs leads to weak interactions without observable initial RNA collapse. Dissociation of coat proteins from these complexes allows them to be captured by the cognate assembly pathway (red arrows, middle). A low yield alternative pathway occurs when non-specifically bound coat proteins nucleate assembly on cellular RNA. Since the coat protein binding sites are not correctly positioned, these RNAs do not collapse and there can be multiple nucleation events leading to misassembled and multishell structures. Such pathways explain the assembly of non-cognate RNAs in vitro at relatively high coat protein concentrations.

SAXS.[94] Genomic RNA compaction by a sub-stoichiometric amount of cowpea chlorotic mottle virus (CCMV) coat protein has also been observed.[95] However, that process was slow and indiscriminate, compacting both viral as well as non-viral RNA, perhaps because the assays were at higher concentrations than those reported here.

A two-stage cooperative mechanism of selective RNA packaging. The new findings challenge the prevailing view that RNA is gradually condensed during capsid polymerization. While non-cognate (e.g., cellular) RNA may be gradually condensed, the cognate viral RNA is rapidly and cooperatively collapsed (c.f. pathways B and C in **Fig.** 4). The collapsed state is relatively stable since it is formed by multiple RNA-CP and CP-CP interactions. The affinities of individual PS-CP complexes within the collapsed state could be very low but are augmented by their relative locations and the protein-protein interactions that they promote. The stability of this nucleation complex facilitates recruitment of the additional coat proteins required to complete the capsid during a second stage of assembly that is protein concentration dependent.

In contrast, coat proteins will bind to non-cognate RNAs at individual stem-loops that resemble cognate packaging sites, but these will be randomly distributed and, thus, unable to support co-operative collapse. Individual coat protein-RNA complexes are therefore likely to dissociate without nucleating assembly. This provides an explanation for the effective discrimination between the cognate and non-cognate RNAs in vivo. Coat proteins binding and dissociating from sites

on cellular RNAs will eventually be captured by the stable nucleation complexes and used to complete capsids (**Fig. 4**, red arrows). The energetic cost of genome confinement in these viruses is paid by the free energy of coat protein-packaging signal interactions.

Simple charge neutralization and condensation by multivalent cations produces collapsed RNAs that assemble poorly at low coat protein concentrations (nanomolar). Since this block can be overcome at high (micromolar) coat protein concentrations (not shown), the cations may simply bind to the packaging sites and compete with coat protein or act indirectly by folding the RNA into structures in which the packaging sites are obscured (**Fig. 4A**). Only high coat protein concentration can overcome this effect and complete assembly. This again highlights the role of multiple packaging sites which are accessible throughout the viral genome and distributed in a way to promote cooperative assembly of the nucleation complex.

Conclusions

The RNAs of ssRNA viruses seem to have evolved multiple packaging signals that are dispersed throughout their genomes facilitating cooperative assembly of an extended nucleation complex that simultaneously compacts the RNA during the first stage of assembly. In the second stage, the collapsed nucleation complex is efficiently encapsidated, effectively outcompeting protein binding to cellular RNA. An analogy of this process is the idea of packing a suitcase. In this instance, the suitcase (capsid) folds its own contents (the viral RNA). Since we see the same phenomenon in two viruses drawn from widely different families, one of which has an extremely widespread coat protein fold,[96] it is highly likely that similar conserved assembly mechanisms exist in many viruses in this class. Note the molecular details driving the hydrodynamic collapse can vary widely in the different cases. Some viruses, especially animal viruses, appear only to package nascent genomes into progeny virions. This behavior could be due to the need for kinetic folding of the PSs in those cases. Even though the genome would not collapse in the same way, virion assembly needs to compensate for the entropic cost of confining its RNA by making the favorable CP-PS contacts. Interrupting the interactions between coat proteins and their cognate packaging signals would inhibit these processes and may provide a novel target for antivirals. Given the co-operative nature of the assembly mechanism described here, such drugs may be more difficult to escape by mutation.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgements

A.B. thanks the Wellcome Trust for support of his postgraduate studentship (089310/Z/09/Z) and support of the facilities used (062164) (090932/Z/09/Z). Single molecule work within the Astbury Centre has been supported by The University of Leeds.

References

1. Belyi VA, Muthukumar M. Electrostatic origin of the genome packing in viruses. Proc Natl Acad Sci USA 2006; 103:17174-8; PMID:17090672; http://dx.doi.org/10.1073/pnas.0608311103.
2. Forrey C, Muthukumar M. Electrostatics of capsid-induced viral RNA organization. J Chem Phys 2009; 131:105101-9; http://dx.doi.org/10.1063/1.3216550.
3. Devkota B, Petrov AS, Lemieux S, Boz MB, Tang L, Schneemann A, et al. Structural and electrostatic characterization of pariacoto virus: implications for viral assembly. Biopolymers 2009; 91:530-8; PMID:19226622; http://dx.doi.org/10.1002/bip.21168.
4. Borodavka A, Tuma R, Stockley PG. Evidence that viral RNAs have evolved for efficient, two-stage packaging. Proc Natl Acad Sci USA 2012; 109:15769-74; PMID:23019360; http://dx.doi.org/10.1073/pnas.1204357109.
5. Adilakshmi T, Ramaswamy P, Woodson SA. Protein-independent folding pathway of the 16S rRNA 5' domain. J Mol Biol 2005; 351:508-19; PMID:16023137; http://dx.doi.org/10.1016/j.jmb.2005.06.020.
6. Russell R, Millett IS, Tate MW, Kwok LW, Nakatani B, Gruner SM, et al. Rapid compaction during RNA folding. Proc Natl Acad Sci USA 2002; 99:4266-71; PMID:11929997; http://dx.doi.org/10.1073/pnas.072589599.
7. Woodson SA. RNA folding pathways and the self-assembly of ribosomes. Acc Chem Res 2011; 44:1312-9; PMID:21714483; http://dx.doi.org/10.1021/ar2000474.
8. Traub P, Nomura M. Structure and function of Escherichia coli ribosomes. VI. Mechanism of assembly of 30 s ribosomes studied in vitro. J Mol Biol 1969; 40:391-413; PMID:4903714; http://dx.doi.org/10.1016/0022-2836(69)90161-2.
9. Traub P, Nomura M. Structure and function of E. coli ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. Proc Natl Acad Sci USA 1968; 59:777-84; PMID:4868216; http://dx.doi.org/10.1073/pnas.59.3.777.
10. Stagg SM, Mears JA, Harvey SC. A structural model for the assembly of the 30S subunit of the ribosome. J Mol Biol 2003; 328:49-61; PMID:12683996; http://dx.doi.org/10.1016/S0022-2836(03)00174-8.
11. Williamson JR. Proteins that bind RNA and the labs who love them. Nat Struct Biol 2001; 8:390-1; PMID:11323709; http://dx.doi.org/10.1038/87540.
12. Talkington MW, Siuzdak G, Williamson JR. An assembly landscape for the 30S ribosomal subunit. Nature 2005; 438:628-32; PMID:16319883; http://dx.doi.org/10.1038/nature04261.
13. Bubunenko M, Korepanov A, Court DL, Jagannathan I, Dickinson D, Chaudhuri BR, et al. 30S ribosomal subunits can be assembled in vivo without primary binding ribosomal protein S15. RNA 2006; 12:1229-39; PMID:16682557; http://dx.doi.org/10.1261/rna.2262106.
14. Bag J, Bhattacharjee RB. Multiple levels of post-transcriptional control of expression of the poly (A)-binding protein. RNA Biol 2010; 7:5-12; PMID:20009508; http://dx.doi.org/10.4161/rna.7.1.10256.
15. Li Y, Kiledjian M. Regulation of mRNA decapping. Wiley Interdiscip Rev RNA 2010; 1:253-65; PMID:21935889; http://dx.doi.org/10.1002/wrna.15.
16. Tuerk C, Gauss P, Thermes C, Groebe DR, Gayle M, Guild N, et al. CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. Proc Natl Acad Sci USA 1988; 85:1364-8; PMID:2449689; http://dx.doi.org/10.1073/pnas.85.5.1364.
17. Toropova K, Basnak G, Twarock R, Stockley PG, Ranson NA. The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. J Mol Biol 2008; 375:824-36; PMID:18048058; http://dx.doi.org/10.1016/j.jmb.2007.08.067.
18. Toropova K, Stockley PG, Ranson NA. Visualising a viral RNA genome poised for release from its receptor complex. J Mol Biol 2011; 408:408-19; PMID:21376055; http://dx.doi.org/10.1016/j.jmb.2011.02.040.
19. Bakker SE, Ford RJ, Barker AM, Robottom J, Saunders K, Pearson AR, et al. Isolation of an asymmetric RNA uncoating intermediate for a single-stranded RNA plant virus. J Mol Biol 2012; 417:65-78; PMID:22306464; http://dx.doi.org/10.1016/j.jmb.2012.01.017.
20. Tang L, Johnson KN, Ball LA, Lin T, Yeager M, Johnson JE. The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. Nat Struct Biol 2001; 8:77-83; PMID:11135676; http://dx.doi.org/10.1038/83089.
21. Larson SB, Koszelak S, Day J, Greenwood A, Dodds JA, McPherson A. Double-helical RNA in satellite tobacco mosaic virus. Nature 1993; 361:179-82; PMID:8421525; http://dx.doi.org/10.1038/361179a0.
22. Bunka DHJ, Lane SW, Lane CL, Dykeman EC, Ford RJ, Barker AM, et al. Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus: implications for the assembly of a T=1 capsid. J Mol Biol 2011; 413:51-65; PMID:21839093; http://dx.doi.org/10.1016/j.jmb.2011.07.063.
23. D'Souza V, Summers MF. How retroviruses select their genomes. Nat Rev Microbiol 2005; 3:643-55; PMID:16064056; http://dx.doi.org/10.1038/nrmicro1210.
24. Hutchinson EC, von Kirchbach JC, Gog JR, Digard P. Genome packaging in influenza A virus. J Gen Virol 2010; 91:313-28; PMID:19955561; http://dx.doi.org/10.1099/vir.0.017608-0.

25. Kim DY, Firth AE, Atasheva S, Frolova EI, Frolov I. Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. J Virol 2011; 85:8022-36; PMID:21680508; http://dx.doi.org/10.1128/JVI.00644-11.
26. Lever A, Gottlinger H, Haseltine W, Sodroski J. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. J Virol 1989; 63:4085-7; PMID:2760989.
27. Qu F, Morris TJ. Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. J Virol 1997; 71:1428-35; PMID:8995668.
28. Rao AL. Genome packaging by spherical plant RNA viruses. Annu Rev Phytopathol 2006; 44:61-87; PMID:16480335; http://dx.doi.org/10.1146/annurev.phyto.44.070505.143334.
29. Terasaki K, Murakami S, Lokugamage KG, Makino S. Mechanism of tripartite RNA genome packaging in Rift Valley fever virus. Proc Natl Acad Sci USA 2011; 108:804-9; PMID:21187405; http://dx.doi.org/10.1073/pnas.1013155108.
30. Routh A, Domitrovic T, Johnson JE. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proc Natl Acad Sci USA 2012; 109:1907-12; PMID:22308402; http://dx.doi.org/10.1073/pnas.1116168109.
31. Cohen SS, Greenberg ML. Spermidine, an intrinsic component of turnip yellow mosaic virus. Proc Natl Acad Sci USA 1981; 78:5470-4; PMID:6946484; http://dx.doi.org/10.1073/pnas.78.9.5470.
32. Raina A, Tuomi K, Mäntyjärvi R. Roles of polyamines in the replication of animal viruses. Med Biol 1981; 59:428-32; PMID:6279979.
33. Nickerson KW, Lane LC. Polyamine content of several RNA plant viruses. Virology 1977; 81:455-9; PMID:898668; http://dx.doi.org/10.1016/0042-6822(77)90160-X.
34. Hagan MF. A theory for viral capsid assembly around electrostatic cores. J Chem Phys 2009; 130:114902; PMID:19317561; http://dx.doi.org/10.1063/1.3086041.
35. Cadena-Nava RD, Hu Y, Garmann RF, Ng B, Zelikin AN, Knobler CM, et al. Exploiting fluorescent polymers to probe the self-assembly of virus-like particles. J Phys Chem B 2011; 115:2386-91; PMID:21338131; http://dx.doi.org/10.1021/jp1094118.
36. van der Schoot P, Bruinsma R. Electrostatics and the assembly of an RNA virus. Phys Rev E Stat Nonlin Soft Matter Phys 2005; 71:061928; PMID:16089786; http://dx.doi.org/10.1103/PhysRevE.71.061928.
37. Wang RY, Nagy PD. Tomato bushy stunt virus co-opts the RNA-binding function of a host metabolic enzyme for viral genomic RNA synthesis. Cell Host Microbe 2008; 3:178-87; PMID:18329617; http://dx.doi.org/10.1016/j.chom.2008.02.005.
38. Novoa RR, Calderita G, Arranz R, Fontana J, Granzow H, Risco C. Virus factories: associations of cell organelles for viral replication and morphogenesis. Biol Cell 2005; 97:147-72; PMID:15656780; http://dx.doi.org/10.1042/BC20040058.
39. Nugent CI, Johnson KL, Sarnow P, Kirkegaard K. Functional coupling between replication and packaging of poliovirus replicon RNA. J Virol 1999; 73:427-35; PMID:9847348.
40. Seo JK, Kwon SJ, Rao ALN. A physical interaction between viral replicase and capsid protein is required for genome-packaging specificity in an RNA virus. J Virol 2012; 86:6210-21; PMID:22438552; http://dx.doi.org/10.1128/JVI.07184-11.
41. Valegård K, Liljas L, Fridborg K, Unge T. The three-dimensional structure of the bacterial virus MS2. Nature 1990; 345:36-41; PMID:2330049; http://dx.doi.org/10.1038/345036a0.
42. Montelius I, Liljas L, Unge T. Structure of EDTA-treated satellite tobacco necrosis virus at pH 6.5. J Mol Biol 1988; 201:353-63; PMID:3138417; http://dx.doi.org/10.1016/0022-2836(88)90143-X.
43. Lane SW, Dennis CA, Lane CL, Trinh CH, Rizkallah PJ, Stockley PG, et al. Construction and crystal structure of recombinant STNV capsids. J Mol Biol 2011; 413:41-50; PMID:21839089; http://dx.doi.org/10.1016/j.jmb.2011.07.062.
44. Stockley PG, Stonehouse NJ, Murray JB, Goodman STS, Talbot SJ, Adams CJ, et al. Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. Nucleic Acids Res 1995; 23:2512-8; PMID:7543200; http://dx.doi.org/10.1093/nar/23.13.2512.
45. LeCuyer KA, Behlen LS, Uhlenbeck OC. Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. Biochemistry 1995; 34:10600-6; PMID:7544616; http://dx.doi.org/10.1021/bi00033a035.
46. Beckett D, Wu HN, Uhlenbeck OC. Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. J Mol Biol 1988; 204:939-47; PMID:3221401; http://dx.doi.org/10.1016/0022-2836(88)90053-8.
47. Johansson HE, Liljas L, Uhlenbeck OC. RNA recognition by the MS2 phage coat protein. Semin Virol 1997; 8:176-85; http://dx.doi.org/10.1006/smvy.1997.0120.
48. Convery MA, Rowsell S, Stonehouse NJ, Ellington AD, Hirao I, Murray JB, et al. Crystal structure of an RNA aptamer-protein complex at 2.8 A resolution. Nat Struct Biol 1998; 5:133-9; PMID:9461079; http://dx.doi.org/10.1038/nsb0298-133.
49. Gell C, Sabir T, Westwood J, Rashid A, Smith DAM, Harris SA, et al. Single-molecule fluorescence resonance energy transfer assays reveal heterogeneous folding ensembles in a simple RNA stem-loop. J Mol Biol 2008; 384:264-78; PMID:18805425; http://dx.doi.org/10.1016/j.jmb.2008.08.088.
50. Horn WT, Tars K, Grahn E, Helgstrand C, Baron AJ, Lago H, et al. Structural basis of RNA binding discrimination between bacteriophages Qbeta and MS2. Structure 2006; 14:487-95; PMID:16531233; http://dx.doi.org/10.1016/j.str.2005.12.006.
51. Ford RJ, Barker AM, Bakker SE, Coutts RH, Ranson NA, Phillips SEV, et al. Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of Satellite Tobacco Necrosis Virus coat protein. J Mol Biol 2013; In press; PMID:23318955; http://dx.doi.org/10.1016/j.jmb.2013.01.004.
52. Elsawy KM, Caves LS, Twarock R. The impact of viral RNA on the association rates of capsid protein assembly: bacteriophage MS2 as a case study. J Mol Biol 2010; 400:935-47; PMID:20562027; http://dx.doi.org/10.1016/j.jmb.2010.05.037.
53. Comas-Garcia M, Cadena-Nava RD, Rao AL, Knobler CM, Gelbart WM. In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. J Virol 2012; 86:12271-82; PMID:22951822; http://dx.doi.org/10.1128/JVI.01695-12.
54. Annamalai P, Apte S, Wilkens S, Rao AL. Deletion of highly conserved arginine-rich RNA binding motif in cowpea chlorotic mottle virus capsid protein results in virion structural alterations and RNA packaging constraints. J Virol 2005; 79:3277-88; PMID:15731222; http://dx.doi.org/10.1128/JVI.79.6.3277-3288.2005.
55. Annamalai P, Rao AL. Dispensability of 3' tRNA-like sequence for packaging cowpea chlorotic mottle virus genomic RNAs. Virology 2005; 332:650-8; PMID:15680430; http://dx.doi.org/10.1016/j.virol.2004.12.009.
56. Adolph KW, Butler PJ. Studies on the assembly of a spherical plant virus. I. States of aggregation of the isolated protein. J Mol Biol 1974; 88:327-41; PMID:4452998; http://dx.doi.org/10.1016/0022-2836(74)90485-9.
57. Gell C, Brockwell D, Smith A. Handbook of Single Molecule Fluorescence Spectroscopy. Oxford: Oxford University Press, 2006.
58. Rigler R, Elson ES, eds. Fluorescence Correlation Spectroscopy. Berlin: Springer, 2001.
59. Rolfsson O. The roles of MS2 RNA in MS2 capsid assembly. Faculty of Biological Sciences. University of Leeds, 2009:176.
60. Rolfsson O, Toropova K, Ranson NA, Stockley PG. Mutually-induced conformational switching of RNA and coat protein underpins efficient assembly of a viral capsid. J Mol Biol 2010; 401:309-22; PMID:20684044; http://dx.doi.org/10.1016/j.jmb.2010.05.058.
61. Stockley PG, Rolfsson O, Thompson GS, Basnak G, Francese S, Stonehouse NJ, et al. A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. J Mol Biol 2007; 369:541-52; PMID:17434527; http://dx.doi.org/10.1016/j.jmb.2007.03.020.
62. Hermanson GT, Hermanson GT. Bioconjugate techniques. Bioconjugate techniques. San Diego: Academic Press, 1996:785.
63. Fauster K, Hartl M, Santner T, Aigner M, Kreutz C, Bister K, et al. 2'-Azido RNA, a versatile tool for chemical biology: synthesis, X-ray structure, siRNA applications, click labeling. ACS Chem Biol 2012; 7:581-9; PMID:22273279; http://dx.doi.org/10.1021/cb200510k.
64. Ishizuka T, Kimoto M, Sato A, Hirao I. Site-specific functionalization of RNA molecules by an unnatural base pair transcription system via click chemistry. Chem Commun (Camb) 2012; 48:10835-7; PMID:23032097; http://dx.doi.org/10.1039/c2cc36293g.
65. Rao H, Sawant AA, Tanpure AA, Srivatsan SG. Posttranscriptional chemical functionalization of azide-modified oligoribonucleotides by bioorthogonal click and Staudinger reactions. Chem Commun (Camb) 2012; 48:498-500; PMID:22006199; http://dx.doi.org/10.1039/c1cc15659d.
66. Winz ML, Samanta A, Benzinger D, Jäschke A. Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. Nucleic Acids Res 2012; 40:e78; PMID:22344697; http://dx.doi.org/10.1093/nar/gks062.
67. Vauléon S, Ivanov SA, Gwiazda S, Müller S. Site-specific fluorescent and affinity labelling of RNA by using a small engineered twin ribozyme. Chembiochem 2005; 6:2158-62; PMID:16276501; http://dx.doi.org/10.1002/cbic.200500215.
68. Vilfan ID, Kamping W, van den Hout M, Candelli A, Hage S, Dekker NH. An RNA toolbox for single-molecule force spectroscopy studies. Nucleic Acids Res 2007; 35:6625-39; PMID:17905817; http://dx.doi.org/10.1093/nar/gkm585.
69. Logsdon N, Lee CG, Harper JW. Selective 5' modification of T7 RNA polymerase transcripts. Anal Biochem 1992; 205:36-41; PMID:1443558; http://dx.doi.org/10.1016/0003-2697(92)90575-R.
70. Kinoshita Y, Nishigaki K, Husimi Y. Fluorescence-, isotope- or biotin-labeling of the 5'-end of single-stranded DNA/RNA using T4 RNA ligase. Nucleic Acids Res 1997; 25:3747-8; PMID:9278501; http://dx.doi.org/10.1093/nar/25.18.3747.
71. Smith GJ, Sosnick TR, Scherer NF, Pan T. Efficient fluorescence labeling of a large RNA through oligonucleotide hybridization. RNA 2005; 11:234-9; PMID:15613536; http://dx.doi.org/10.1261/rna.7180305.

72. Heppell B, Blouin S, Dussault AM, Mulhbacher J, Ennifar E, Penedo JC, et al. Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. Nat Chem Biol 2011; 7:384-92; PMID:21532599; http://dx.doi.org/10.1038/nchembio.563.
73. Eschbach SH, St-Pierre P, Penedo JC, Lafontaine DA. Folding of the SAM-I riboswitch: a tale with a twist. RNA Biol 2012; 9:535-41; PMID:22336759; http://dx.doi.org/10.4161/rna.19648.
74. Stone MD, Mihalusova M, O'connor CM, Prathapam R, Collins K, Zhuang X. Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. Nature 2007; 446:458-61; PMID:17322903; http://dx.doi.org/10.1038/nature05600.
75. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem 2012; 81:145-66; PMID:22663078; http://dx.doi.org/10.1146/annurev-biochem-051410-092902.
76. Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. Nat Rev Genet 2011; 12:136-49; PMID:21245830; http://dx.doi.org/10.1038/nrg2904.
77. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 2010; 464:1071-6; PMID:20393566; http://dx.doi.org/10.1038/nature08975.
78. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007; 129:1311-23; PMID:17604720; http://dx.doi.org/10.1016/j.cell.2007.05.022.
79. Gralla J, Steitz JA, Crothers DM. Direct physical evidence for secondary structure in an isolated fragment of R17 bacteriophage mRNA. Nature 1974; 248:204-8; PMID:4819414; http://dx.doi.org/10.1038/248204a0.
80. Thomas GJ Jr., Prescott B, McDonald-Ordzie PE, Hartman KA. Studies of virus structure by laser-Raman spectroscopy. II. MS2 phage, MS2 capsids and MS2 RNA in aqueous solutions. J Mol Biol 1976; 102:103-24; PMID:1271459; http://dx.doi.org/10.1016/0022-2836(76)90076-0.
81. Thomas GJ, Hartman KA. Laser-Excited Raman-Spectra of R17 Phage and R17 RNA. Fed Proc 1974; 33:1372-3.
82. Thomas GJ Jr., Hartman KA. Raman studies of nucleic acids. 8. Estimation of RNA secondary structure from Raman scattering by phosphate-group vibrations. Biochim Biophys Acta 1973; 312:311-32; PMID:4579230; http://dx.doi.org/10.1016/0005-2787(73)90376-6.
83. Hartman KA, Clayton N, Thomas GJ Jr. Studies of viral structure by Raman spectroscopy. I. R17 virus and R17 RNA. Biochem Biophys Res Commun 1973; 50:942-9; PMID:4569889; http://dx.doi.org/10.1016/0006-291X(73)91336-3.
84. Beekwilder J, Nieuwenhuizen R, Poot R, van Duin J. Secondary structure model for the first three domains of Q beta RNA. Control of A-protein synthesis. J Mol Biol 1996; 256:8-19; PMID:8609616; http://dx.doi.org/10.1006/jmbi.1996.0064.
85. Skripkin EA, Adhin MR, de Smit MH, van Duin J. Secondary structure of the central region of bacteriophage MS2 RNA. Conservation and biological significance. J Mol Biol 1990; 211:447-63; PMID:2407856; http://dx.doi.org/10.1016/0022-2836(90)90364-R.
86. Schmidt BF, Berkhout B, Overbeek GP, van Strien A, van Duin J. Determination of the RNA secondary structure that regulates lysis gene expression in bacteriophage MS2. J Mol Biol 1987; 195:505-16; PMID:3656423; http://dx.doi.org/10.1016/0022-2836(87)90179-3.
87. Jacobson AB. Studies on secondary structure of single-stranded RNA from bacteriophage MS2 by electron microscopy. Proc Natl Acad Sci USA 1976; 73:307-11; PMID:1061134; http://dx.doi.org/10.1073/pnas.73.2.307.
88. Ribitsch G, De Clercq R, Folkhard W, Zipper P, Schurz J, Clauwaert J. Small-angle X-ray and light scattering studies on the influence of Mg2+ ions on the structure of the RNA from bacteriophage MS2. Z Naturforsch C 1985; 40:234-41; PMID:4002831.
89. Zipper P, Folkhard W. A small-angle x-ray scattering investigation on the structure of the RNA from bacteriophage MS2. FEBS Lett 1975; 56:283-7; PMID:1157947; http://dx.doi.org/10.1016/0014-5793(75)81110-0.
90. Yoffe AM, Prinsen P, Gopal A, Knobler CM, Gelbart WM, Ben-Shaul A. Predicting the sizes of large RNA molecules. Proc Natl Acad Sci USA 2008; 105:16153-8; PMID:18845685; http://dx.doi.org/10.1073/pnas.0808089105.
91. Gopal A, Zhou ZH, Knobler CM, Gelbart WM. Visualizing large RNA molecules in solution. RNA 2012; 18:284-99; PMID:22190747; http://dx.doi.org/10.1261/rna.027557.111.
92. Olsthoorn RCL. Structure and Evolution of RNA Phages. Leiden: Leiden University, 1996.
93. Groeneveld H. Secondary Structure of Bacteriophage MS2 RNA: Translational Control by Kinetics of RNA folding. Leiden: Leiden University, 1997.
94. Fang X, Littrell K, Yang XJ, Henderson SJ, Siefert S, Thiyagarajan P, et al. Mg2+-dependent compaction and folding of yeast tRNAPhe and the catalytic domain of the B. subtilis RNase P RNA determined by small-angle X-ray scattering. Biochemistry 2000; 39:11107-13; PMID:10998249; http://dx.doi.org/10.1021/bi000724n.
95. Johnson JM, Willits DA, Young MJ, Zlotnick A. Interaction with capsid protein alters RNA structure and the pathway for in vitro assembly of cowpea chlorotic mottle virus. J Mol Biol 2004; 335:455-64; PMID:14672655; http://dx.doi.org/10.1016/j.jmb.2003.10.059.
96. Abrescia NGA, Bamford DH, Grimes JM, Stuart DI. Structure unifies the viral universe. Annu Rev Biochem 2012; 81:795-822; PMID:22482909; http://dx.doi.org/10.1146/annurev-biochem-060910-095130.

Evidence that viral RNAs have evolved for efficient, two-stage packaging

Alexander Borodavka, Roman Tuma[1], and Peter G. Stockley[1]

Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom

Edited by Michael G. Rossmann, Purdue University, West Lafayette, IN, and approved August 15, 2012 (received for review March 21, 2012)

Genome packaging is an essential step in virus replication and a potential drug target. Single-stranded RNA viruses have been thought to encapsidate their genomes by gradual co-assembly with capsid subunits. In contrast, using a single molecule fluorescence assay to monitor RNA conformation and virus assembly in real time, with two viruses from differing structural families, we have discovered that packaging is a two-stage process. Initially, the genomic RNAs undergo rapid and dramatic (approximately 20–30%) collapse of their solution conformations upon addition of cognate coat proteins. The collapse occurs with a substoichiometric ratio of coat protein subunits and is followed by a gradual increase in particle size, consistent with the recruitment of additional subunits to complete a growing capsid. Equivalently sized nonviral RNAs, including high copy potential in vivo competitor mRNAs, do not collapse. They do support particle assembly, however, but yield many aberrant structures in contrast to viral RNAs that make only capsids of the correct size. The collapse is specific to viral RNA fragments, implying that it depends on a series of specific RNA–protein interactions. For bacteriophage MS2, we have shown that collapse is driven by subsequent protein–protein interactions, consistent with the RNA–protein contacts occurring in defined spatial locations. Conformational collapse appears to be a distinct feature of viral RNA that has evolved to facilitate assembly. Aspects of this process mimic those seen in ribosome assembly.

fluorescence correlation spectroscopy | RNA folding | RNA condensation | kinetics | hydrodynamic radius

Positive-sense, single-stranded (ss)RNA viruses are ubiquitous pathogens in all kingdoms of life (1), causing significant human disease and major financial losses (2). Therapeutic strategies are currently limited and the ideal of vaccination will only ever be practical in a minority of the human and animal viruses. In addition, recent work has highlighted the potential problems that might arise by misincorporation of nonviral RNAs in virus-like particles (VLPs) (3) that are being considered as synthetic vaccines against both pathogens and oncogenic viruses (4). A more thorough understanding of the molecular events central to viral lifecycles is therefore needed. Such studies may reveal novel strategies for therapeutic intervention.

Genomic RNAs play essential roles during the viral lifecycle, adopting different metastable conformations in order to be replicated, translated and packaged into the virion (5, 6). The latter process, the final step in the production of infectious viral progeny, is the topic of the experiments described here. For isometric, nonenveloped virions, protein capsids self-assemble around their genomes, resulting in their confinement at relatively high packing densities (7). It has been proposed that this is a spontaneous process because RNA molecules are branched polymers and there is no barrier to compaction due to large persistence lengths, as seen in dsDNA phages. Electrostatic neutralisation of the nucleic acid charge has been seen as the principal driving force for this confinement step (8–12), consistent with the packaging of noncognate RNAs or protein-alone assembly in vitro (13).

Here we have investigated the molecular mechanism controlling this vital step in the viral lifecycle by taking advantage of single molecule fluorescence correlation spectroscopy (smFCS) to selectively monitor coat protein (CP) or viral RNA components in in vitro reassembly reactions. Such assays can be carried out at low concentrations ($\leq$1 μM), allowing observation of mechanistic features that are not dominated by high CP concentrations. We have applied the smFCS assays to two viral model systems (Fig. 1), the RNA bacteriophage MS2 of the *Leviviridae* family and Satellite Tobacco Necrosis virus (STNV), which is representative of a large number of plant viruses. These were chosen because our previous studies (14–22) have revealed potential roles played by the RNA in coat protein quasi-conformer switching, required for correct assembly of the MS2 $T = 3$ capsid and the presence of multiple putative preferred coat protein binding sites (packaging signals) within the STNV genome, which gets packaged into a $T = 1$ capsid in which all protein conformers are identical. The two viral coat protein architectures are distinct, the STNV CP having a positively charged N-terminal extension to its globular body which is essential for assembly (17) and partially disordered in X-ray structures, whilst the MS2 CP lacks these features. These differences allow us to identify conserved and distinct mechanistic processes.

SmFCS reveals that there is no simple correlation between RNA length and hydrodynamic radius (R_h) and this property cannot be used to discriminate between viral and nonviral RNAs for these systems. The viral RNAs are larger in the absence of their CPs than the capsids into which they must eventually fit. Remarkably instead of a steady condensation of the RNA, which would be expected by a charge neutralisation mechanism, addition of CPs to their cognate RNAs results in a rapid collapse (<1 min) in the solution conformation. Collapse depends on protein–protein interactions, and does not occur on nonviral RNA controls or with the noncognate viral RNA, showing that it depends on specific RNA-CP interactions mediated by the sequence and structure of each genome. The collapsed state is smaller than the capsid and appears to consist of complexes with substoichiometric amounts of coat proteins with respect to capsids, but with roughly the correct shell curvature. The full complement of CPs is recruited in a second slower stage of assembly. Nonviral RNAs support assembly inefficiently and with much lower fidelity.

Results and Discussion

During assembly the sizes and hydrodynamic properties of viral RNAs and CP aggregates change as a result of their co-polymerisation. SmFCS is particularly suitable for determining the hydrodynamic sizes of species in solution (see *SI Methods* for a brief introduction and further references). By fluorescent labeling of the CP component one can selectively follow capsid shell poly-

Author contributions: A.B., R.T., and P.G.S. designed research; A.B. performed research; A.B., and R.T. contributed new reagents/analytic tools; A.B., R.T., and P.G.S. analyzed data; and A.B., R.T., and P.G.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: r.tuma@leeds.ac.uk or p.g.stockley@leeds.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1204357109/-/DCSupplemental.

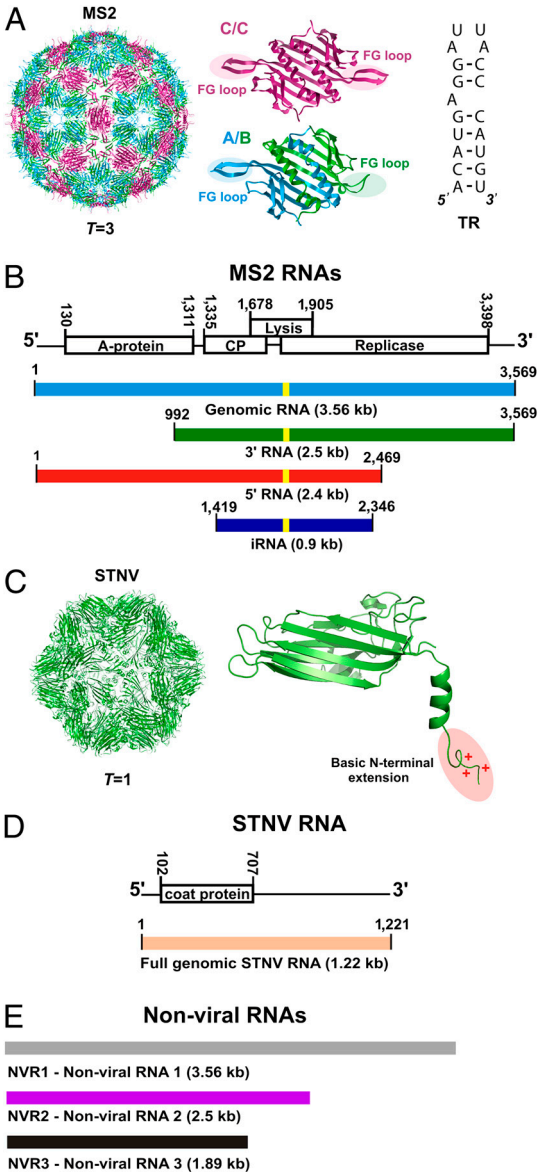


Fig. 1. The bacteriophage MS2 and Satellite Tobacco Necrosis Virus (STNV) systems. (*A*) Structure of the $T = 3$ MS2 bacteriophage (PDB ID 2ms2). Structures of the A/B and C/C quasiequivalent coat protein dimers (PDB ID 1zdh) that differ in the conformations of their FG-loops (highlighted). Sequence and secondary structure of the high-affinity 19 nt TR stem-loop, which is known to cause a conformational switch in coat protein from the C/C to the A/B dimer. Capsid assembly is believed to be initiated at this site on genomic RNA. (*B*) Genetic map of the MS2 genome (GenBank Accession NC001417) and the RNAs used in assembly studies (color-coded here and throughout). The locations of TR are indicated by the yellow stripes. (*C*) Structure of the $T = 1$ STNV capsid. Structure of the STNV coat protein monomer (PDB ID 3RQV) shown with its positively charged N-terminal extension (highlighted). (*D*) Genetic map of the STNV genome (strain C, GenBank Accession AJ000898) and the RNA used in assembly studies. (*E*) Nonviral RNA controls used in the study (color-coded here and throughout).

merisation, whilst labeling of the viral RNA monitors any conformational changes, such as compaction, that occur (23). As described in *Materials and Methods* and Fig. S1, we labeled MS2 coat proteins and transcript RNAs with Alexa-fluor-488. These molecules retained the ability to self-assemble into $T = 3$ capsids, as described in *SI Materials and Methods* and in Fig. S2. In experiments using TR oligonucleotides we were able to show that the smFCS assays reproduced the features of the MS2 assembly reaction seen previously by mass spectrometry (20). The derived hydrodynamic radii of the species involved matched very closely values for the same species obtained by other techniques (Table S1).

Investigating the Solution Conformations of RNAs Via smFCS. Unlike ensemble methods smFCS data can be obtained in very dilute solutions, effectively eliminating artefacts due to aggregation, including nonspecific interactions between large RNA molecules. Fig. 2*A* shows the average R_h values for the dye-labeled genomic and subgenomic RNAs from MS2 and STNV, together with three nonviral RNA controls (Fig. 1 and Figs. S1 and S2) as a function of their nucleotide lengths. The nonviral control RNAs (NVR1-3) were chosen because they had similar lengths to some of the viral RNAs being used. NVR1 is a fragment of the *E.coli* RNA polymerase B subunit (*rpoB*), a high copy number transcript present in the same cells in which MS2 replicates. NVR2 is a transcript from pGEM-3Zf(+) encompassing most of the plasmid. NVR3 is a transcript encompassing a eukaryotic ORF.

There is a general increase of R_h with the length of the RNAs, but they are not directly proportional. For example gRNA and the 5'RNA have similar average R_h values (approximately 12 nm) despite the fact that the latter is approximately 30% shorter. Similarly, the 3'RNA fragment has an even larger R_h (approximately 14 nm), implying that there are long range RNA-RNA contacts that are preserved in both the gRNA and 5'RNAs that stabilize their more compact conformations (24, 25). Contrary to previous predictions (8) the viral RNAs are not significantly more compact than their nonviral counterparts (*c.f.* 3'RNA and NVR2, STNV and NVR3).

SmFCS is sensitive to variations in R_h due to different conformations and these can be seen in time dependent R_h distributions (Fig. S3 and Table S2). The distribution width is a measure of the conformational heterogeneity for each RNA and is shown as error bars in Fig. 2*A*. It is clear that even the relatively compact and folded MS2 genomic RNA is an ensemble of molecules with a wide range of conformations.(R_h approximately 10–16 nm). Even for conformers that are relatively compact the sizes of the majority of viral RNAs are larger than the volumes of their cognate capsids. Hence, the genomic RNAs need to be condensed during packaging.

Assembly Kinetics are Dependent on the Sizes and Sequences of the RNA Being Encapsidated. We then examined the assembly of labelled MS2 CPs in the presence of genomic and various subgenomic RNAs (Fig. 1 and 2*B*), all of which encompass the TR site and presumably initiate assembly from that point. These fragments have been shown previously to support assembly of $T = 3$ capsids (23) and EMs of the smFCS reaction end-points confirmed this (Fig. S2). Assembly time courses were measured by acquiring time-dependent correlation functions, which were then reduced to plots of apparent R_h *vs.* time (Fig. 2*B*). As expected, there is time-dependent variation in the apparent R_h values (shown by the black dotted lines) but the trends of each data set are obvious after data smoothing (solid lines, see *Materials and Methods*). The apparent rates of assembly vary between RNA fragments but are not directly related to RNA fragment length. Instead, the rates increase roughly with the $(R_h)^3$ values of each RNA; i.e., their volumes (Fig. 2*C*), suggesting that the rate is determined by complex factors such as compactness and shape. In contrast, coat protein assembles very slowly if at all with two of the nonviral RNAs, confirming that RNA sequence/structure determines assembly efficiency.

RNA Conformational Changes During Encapsidation. We then examined the process of RNA compaction using labelled RNAs. Fig. 3 shows time-resolved $R_h(t)$ values for RNAs in assembly buffer. Fluctuations in the initial $R_h(t)$ values correspond to the conformational variations discussed above for protein-free RNA.

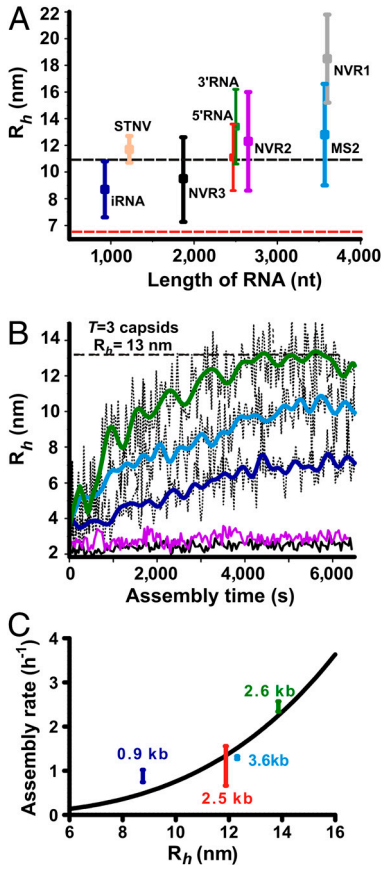


Fig. 2. (*A*) Size distribution of the test RNAs (color-coded as on Fig. 1) measured by FCS as a function of their lengths in nucleotides. Vertical error bars correspond to the full width at half maximum of R_h distributions for each of the RNAs. The black dashed line represents the inner capsid radius of MS2, whilst the red one corresponds to the inner radius of STNV capsid. (*B*) Assembly assays with fluorescently labeled MS2 CP_2. Capsid assembly kinetics with genomic (blue), and subgenomic 3′RNA (green) and iRNA (dark blue). The kinetics with nonviral RNA controls are shown in magenta (NVR2) and black (NVR3). Apparent hydrodynamic radii, R_h, are plotted as a function of time from assembly initiation (assembly time). FFT smoothed data are shown as solid lines overlaid on the original traces (thin dotted lines). The gray dashed line represents the hydrodynamic radius of the MS2 virion. (*C*) Apparent assembly rates obtained from a single exponential approximation of $R_h(t)$ for different MS2 RNAs as a function of their hydrodynamic size. Vertical error bars (rate) represent the standard error of repeated experiments.

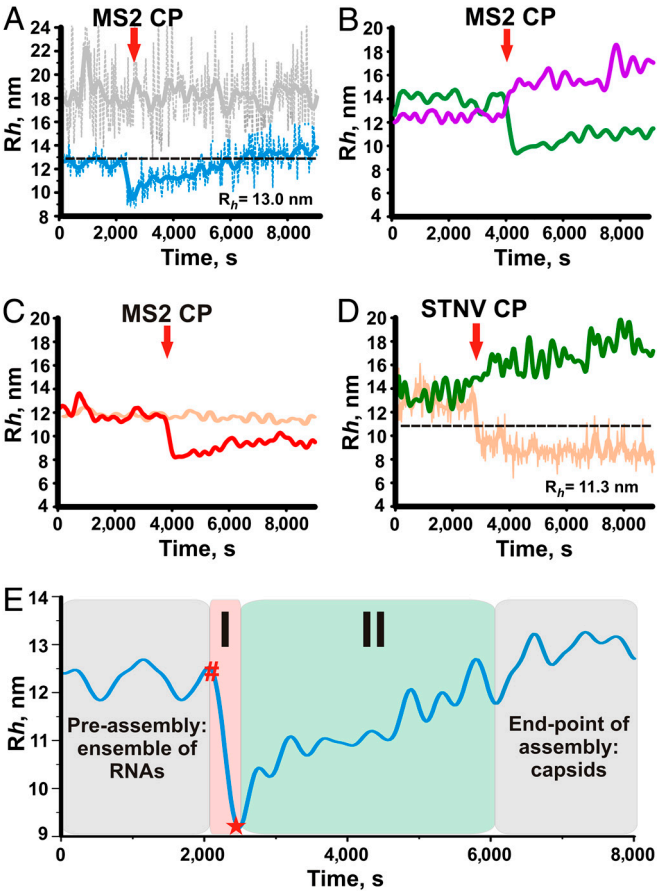


Fig. 3. Assembly with labelled long RNA was followed by $R_h(t)$ and performed at 1:200 molar ratio RNA:CP_2 for MS2 or at 1:60 molar ratio for STNV under the conditions described in *Materials and Methods* . Data collection was started for RNA alone and the coat protein (CP) was mixed with RNA at the time points indicated by the red arrows. (*A*) gRNA (light blue) and NVR1 (gray). The black dashed line represents the R_h of the MS2 virion. (*B*) 3′RNA (green) and NVR2 (magenta). (*C*) 5′RNA (red) and STNV RNA (orange). (*D*) 3′RNA (green) & STNV RNA (orange) upon addition of STNV CP. The black dashed line represents the R_h of the STNV virion. (*E*) Changes in R_h for viral RNAs upon addition of cognate coat protein. Red hash (#) indicates the precollapse state of the RNA-protein complex, which is not detectable by FCS due to the fast kinetics of the CP binding reaction. The initial (ensemble of RNA molecules) and end-points of the assembly reaction (capsids) are highlighted in gray. The collapse stage (I) is highlighted in pink and the collapsed RNA intermediate is marked with a red star (*). The capsid assembly stage (II) is highlighted in green.

At the times indicated by arrows in Fig. 3*A*–*C*, capsid assembly was initiated on each RNA by the addition of a 200-fold molar excess of MS2 CP_2; i.e., enough CP to complete capsid formation around each RNA. Fig. 3*A* compares the data for gRNA (approximately 3.5 kb) and the NVR1 control (approximately 3.5 kb). Both protein-free RNAs have dramatically different R_h values although they are roughly the same length. Although the viral RNA is the smaller, it is larger than the inner capsid radius (approximately 10 nm). Strikingly, these RNAs behave very differently upon addition of coat protein. There is an immediate collapse by approximately 30% of the R_h of the viral RNA, whereas very little happens to NVR1. The smoothing of the $R_h(t)$ traces disguises the kinetics of the collapse with the viral RNA, which happens within the experimental dead time (<1 min) and is presumably a consequence of the rapid formation of RNA-CP interactions.

Following the collapse, gRNA increase its apparent hydrodynamic radius reaching a plateau close to the size expected for $T = 3$ capsids (13 nm, Fig. 3*A*). In contrast, the nonviral control, NVR1, appears largely unaffected by the presence of coat protein. Both RNAs lead to the formation of capsids or capsid-like aggregates, presumably with different yields, as judged by TEM, with the nonviral RNA showing many more aggregates than gRNA. The R_h distributions of the assembly reactions show that gRNA results in a tight distribution similar to bona fide samples of capsid (Fig. S3). In contrast, the NVR1 distribution is slightly larger than that for the free RNA with a peak much larger than the diameter of the $T = 3$ MS2 capsid. The data support the ideas that faithful assembly requires compaction of the protein-free RNA and that this is facilitated by the viral RNA sequence compared to a similar length control.

In order to assess the roles of viral RNA sequences in this collapse and the faithful assembly process we then examined two subgenomic RNA fragments in similar assays. The MS2 3′RNA and the control NVR2 are similar lengths but the MS2 RNA has a larger initial R_h. Strikingly it also shows a CP-induced collapse whilst the control does not collapse (Fig. 3*B*). In this case both RNAs show a gradual $R_h(t)$ increase after addition of the coat protein consistent with further assembly of RNA-protein complexes, which was confirmed by TEMs (Fig. S3). Once again

the viral RNA sequence leads to faithful production of $T = 3$ shells whilst the nonviral RNA control yields large fused-shell structures or aggregates (Fig. S3). This result confirms that it is the viral RNA sequence rather than its size that governs the initial collapse and subsequent assembly. Similar behavior, for example, lack of initial collapse and aberrant assembly, occurs for the other nonviral RNA NVR3 (Fig. S3).

As a final control we compared two viral RNA fragments, the MS2 5'RNA and the STNV genomic RNA, that both must undergo compaction during assembly of their cognate capsids (see Fig. 2). The data (Fig. 3*C*) show that the cognate interaction between MS2 coat protein and its RNA causes the collapse to occur whilst the STNV RNA behaves similarly to the nonviral RNA controls (Fig. S3). This confirms that the initial collapse requires sequence-specific RNA recognition by the coat protein. Above a certain threshold CP_2 concentration (approximately 50 nM) the magnitude of this collapse does not depend on protein concentration (Fig. S4). This result is not compatible with a simple charge neutralization in which increasing protein concentration would produce a gradual compaction of the RNA. Instead it suggests a cooperative binding of a subset of CP_2 to the RNA. Using an MS2 coat protein mutant (W82R) that retains the ability to bind TR sequences but not to assemble beyond that point (26), we were able to show that the collapse depends on protein–protein as well as protein–RNA interactions (Fig. S5).

Thus, it appears that specific recognition of multiple sites on MS2 viral RNAs by a subset of the coat protein subunits required to form the capsid triggers a collapse, which facilitates assembly. In order to assess whether this assembly mechanism is specific to MS2 and related bacteriophages, or also occurs more widely, we carried out similar assays using STNV CP and STNV genomic RNA (Fig. 3*D*). The comparison in this case was between STNV RNA and MS2 3'RNA which is of a similar R_h, although it is much longer. In this case, a rapid collapse occurs when STNV coat is mixed with its cognate RNA whilst the MS2 RNA behaves similarly to the previous nonviral controls. TEMs and R_h distributions (Fig. S3) show that the cognate protein–RNA interaction yields mostly $T = 1$ capsids, whereas reassembly of MS2 RNA by STNV protein yields mostly fused shells and aggregates. These very similar outcomes for two different viruses, especially since they represent very different coat protein folds, MS2 lacking the positively charged N-terminal region of STNV CP, suggests that viral RNA collapse is likely to be a widespread phenomenon and is a prerequisite for correct assembly.

A Two-Stage Assembly Mechanism. A typical path of a viral RNA through packaging is epitomised by MS2 gRNA (Fig. 3*E*). In the preassembly state, the protein-free RNA exists in an ensemble of conformations. These are relatively compact and likely possess folded structures that favor assembly, since heat denaturation followed by rapid quenching of this RNA on ice produces an RNA with decreased assembly rate and yield (Fig. S6). Rapid coat protein binding to RNA initiates the first stage of assembly (region I in Fig. 3*E*), namely a cooperative RNA collapse, which results in a compact initiation complex. This stage is dependent both on specific recognition of viral RNA sequences and protein–protein interactions. In the second stage of assembly (II in Fig. 3*E*), additional CP subunits are recruited to the growing shell in a process that is CP concentration dependent (Fig. S4). The similarity in behavior of the MS2 and STNV genomic RNAs in the presence of their cognate coat proteins suggests that this two-stage mechanism is likely adopted by many ssRNA viruses. Nonspecific RNAs, on the other hand, may be encapsidated using a different mechanism which bypasses stage I via formation of heterologous nucleation events from multiple sites, often leading to fused or partial capsids.

Biological Implications. The data described above, together with our extensive previous studies on the assembly of MS2 (20–23, 27), allow us to propose a molecular model for capsid assembly (Fig. 4) that amounts to a paradigm shift with respect to ssRNA viruses. Current views of assembly are dominated by protein-centric models based on the observations that many viral coat proteins can assemble in the absence of RNA, around nonviral RNA or polyanions (28, 29) and even nanoparticles (30). This has led to the conclusion that viral proteins are the key component in the assembly process (31). In vivo, however, most viruses package their own genomes highly efficiently into defined capsids even in the presence of competitor cellular RNA. There must therefore be some mechanism promoting both specificity and assembly fidelity. We have argued that coat proteins and RNA act as mutual chaperones altering their conformations due to complex formation in a context specific way as the capsid shell grows from an initiation complex (23).

The smFCS data are entirely consistent with this view and reveal several novel features to the assembly mechanism. In the first stage of assembly, these RNAs undergo a rapid collapse in their solution conformations. The threshold concentration at which this occurs for MS2 suggests that this is driven by the formation of multiple CP-RNA contacts, presumably at TR and TR-like degenerate sites (shown in red in Fig. 4). The K_d value of the highest affinity MS2 CP binding site, TR, is low nanomolar and secondary sites would be expected to be lower affinity (32), consistent with the effect being mediated by RNA-protein affinities in this concentration range (i.e., 50 nM, see Fig. S4). This idea is further supported by the lack of collapse for the smallest iRNA (Fig. S6*B* and *C*), which encompasses the high affinity TR site but may lack some secondary binding sites. Indeed, a set of dispersed, TR-like binding sites located throughout the genome has recently been identified for MS2 and a related phage GA [Dykeman, Stockley and Twarock, unpublished]. That bioinformatics analysis also demonstrated that such a disposition of binding sites is

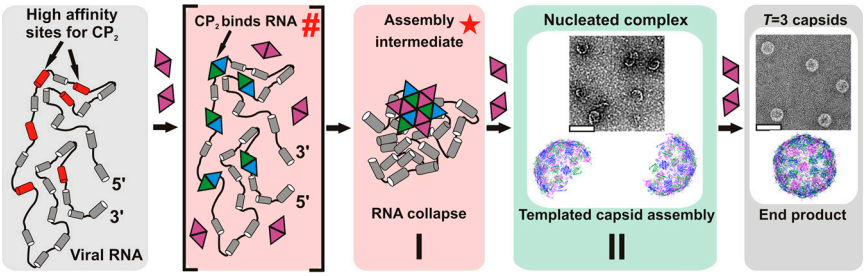


Fig. 4. Proposed two-stage assembly model for MS2. A pre-assembly ensemble of protein-free genomic RNAs (gray box) presents a number of dispersed high affinity sites for CP_2 (red, not to scale), having an R_h approximately 20–30% larger than the inner space of the $T = 3$ capsid. CP_2 binds to these putative packaging signals creating an initiation complex denoted as a (hash) as in Fig. 3*E*. Protein–protein interactions within this complex trigger RNA collapse into the assembly intermediate with an R_h small enough to fit within the confines of the capsid shell, denoted by the (red star). This completes stage I of the assembly process (pink box). During stage II the stable collapsed complex recruits additional CP_2 to populate shell-like species of the correct curvature (green box, see inset for EM) and allows efficient completion of capsids (end-point of assembly, gray box).

unlikely to occur in cellular RNAs. Similarly, degenerate CP binding sites have been identified in STNV (17) and other ssRNA viruses (33). We propose that such coat protein binding sites distinguish viral RNAs from their cellular competitors by facilitating the initial collapse, which also promotes packaging at coat protein concentrations significantly lower than those necessary for formation of empty shells in vitro. Assembly at low protein concentrations would also minimize nonspecific packaging of cellular RNAs in infected cells. Assembly has an intrinsic entropic cost as the ensemble of solution RNA conformations and multiple CP subunits become ordered in the virion. Clearly for the two test viral RNAs used here, most of this cost is compensated by the enthalpic gains from the initial CP-RNA contacts and further augmented by protein–protein interactions in the initiation complex. Hence, most of the energy cost may be recouped during the formation of the specific initiation complex and consequently stage II may be more favorable for packaging of specific rather than for nonspecific RNAs. This is clearly not the case for noncognate RNAs where coat proteins would bind at more or less random places along each RNA, resulting in heterologous nucleation and complexes of low stability. The constant break-down and reformation of weak nucleation complexes would slow the rate of capsid formation (34) for such noncognate RNAs, which is indeed the case (Fig. 2*B*). This mechanism also explains why viral RNA gets efficiently packaged even in the presence of a large number of cellular competitors. The viral RNA sequence and structure, fine tuned by evolution, decreases the entropic penalty for the formation of the collapsed/compact initiation complex, which is stable and competes efficiently for additional coat proteins from the less stable heterologously initiated cellular competitors.

The collapsed RNAs have sizes ($R_h \sim 9$–10 or 8 nm, for MS2 and STNV, respectively) similar to that of their cognate capsid interiors [$R_{\text{inner}} \sim 11$ and 6 nm (15, 18, 35)] and may act as a scaffold during stage II, facilitating formation of complete capsids with the correct curvature. Negative stain EMs of the early products of assembly with MS2, equivalent to the start of stage II, show incomplete shells of approximately the correct curvature, consistent with growth from an initiation complex (Fig. 4). At later times all the viral RNAs result in high yields of $T = 3$ capsids with few misassembled species. Because this is also true of subgenomic RNA fragments, it implies that for the collapsed complex CP-CP interactions are sufficient to complete the capsid shell.

This behavior is remarkably similar to results emerging from studies of RNA folding and particularly ribosome assembly (36). Folding of the *Tetrahymena* ribozyme shows that it can undergo very rapid collapse into its native conformation (in approximately 1 s). However, the bulk of the RNA collapses into nonnative folds that then rearrange over a much longer time period (min). In contrast, ribosomal 16S rRNA folding is both rapid and mediated by interactions with the 30S proteins. The result is a highly ordered and accurate assembly reaction brought about by induced conformational changes within the ribonucleoprotein complex, although in that case there are few direct protein–protein contacts. It appears that with significantly larger RNAs to fold into the restricted volumes of their capsids, RNA viruses exploit similar basic mechanisms but with the additional co-operativity of the CP lattice that forms around the RNA. The smFCS assays described here will be an important tool in the armoury of techniques needed to investigate the behaviour of both viral and other large RNAs (37).

Materials and Methods

MS2 virus-like particles (VLPs) were produced and purified as described previously (38). Purified VLPs were covalently labeled with Alexa Fluor 488-SDP ester (Invitrogen) via surface lysines (39). RNA-free CP, having approximately 0.3 mole dye/mole CP_2, was isolated as previously described (40) and used in assembly assays.

RNA 19-mer, TR, encompassing the high affinity packaging signal (Fig. 1*C* and ref. 20, 27) and carrying a 3′-amino group, was produced by solid-phase synthesis and labeled using Alexa Fluor 488-SDP ester. The genomic and subgenomic RNAs (Fig. 1*D*) were produced by transcription of cDNA clones (23). Genomic RNA extracted from the phage using RNeasy mini kit (QIAGEN) and produced similar results in assembly assays to the corresponding transcript; hence transcripts were used throughout unless otherwise indicated. Nonviral control RNAs were transcribed from linearized templates as described in Supporting Information. All RNAs were purified using RNeasy mini kit (QIAGEN). 3′-end fluorescent labelling of RNA transcripts was achieved via ligation of a 5′-phosphorylated-oligo-A(25-mer)-3′-AF488 oligonucleotide (subgenomic and nonviral RNAs) or incorporation of a reactive amine group at the 3′-end using poly-A polymerase, followed by chemical coupling of the amine to AF488 SDP-ester (MS2 full genomic RNA, STNV). STNV genome was also trace labelled by transcription in the presence of amino-allyl UTP followed by chemical coupling of the amine to AF488 SDP-ester. Denaturing formaldehyde agarose gels were used to assess RNA integrity before and after experiments.

Assembly conditions (23) were optimised for the single molecule assays. RNA (0.5 nM) was mixed with AF488-labeled CP_2 (final concentrations 25 nM–100 nM) to achieve the desired CP_2 : RNA ratio. Similarly, labeled RNAs (long RNAs at 1 nM, TR at 10 nM) were mixed with unlabeled CP_2 (50 nM-5 μM). All assembly reactions were performed in assembly buffer [50 mM ammonium acetate, 1 mM EDTA, 1 mM DTT, 0.05% (v/v) Tween20, pH 7.5] at 21 °C. Manual mixing caused approximately 1 min delay at the start of time-resolved FCS data collection. FCS measurements were made using a custom-built FCS setup (41) with 10 s data accumulation per each autocorrelation function (CF). Individual CFs were decomposed into triplet state relaxation and quenching (if necessary) and diffusion (characterized by diffusion time, t_D) components and the latter was converted into an apparent hydrodynamic radius R_h (42). Electron microscopy grids were prepared at the end of FCS data collection to monitor formation of assembly products by imaging in negative stain. Further details are given in *SI Methods*.

ACKNOWLEDGMENTS. We thank Professor Dave Rowlands and Professor Reidun Twarock and her group (University of York) for many helpful comments on these results, Dr. Robert Coutts for providing the STNV-C genome and Amy Barker for synthesis of RNA oligonucleotides. We thank Dr. Tomas Fessl (University of Dundee) for help with processing of FCS data. AB thanks the Wellcome Trust for support of his postgraduate studentship [089310/Z/09/Z] and support of the facilities used [062164][090932/Z/09/Z]. Single molecule work within the Astbury Centre has been supported by The University of Leeds.

1. Schneemann A (2006) The structural and functional role of RNA in icosahedral virus assembly. *Annu Rev Microbiol* 60:51–67.
2. Lomonossoff GP (1995) Pathogen-derived resistance to plant-viruses. *Annu Rev Phytopathol* 33:323–343.
3. Routh A, Domitrovic T, Johnson JE (2012) Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc Natl Acad Sci USA* 109:1907–12.
4. Tumban E, Peabody J, Peabody DS, Chackerian B (2011) A pan-HPV vaccine based on bacteriophage PP7 VLPs displaying broadly cross-neutralizing epitopes from the HPV minor capsid protein, L2. *Plos One* 6:e23310.
5. Simon AE, Gehrke L (2009) RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. *Biochim Biophys Acta* 1789:571–583.
6. Venter PA, Krishna NK, Schneemann A (2005) Capsid protein synthesis from replicating RNA directs specific packaging of the genome of a multipartite, positive-strand RNA virus. *J Virol* 79:6239–6248.
7. Speir JA, Johnson JE (2012) Nucleic acid packaging in viruses. *Curr Opin Struc Biol* 22:65–71.
8. Yoffe AM, et al. (2008) Predicting the sizes of large RNA molecules. *Proc Natl Acad Sci USA* 105:16153–16158.
9. Bruinsma RF (2006) Physics of RNA and viral assembly. *Eur Phys J E* 19:303–310.
10. Rudnick J, Bruinsma R (2005) Icosahedral packing of RNA viral genomes. *Phys Rev Lett* 94:038101–038104.
11. van der Schoot P, Bruinsma R (2005) Electrostatics and the assembly of an RNA virus. *Phys Rev E* 71:061928–061939.
12. Belyi VA, Muthukumar M (2006) Electrostatic origin of the genome packing in viruses. *Proc Natl Acad Sci USA* 103:17174–17178.
13. Balint R, Cohen SS (1985) The incorporation of radiolabeled polyamines and methionine into turnip yellow mosaic-virus in protoplasts from infected plants. *Virology* 144:181–193.
14. Toropova K, Stockley PG, Ranson NA (2011) Visualising a viral RNA genome poised for release from its receptor complex. *J Mol Biol* 408:408–419.
15. Lane SW, et al. (2011) Construction and crystal structure of recombinant STNV capsids. *J Mol Biol* 413:41–50.

16. Dykeman EC, et al. (2011) Simple rules for efficient assembly predict the layout of a packaged viral RNA. *J Mol Biol* 408:399–407.
17. Bunka DHJ, et al. (2011) Degenerate RNA packaging signals in the genome of satellite tobacco necrosis virus: Implications for the assembly of a $T = 1$ capsid. *J Mol Biol* 413:51–65.
18. Toropova K, et al. (2008) The three-dimensional structure of genomic RNA in bacteriophage MS2: Implications for assembly. *J Mol Biol* 375:824–836.
19. Rolfsson O, et al. (2008) RNA packing specificity and folding during assembly of the bacteriophage MS2. *Comput Math Methods Med* 9:339–349.
20. Stockley PG, et al. (2007) A simple, RNA-mediated allosteric switch controls the pathway to formation of a $T = 3$ viral capsid. *J Mol Biol* 369:541–552.
21. Basnak G, et al. (2010) Viral genomic single-stranded RNA directs the pathway toward a $T = 3$ capsid. *J Mol Biol* 395:924–936.
22. Dykeman EC, Stockley PG, Twarock R (2010) Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J Mol Biol* 395:916–923.
23. Rolfsson O, Toropova K, Ranson NA, Stockley PG (2010) Mutually-induced conformational switching of RNA and coat protein underpins efficient assembly of a viral capsid. *J Mol Biol* 401:309–322.
24. Lodish HF, Zinder ND (1966) Mutants of the bacteriophage f2. 8. Control mechanisms for phage-specific syntheses. *J Mol Biol* 19:333–348.
25. Klovins J, Berzins V, Van Duin J (1998) A long-range interaction in Q(beta) RNA that bridges the thousand nucleotides between the M-site and the 3′ end is required for replication. *RNA* 4:948–957.
26. Peabody DS, Ely KR (1992) Control of translational repression by protein–protein interactions. *Nucleic Acids Res* 20:1649–1655.
27. Morton VL, et al. (2010) The impact of viral RNA on assembly pathway selection. *J Mol Biol* 401:298–308.
28. Cadena-Nava RD, et al. (2011) Exploiting fluorescent polymers to probe the self-assembly of virus-like particles. *J Phys Chem B* 115:2386–2391.
29. Hu YF, et al. (2008) Packaging of a polymer by a viral capsid: The interplay between polymer length and capsid size. *Biophys J* 94:1428–1436.
30. Chen C, et al. (2006) Nanoparticle-templated assembly of viral protein cages. *Nano Lett* 6:611–615.
31. Zlotnick A, Mukhopadhyay S (2011) Virus assembly, allostery and antivirals. *Trends Microbiol* 19:14–23.
32. Lago H, et al. (1998) Dissecting the key recognition features of the MS2 bacteriophage translational repression complex. *Nucleic Acids Res* 26:1337–1344.
33. Kim DY, et al. (2011) Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. *J Virol* 85:8022–8036.
34. Prevelige PE, Thomas D, King J (1993) Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys J* 64:824–835.
35. Chauvin C, Jacrot B, Witz J (1977) Structure and molecular-weight of satellite tobacco necrosis virus—neutron small-angle scattering study. *Virology* 83:479–481.
36. Woodson SA (2011) RNA folding pathways and the self-assembly of ribosomes. *Acc Chem Res* 44:1312–1319.
37. Gopal A, Zhou ZH, Knobler CM, Gelbart WM (2012) Visualizing large RNA molecules in solution. *RNA* 18:284–299.
38. Stonehouse NJ, Stockley PG (1993) Effects of amino-acid substitution on the thermal-stability of Ms2 capsids lacking genomic RNA. *Febs Lett* 334:355–389.
39. Hermanson GT, Hermanson GT (1996) *Bioconjugate techniques* (Academic Press, San Diego) p 785.
40. Sugiyama T, Nakada D (1967) Control of translation of Ms2 Rna cistrons by Ms2 coat protein. *Proc Natl Acad Sci USA* 57:1744–1750.
41. Gell C, et al. (2008) Single-molecule fluorescence resonance energy transfer assays reveal heterogeneous folding ensembles in a simple RNA stem-loop. *J Mol Biol* 384:264–278.
42. Gell C, Brockwell D, Smith A (2006) *Handbook of Single Molecule Fluorescence Spectroscopy* (Oxford University Press, Oxford).

Revealing the density of encoded functions in a viral RNA

Nikesh Patel[a], Eric C. Dykeman[b,c], Robert H. A. Coutts[d,1], George P. Lomonossoff[e], David J. Rowlands[a], Simon E. V. Phillips[f], Neil Ranson[a], Reidun Twarock[b,c], Roman Tuma[a], and Peter G. Stockley[a,2]

[a]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom; [b]York Centre for Complex Systems Analysis, and [c]Departments of Mathematics and Biology, University of York, York YO10 5DD, United Kingdom; [d]Division of Biology, Imperial College London, London SW7 2AZ, United Kingdom; [e]Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norfolk NR4 7UH, United Kingdom; and [f]Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Oxford, Didcot, Oxon OX11 0FA, United Kingdom

Edited by John E. Johnson, The Scripps Research Institute, La Jolla, CA, and accepted by the Editorial Board January 5, 2015 (received for review October 30, 2014)

We present direct experimental evidence that assembly of a single-stranded RNA virus occurs via a packaging signal-mediated mechanism. We show that the sequences of coat protein recognition motifs within multiple, dispersed, putative RNA packaging signals, as well as their relative spacing within a genomic fragment, act collectively to influence the fidelity and yield of capsid self-assembly in vitro. These experiments confirm that the selective advantages for viral yield and encapsidation specificity, predicted from previous modeling of packaging signal-mediated assembly, are found in Nature. Regions of the genome that act as packaging signals also function in translational and transcriptional enhancement, as well as directly coding for the coat protein, highlighting the density of encoded functions within the viral RNA. Assembly and gene expression are therefore direct molecular competitors for different functional folds of the same RNA sequence. The strongest packaging signal in the test fragment, encodes a region of the coat protein that undergoes a conformational change upon contact with packaging signals. A similar phenomenon occurs in other RNA viruses for which packaging signals are known. These contacts hint at an even deeper density of encoded functions in viral RNA, which if confirmed, would have profound consequences for the evolution of this class of pathogens.

virus assembly | single-molecule fluorescence correlation spectroscopy | satellite tobacco necrosis virus | packaging signal

Almost all natural nucleic acid sequences appear in the form of RNA polynucleotides at some point in their existence. The concept of "one RNA, one function" (e.g., for tRNA or mRNA) is increasingly seen as far too simplistic with RNAs known to alter their secondary/tertiary structures in regulated ways to control the appearance of distinct functional states (1). This multifunctionality is also true of the genomes of single-stranded (ss) RNA viruses, where refolding to promote replication, translation and control of gene expression are well known (2). Here we describe additional layers of previously unsuspected functions within the genome of the satellite plant virus, Satellite Tobacco Necrosis Virus (STNV). Aspects of these functions are potential drug targets for this class of viruses, which include major human pathogens and a significant fraction of emerging viral diseases (3, 4). Virus-like particles (VLPs) are also increasingly being used both for vaccines and for delivery of imaging, diagnostic, or therapeutic agents (5, 6). Thus, understanding the detailed molecular mechanisms used by these viruses for assembly is of widespread interest.

Previously we showed, using single molecule fluorescence correlation spectroscopy (smFCS), that at low (nanomolar) concentrations the coat proteins (CPs) of STNV and bacteriophage MS2 package their cognate genomes preferentially (7). Cognate packaging in vitro is accompanied by a characteristic drop in the hydrodynamic radius (Rh) of the viral RNA, making it smaller than its capsid as a subset of CPs bind. We have shown that this behavior is consistent with an assembly mechanism (8) controlled by the interaction of multiple CPs with degenerate RNA sequence/structure sites we term packaging signals (PSs), distributed across the genomes. PSs are bound sequence-specifically by their cognate CPs, increasing the yield, rate and/or fidelity of assembly. PS-mediated assembly has many potential advantages for the virus, including solving a viral version of the Levinthal Paradox (9). For the $T = 3$ MS2, there is a known assembly origin, the 19-nt-long coat protein gene translational operator (TR) stem-loop (10). TR functions as an allosteric effector, switching the dimeric viral capsomere between the two quasi-conformers required to build the capsid (11). This mechanism requires 60 TR-like sites within the genomic RNA for CP contacts (12). These defined RNA–CP contacts require the RNA in every particle be in a very similar conformation, and an asymmetric tomographic reconstruction is consistent with this view (13).

For STNV, there were no known PSs, so we used RNA SELEX combined with a theoretical analysis protocol we designed for the identification of PSs from SELEX data, to isolate preferred sequences for CP binding. This dual approach identified an aptamer whose sequence matches a region of the cognate genome (14). Both the aptamer (B3) and genomic sites (B3-like, PS3, Fig. 1 *A* and *B*) are predicted to fold into stem-loops. An X-ray structure of VLPs reassembled around B3 shows that the presence of multiple copies of a preferred RNA sequence increases the ordering of

Significance

Single-stranded RNA viruses self-assemble protective protein containers around their cognate genomes rapidly and efficiently at low concentrations. RNA encapsidation in vivo occurs preferentially with the cognate genome, in contrast to many in vitro reassembly experiments. We describe in molecular detail how this specificity and efficiency is accomplished using multiple contacts between coat proteins and dispersed packaging signals in the viral genome. The sequences and relative positioning of the packaging signals are important for this mechanism, creating a strong evolutionary constraint. Packaging signals overlap untranslated and coding regions ensuring assembly is in competition with other functions of the genome. Disrupting these contacts has deleterious consequences for capsid assembly identifying a novel antiviral drug target.

Author contributions: N.P., R. Tuma, and P.G.S. designed research; N.P. performed research; E.C.D., N.R., R. Twarock, and R. Tuma contributed new reagents/analytic tools; N.P., N.R., R. Twarock, R. Tuma, and P.G.S. analyzed data; and N.P., E.C.D., R.H.A.C., G.P.L., D.J.R., S.E.V.P., N.R., R. Twarock, R. Tuma, and P.G.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.E.J. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

[1]Present address: Geography, Environment and Agriculture Division, Department of Human and Environmental Sciences, School of Life and Medical Sciences, University of Hertfordshire, Hatfield AL109AB, United Kingdom.

[2]To whom correspondence should be addressed. Email: p.g.stockley@leeds.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1420812112/-/DCSupplemental.

the N-terminal region of the STNV CP. The N-terminal helix extends by roughly one turn to include amino acid residue 8, in contrast to the virion or mRNA containing VLP, in which the electron density shows that the polypeptide is disordered before residue 12. These data suggest that B3 acts like a PS (15) (Fig. 1).

SmFCS confirms that B3 binds STNV CP with high affinity (low nM) to form a capsomere, probably a CP trimer with at least one B3 bound, that can then be assembled into $T = 1$ capsids in response to increasing concentrations of CP. This reaction is highly sequence-specific, because oligos containing B3 variant in the recognition motif fail to assemble under identical conditions. Despite the catalytic role of B3, assembly is highly dependent on electrostatic interactions: The VLPs formed are sensitive to both the ionic strength and the removal of structural calcium ions. Using a 5′ genomic fragment (127 nt long) encompassing the B3-like PS3 and two additional putative PSs on either side, we have confirmed that CP-induced reduction in Rh occurs as a consequence of binding PS3. The five PSs work collectively to promote successful capsid assembly and their spacing within the fragment is also important. These functions are encoded within a region heavily implicated in enhancing translation and viral RNA replication (16–19). In addition, PS3 is located within the CP region that encodes the amino acid sequence undergoing conformational change in response to PS binding. If similar contacts occur more widely for RNA viruses there may be even deeper embedding of function in viral RNAs.

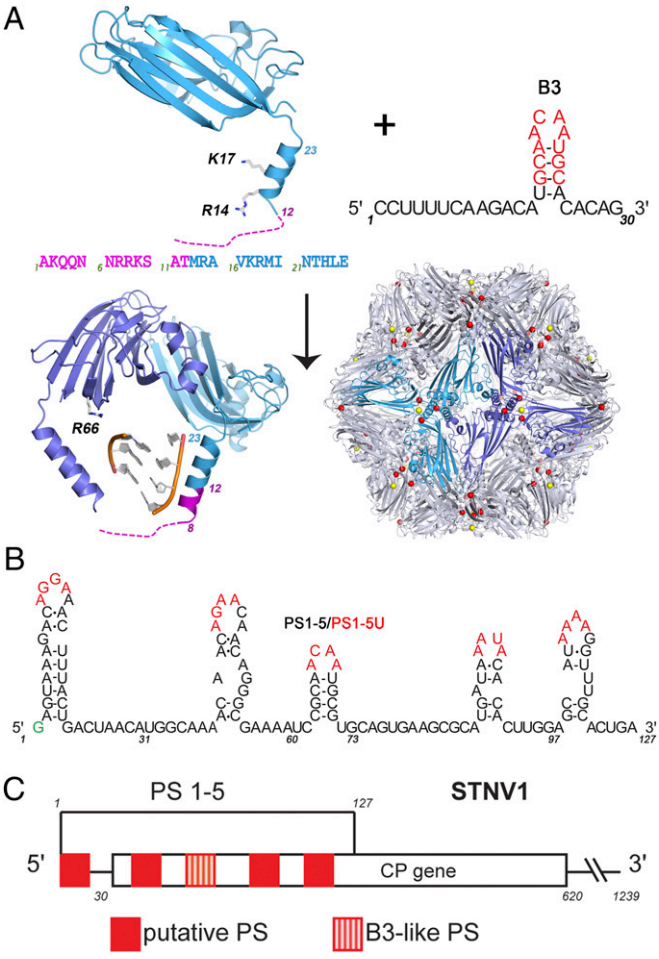


Fig. 1. The components of the STNV system. (*A*) (*Upper Left*) Ribbon diagram of the STNV CP subunit structure seen in virions [Protein Data Bank (PDB) ID 4V4M] with the disordered N-terminal amino acid sequence shown dashed, with the sequence of the first 25 amino acids below. When this assembles around the B3 aptamer, sequence and secondary structure shown in *Upper Right* (nucleotide identities with the STNV-1 genome in red), it forms a $T = 1$ capsid (*Lower Right*, PDB 3S4G), in which there is additional ordering of the CP N-terminal region (magenta helical turn) in response to binding the preferred RNA sequence. The virion is stabilized by multiple Ca^{2+} ions that bind around the differing symmetry axes (shown as color-coded dots). (*B*) The sequence and putative secondary structure of the 5′ genomic fragment encompassing five putative PSs, named in order 5′ to 3′ as PS1-5, respectively. Red nucleotides indicate the potential CP recognition motif (AXXA) and are the sites mutated to U to create variant fragments. The green G is a nonviral nucleotide added for ease of dye-labeling. (*C*) Map of the STNV1 genome showing the positions of these putative packaging signals (red) with respect to the CP gene.

Sequence-Specificity of Assembly

To study in vitro reassembly of STNV VLPs at low CP concentrations, we used fluorescently labeled oligonucleotides (see Fig. 1 *A* and *B* and Table S1 for details). This method enabled smFCS assembly assays at nanomolar RNA concentrations. Initially we monitored CP titrations into labeled B3 (1 nM) under reassembly conditions. This experiment reveals formation of a high affinity RNA–CP complex that begins to form at 1 nM CP and is fully formed by 5 nM CP. It has an apparent Rh of ∼5 nm (Fig. 2*B*), consistent with it being a CP trimeric capsomere based on the X-ray structure of the B3 VLP. Photon counting statistics (*SI Materials and Methods*) were used to estimate the affinity of the CP–B3 interaction, yielding a K_d of 0.8 ± 0.1 nM, and photon burst statistics suggest that the complex formed most likely contains a single B3 RNA (Fig. S1). Once formed, this complex is stable for extended periods (Fig. 2*C*) and indifferent to the addition of extra CP monomer until a threshold concentration ∼50 nM is reached. Thereafter assembly of $T = 1$ capsids occurs slowly in a CP concentration-dependent process that remains incomplete below 1 μM CP. These VLPs are also stable over extended periods, indifferent to the addition of more CP and protect the labeled RNA from RNase digestion (Fig. S2). This second phase of assembly must be driven by CP–CP interactions between the complex nucleated by the initial RNA–CP contact and incoming CPs. Note, at much higher concentrations (μM) the STNV CP is a monomer with no tendency to self-assemble in the absence of RNA (15). The slow kinetics of capsid assembly from the B3 capsomere suggests that the CPs must overcome a barrier to complex formation which we previously speculated was electrostatic, i.e., that the RNA reduces unfavorable interactions between the positively charged N-terminal regions of CP subunits (Fig. 1*A*).

To see whether the RNA effect is purely electrostatic or sequence-specific we set up an assembly titration with B34U, a sequence variant of B3 of the same length predicted to form a B3-like stem-loop, but with the CP recognition motif, -ACAA-, in the loop replaced by -UUUU-. RNA SELEX suggests that the minimal motif consists of a base-paired stem topped by a loop containing -AXXA-, where X is any nucleotide (14). The result (Fig. 2*A*, *Lower*) shows very clearly the importance of this sequence for CP recognition. At ∼twofold higher CP concentration than for B3, the variant shows an increased Rh suggesting formation of the capsomere. However, no further assembly occurs until the CP concentration reaches 3 μM, when there is a sharp increase in Rh followed by a steady decline, consistent with formation of unstable nonspecific aggregates rather than VLPs, and the RNA remains accessible to RNase (Fig. S2). The initial B34U-CP complex is not a viable on-pathway assembly intermediate despite the RNA being of the same length and therefore having the same negative charge. This result shows that capsomere assembly is sequence-specific and that only correctly formed species are intermediates in VLP formation. Because B34U RNA forms a complex with CP, sequence recognition is likely to involve binding of the base-paired stem, which is then orientated correctly for further assembly by sequence-specific contacts with the loop motif. B34U is able to make the first of these contacts, but not the latter. Transmission electron microscopy (TEM) images (Fig. S3*A*) and Rh distribution plots (Fig. 2*B*) support these interpretations. In conclusion, aptamer B3 has the properties expected of a PS.

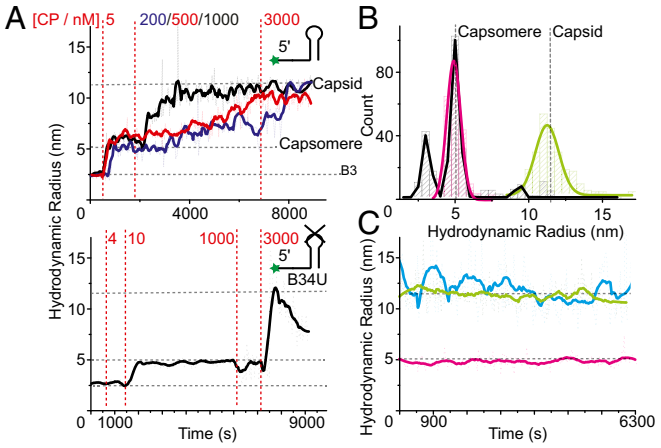


Fig. 2. RNA dependence and sequence-specificity of STNV assembly. (*A*) STNV CP was titrated into 1 nM of 5′ Alexa Fluor-488–labeled B3 (*Upper*) or B34U RNA (*Lower*) (see inset cartoons) in assembly buffer and the resulting hydrodynamic radius monitored using smFCS over time. Dotted red lines indicate titration points with the labels showing the final monomer CP concentrations reached. FFT smoothed data (thick lines) are shown for clarity, with horizontal gray dashed lines showing the sizes of capsid, B3, and capsomere. For B3 the curves for the differing CP concentrations post-formation of the capsomere are color coded and displaced slightly along the time axis. Samples at various points in the titrations were analyzed by TEM (Fig. S3). (*B*) Size distributions of the species in solution for capsomere (magenta) and B3 containing VLP (green) together with the STNV CP + B34U (black) containing species in plot *A*. Gaussian peak fits are shown by thick lines, with vertical gray dashed lines showing expected sizes. (*C*) 1 nM B3 containing capsomeres (magenta) and VLPs (green) were assembled using 4 and 1,000 nM STNV CP, respectively, and their Rh measured over 2 h. The results are compared with similar measurements of 300 nM mRNA VLP (cyan) labeled on the CP with Alexa Fluor-488.

Natural PSs Act Cooperatively

To determine whether assembly with the genomic STNV PSs occurs similarly, and explore potential cooperative effects of having multiple PSs on the same RNA fragment, we carried out CP titrations using a dye-labeled 127 nt RNA. This fragment starts at the 5′ end of the STNV-1 genome (Fig. 1 *B* and *C*) and encompasses five putative PSs, namely sequences that can be folded into stable stem-loops displaying -AXXA- motifs (14). Aptamer B3 has the highest sequence identity with nucleotides 59–72, which is the middle PS on this fragment. We term this fragment PS1-5. We compared reassembly with identical RNA concentrations (10 nM) of PS1-5 and B3 alone (Fig. 3*A*). B3 forms the capsomere at low concentration, ~5 nM and begins to assemble VLPs at 200 nM, although the kinetics of capsid formation are relatively slow and remain incomplete (>4,000 s). On titration to 3,000 nM CP, assembly goes slowly to completion. In contrast, PS1-5 appears to start binding CP only at 120 nM, resulting in a drop of its Rh by 20–30%, reminiscent of the behavior of the full-length genome (7). At 200 nM it rapidly assembles (~250 s) into intact capsids that are nuclease resistant (Fig. S2), suggesting that the multiple PSs promote cooperative formation of the CP–CP contacts found in the capsid.

This difference in behavior between B3 and PS1-5 could be due to the differences in the sequences/structures of their PSs, or simply because of the length of the PS1-5 fragment. To explore this behavior by the longer fragment, we examined the properties of variant fragments in which all of the PSs were mutated by conversion of 4 nt in their loop regions to uridines, or mutating just PS3, or ablating all of the PSs except PS3 (Fig. 3*B*). The all mutant PS1-5 binds CP at 120 nM like the wild-type fragment, but this leads to an increase in its Rh, not a decrease. It does not form $T = 1$ capsids or become nuclease resistant, even when the CP concentration is raised to 3 μM (Fig. 3 *B* and *D*). TEMs suggest that the complexes produced are aggregates rather than partially formed capsids (Fig. S4*A*). This outcome confirms that the loop sequence of PS3 at least is important for CP recognition. When only the putative flanking PSs (1 and 2, and 4 and 5) are mutated, binding again occurs at 120 nM CP and the Rh decreases, showing that this effect is due to initial binding at PS3. However, subsequent assembly is very slow and formation of complete capsids is only achieved at 3,000 nM CP. The result confirms that PS sites other than PS3 play important roles in making capsid assembly efficient, and supports the idea that multiple PSs act cooperatively. When only PS3 is mutated in the PS1-5 fragment, there is no decrease in Rh, confirming that CP binding to PS3 triggers this event, and assembly does not occur until 200 nM CP and does not result in formation of capsids (Fig. 3 *B* and *D*).

A hierarchy of PS CP affinities is an essential requirement of the PS-mediated assembly model (9) and is consistent with the sequence/structure variations in PSs PS1-5 on the genomic fragment. PS3 is the highest affinity site and the other PSs contribute to assembly efficiency but cannot trigger VLP formation alone. The putative STNV PS sites were identified by looking at the short range folding potential of genomic RNA, specifying a stem-loop containing the -AXXA- motif (14), and all such sites with a negative free energy of formation were identified as putative PSs. This simple approach appears successful. Of course, PSs with small folding free energies may only form active conformations in response to CP binding.

Another predicted property of PSs is that their relative locations within the genome is such that they facilitate the CP–CP interactions seen in the capsid, implying that their relative separation in the RNA sequence, which determines their spatial locations in three-dimensions, should be important. To test this, we produced three additional variants of PS1-5 in which 10-nt spacers, chosen in each case to avoid altering the potential secondary structures, were added either 5′ to PS3, i.e., between PS2 and PS3, or 3′ to it, between PS3 and PS4, or at both of these sites. These RNAs are designated 5′ or 3′ spacer PS1-5 and 5′ and 3′ spacer PS1-5, respectively. The results of their titrations with CP are shown in Fig. 3*E*. All three support fairly rapid assembly beyond 120 nM CP, but without an initial Rh decrease. The resulting smFCS traces are very noisy and their distributions suggest that these species form mixtures with few correctly assembled capsids. TEMs (Fig. S4*B*) suggest that although some $T = 1$ capsids are formed by these RNAs, there are many partially formed and malformed species present as well. RNase treatment leads to large changes in these Rh distributions (Fig. S2), confirming that the natural spacing of the five PS sites in the 5′ genomic fragment is important for accurate and efficient capsid assembly. The lack of Rh decreases in fragments with altered PS spacing but containing wild-type PS3 sequences suggests a more complex explanation of the phenomenon in the PS1-5 fragment. CP binding to B34U shows that capsomere-like species can form with high affinity for stem-loops lacking the loop recognition motif. The PS1-5 variants with a wild-type PS3 could allow CPs interacting at that site to contact the neighboring RNA stem-loops, each equivalent to B34U. Although these complexes are not efficient at ensuring cooperative capsid assembly, they could explain the initial decrease in Rh. The spacer variants suggest that both neighboring stem-loops must be contacted to cause this effect, emphasizing the importance of PS positioning in the genome.

Roles of Electrostatics in Assembly and Stability

The sequence-specificity of assembly initiation revealed by the experiments with B3, PS1-5, and their variants, is consistent with PS-mediated assembly, but not with expectations that electrostatic interactions are the principal driving forces behind assembly in this class of viruses (20–25). The PS fragments and the variant sites, with the exception of the insert variants, are of the same length, have the same net charge and are predicted to share common secondary structures. Under ensemble assembly conditions (CP ≥ 10 μM) B34U triggers $T = 1$ capsid formation,

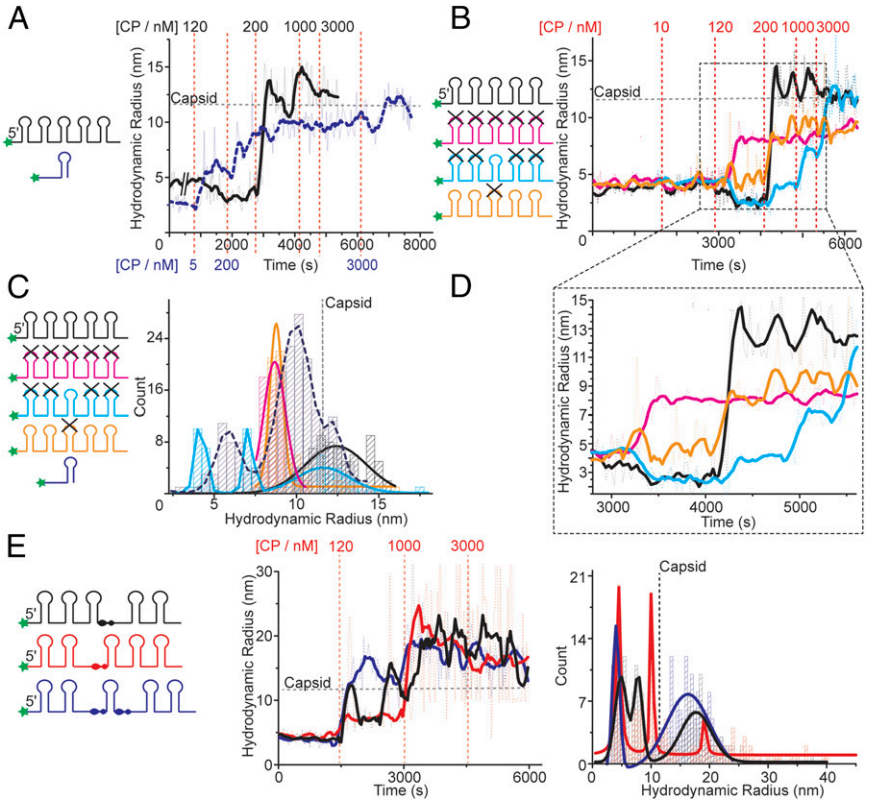


Fig. 3. Packaging signals function cooperatively. (*A*) CP was titrated into 10 nM Alexa Fluor-488–labeled PS1-5 (black) or B3 (dashed - blue). Details as in legend Fig. 2*A*. Titration points are color coded for each RNA. (*B*) CP titration into 10 nM of Alexa Fluor-488–labeled PS1-5 and its sequence variants (X), shown as cartoons (*Left*). (*C*) Color-coded size distributions of species present in the titrations shown in *A* and *B* (substrate key, *Left*). Data shown are taken from the titration of 200 nM CP onwards in *A* and *B*. For PS1-5 and its variants this is 4,000–6,000 s, and for B3 from 2,000 to 8,000 s. (*D*) Expanded section of *B* showing the effects of CP addition on each variant. (*E*) CP titration into spacing variants (see cartoon; *Left*) and their size distribution plots (*Right*). Data shown are from 0 to 5,000 s.

although much less efficiently than B3 (15), i.e., the sequence-specific RNA discrimination is dramatically reduced at higher concentration, and reassembly can be driven almost exclusively by electrostatic interactions (26). In contrast, the outcomes of single-molecule in vitro reassembly, carried out at nanomolar concentrations, are more sensitive to sequence-specific effects and may reflect more accurately conditions in vivo. To examine the effects of charge neutralization on assembly, we repeated the B3 titration at higher RNA concentration (10 nM vs. 1 nM, Fig. S3*B*). More RNA increases the proportion of $T = 1$ capsid and larger intermediates present at lower CP concentrations, but assembly is still relatively slow and requires CP > 1 μM to be complete. It seems therefore that increasing the numbers of phosphodiesters lowers the barrier to assembly, consistent with it being electrostatic in nature.

As a test of this idea, we compared the salt stability of VLPs assembled at either concentration of B3 or with PS1-5, with that of the VLP from *Escherichia coli*, which packages a modified mRNA encoding the CP gene (Fig. 4*A* and Fig. S5*A*). Addition of NaCl to the reassembled particles leads to rapid disassembly of both the 1 and 10 nM B3 VLPs, which are fully disassembled by ∼150 and ∼400 mM, respectively, whereas the PS1-5 and mRNA VLPs remain stable up to 700 mM. STNV virions contain structural calcium ions that bind to three distinct sites located at the particle symmetry axes (Fig. 1*A*). Note, not all these sites are occupied in the crystal structure of the B3 VLP (15). Because these should contribute to electrostatic stability, we repeated the salt titration in the presence of 5 mM EDTA (Fig. 4*A* and Fig. S5*B*). Both B3 VLPs are fully disassembled under these conditions before addition of extra salt, whereas the PS1-5 VLP partially disassembles at 50 mM NaCl, with only the mRNA VLP remaining stable throughout the titration range. These results confirm that particle stability is strongly dependent on formation of favorable, relatively local, electrostatic contacts, following a sequence-specific assembly initiation. Clearly, the presence of longer RNAs encompassing multiple PS sites stabilizes the CP–CP contacts relative to the VLPs formed with the short RNA.

The results described above are consistent with a two-stage assembly model following a free energy path similar to that shown in Fig. 4*B*. Capsomere formation is rapid, high-affinity, and spontaneous, but for the RNA–CP complex to be on-pathway to capsid assembly, the PS stem-loop must be orientated precisely via sequence-specific interactions with the loop motif. In longer RNAs, this initial event is coupled to a decrease in the size of the viral RNA driven by CP binding to correctly placed additional PSs, a vital step in preparing it for specific encapsidation (7, 8). On-pathway capsomeres with short RNAs are prevented from self-assembly to capsid by an electrostatic activation barrier created by the repulsion of clustering six positively charged amino acid side chains into the small volume around the particle threefold helices (Fig. 1*A*). This inhibitory effect can be overcome using more RNA, either in the form of higher concentrations of single PSs or as natural fragments encompassing multiple, appropriately positioned PS sites. The latter are more effective at lowering the activation barrier, allowing assembly to be cooperative once it initiates. Note, the cooperativity for the complete genomic fragment, with up to 30 PSs, would be even more pronounced than seen here for the 5′ fragment with only five such sites.

The Relationship of the Viral RNA to its Coat Protein

By definition there is a close relationship between the genomic RNA of ssRNA viruses and their coat proteins; one is encoded by the other. For satellite viruses, this is heightened by the fact that the only "viral" RNA encapsidated is an extended mRNA encoding the CP subunit. The polymerase responsible for replication is supplied by the helper virus, Tobacco Necrosis Virus. The STNV genome therefore encompasses at least two functions: It is a substrate for both replication and translation. Previous studies have shown that sequences within the 5′ and 3′ untranslated regions (UTRs) are responsible for making both these processes efficient via complementary base pairing between the UTR residues 24–55 and 979–1009 (27) within the RNA. The site in the 5′UTR overlaps the PS containing region. It is therefore

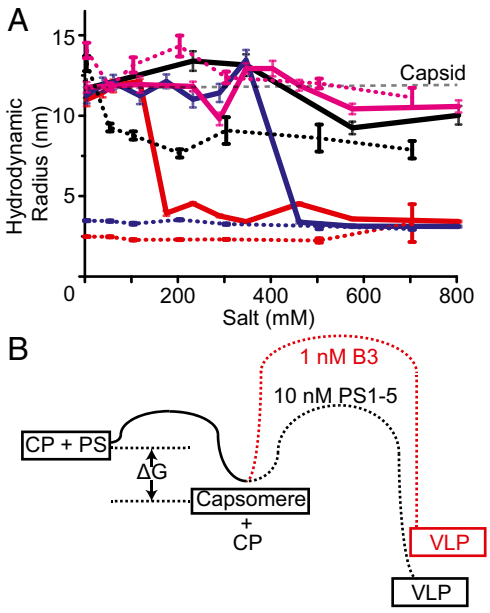


Fig. 4. The roles of electrostatic interactions in assembly of capsomere and capsid. (*A*) Salt titration into mRNA VLPs (magenta), B3 VLPs, assembled at 1:1,000 nM (red) or 10:1,000 nM (blue) RNA:CP, and PS1-5 VLPs assembled at 10:1,000 nM (black). Rh was measured over 10 min at each NaCl concentration. The average and SE of the Rh's are shown. The same experiments, performed in the presence of 5 mM EDTA, are shown as dotted lines. (*B*) Schematic of relative free energy barriers associated with formation of the capsomere and $T = 1$ capsid.

clear that replication/translation and virion assembly are distinct processes brought about by differing RNA conformations.

A global secondary structure of the entire genome calculated using Mfold suggests that only PS1 in the 5′ fragment used here would be formed spontaneously (14). However, if the RNA is folded allowing only for short-range interactions, i.e., in 50-nt sections, then all of the predicted PSs can be formed. The data with the 5′ fragment suggests that STNV packaging occurs on newly replicating RNAs as they emerge from the polymerase. The PS-mediated assembly mechanism also implies that there is a defined spatial relationship between the protein shell of the capsid and the PSs within the genomic RNA. Because ssRNA viruses use RNA-dependent RNA polymerases that lack proof reading functions, the maintenance of such PS sites is a significant challenge. In other systems it has been shown that the coding variants within a viral quasi-species are likely to increase the misincorporation of noncognate RNAs, leading to the suggestion that ssRNA can be horizontal vectors for gene transfer (28, 29). It is thus likely that strategies have evolved that favor conservation of PSs in viral evolution, thus enhancing packaging specificity.

Discussion

There are different ways in which ssRNA viruses may counter the challenge of enhancing packaging of their cognate genomic RNAs against the background of cellular competitor RNAs. One solution could be the physical separation of viral components from cellular factors, but the evidence of significant misincorporation even in viruses believed to operate this way (29) suggests that other mechanisms are likely to have selective advantages. An alternative possibility is that packaging of cognate genomes is the result of specificity between viral genomes and their cognate CPs. Such a mechanism would mirror a phenomenon occurring more widely, in particular in early events in the development of the RNA–protein world: a bias toward interactions between encoded proteins and their mRNAs, even when those proteins are not thought to function via RNA binding (30–32).

We speculate that a similar phenomenon may give rise to PS conservation, and thus ensure packaging specificity, in STNV. In particular, the molecular details of the PS3–CP interaction illustrate how this can be achieved. PS3 is located at nt60 in the STNV-1 genome, and forms part of the codon for amino acid residue 10 in the CP, or 9 neglecting the N-terminal methionine (Fig. 1*A*). Thus, the strongest PS forms a section of the gene encoding the amino acids in the CP that alter their conformation upon PS binding (Fig. 5). This match may of course be coincidental, but a statistical analysis (*SI Materials and Methods*) suggests that it would only occur ∼9% of the time by chance. PS3 is also one of only three PS sites, with PS11 and PS12, conserved in all three known STNV strains (Fig. S6). Testing this idea experimentally would be challenging, so we compared STNV with the two other ssRNA viruses for which we have detailed CP–RNA structural data, namely MS2 and TMV (33, 34) (Fig. S7). The former assembles via a mechanism mediated by multiple PSs, whereas the latter is a helical virus that assembles from a unique PS-RNA contact. There are four known PSs within the MS2 CP gene. One of them encodes a 6-aa peptide that plays a significant role in recognizing the strongest PS in that genome, TR (Fig. 5*B* and Fig. S7 *A* and *C*). The probability that this match occurs by chance in that case is 3.1% (*SI Materials and Methods*). In Group 2 Tobacco Mosaic Viruses, the origin of assembly, a PS, encodes one of the two peptides forming the RNA-binding site in the CP (35–37), suggesting similar relationships occur even in a helical virus (Fig. S7*D*). The occurrence of this phenomenon in a number of viruses suggests that PS encoding features of the protein sites they are in contact with could be an unappreciated aspect of viral CP–genome interactions.

STNV CP shares a common fold with many other ssRNA viruses, most of which are more complex, having quasi-equivalent capsids, in part defined by the conformations of their positively charged N-terminal regions that often contain arginine-rich motifs, ARMs (38). Removal of this region in STNV prevents assembly (39), whereas in other viruses it promotes RNA-independent assembly of $T = 1$ capsids (40). Here we have used smFCS assays to probe the roles of RNA in STNV assembly at nanomolar concentrations, showing that it follows a two-stage PS-mediated mechanism, thus preventing assembly around noncognate RNAs. Even with relatively short PS fragments we have recreated assembly features seen with the intact genome, showing that CP recognition of degenerate PSs, and the latter's relative spacing along the genomic RNA, both contribute to efficient cooperative capsid assembly (9, 41). These assays also validate the use of RNA SELEX against viral CPs, combined with theoretical analysis techniques pioneered by us, as a route for identifying PSs (15).

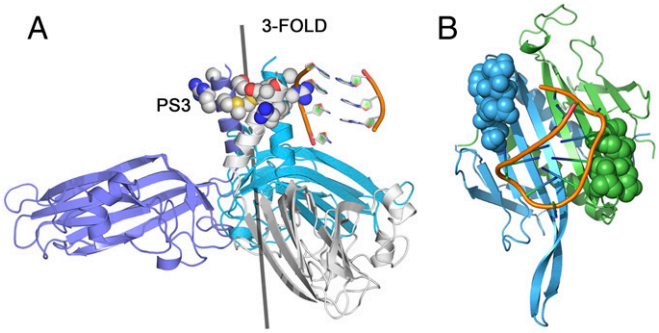


Fig. 5. PSs encode critical features of viral CPs involved in PS RNA recognition. (*A*) View of the contacts between the ordered portion of the B3 RNA (cf. PS3) and the PS3-encoded amino acids (space filling representation) in the B3 VLP structure (PDB ID 4BCU). (*B*) The MS2 coat protein dimer in A/B conformation in complex with the highest affinity genomic PS, the TR stem-loop (34). The amino acids in the A (blue) and B (green) chains that are encoded by one of the four PSs within the CP gene (Fig. S7) are shown in space-filling representation.

It is tempting to speculate that, because of the multiple selective advantages for viruses assembling via such mechanisms, viruses with similar CP folds will also assemble by similar routes (14). The requirement to encode PSs, and perhaps the peptide functions they encode, introduces a previously unsuspected constraint on viral evolution, especially given the density of functions located within the same RNA sequences. Disrupting this assembly mechanism would therefore be a novel antiviral strategy. Previously, we showed that hydrodynamic collapse of viral genomes is a signature of cognate encapsidation, and has many similarities to the assembly of the ribosome. The roles of the multiple genomic PSs revealed here also have similarities to protein sequestration by noncoding (nc) RNAs, "RNA sponges" (42), where proteins bind sequentially, specifically, and cooperatively to stem-loops, leading to protection of a ncRNA from degradation. Assembly of ssRNA viruses may therefore have more in common with cellular RNA biology than previously suspected.

Materials and Methods

STNV CP was prepared by dissociation of recombinant VLPs produced in *E. coli* (15). RNAs were either chemically synthesized with a dye attached to the 5′ end, (DNA technology A/S, Denmark) or transcribed in the presence of 5′-amino-GMP which was subsequently chemically coupled to a dye using an amine reactive Alexa Fluor 488 SDP ester (Invitrogen). SmFCS assays and data analysis were performed as described (8).

ACKNOWLEDGMENTS. We thank Prof. Arwen Pearson for access to the X-ray density and model of the B3 VLP and for helpful comments, and Prof. Cheng Kao and his colleagues at Indiana University for access to their unpublished CLIP-SEQ data on MS2. We thank the University of Leeds for its support of the Single Molecule Facility within the Astbury Centre. This work was supported by UK Biotechnology and Biological Science Research Council Grants BB/J00667X/1, BB/L022095/1, BB/L021803/1, BB/J004596/1 and the Wellcome Trust (089311/Z/09/Z and 090932/Z/09/Z). R. Twarock acknowledges funding via a Royal Society Leverhulme Trust Senior Research fellowship (LT130088), and E.C.D. acknowledges funding via an Early Career Leverhulme Trust fellowship (ECF-2013-019).

1. Woodson SA (2011) RNA folding pathways and the self-assembly of ribosomes. *Acc Chem Res* 44(12):1312–1319.
2. Simon AE, Gehrke L (2009) RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. *Biochim Biophys Acta* 1789(9-10):571–583.
3. Schneemann A (2006) The structural and functional role of RNA in icosahedral virus assembly. *Annu Rev Microbiol* 60:51–67.
4. Enserink M (2014) Infectious diseases. Crippling virus set to conquer Western Hemisphere. *Science* 344(6185):678–679.
5. Lehtinen M, Dillner J (2013) Clinical trials of human papillomavirus vaccines and beyond. *Nature reviews. Clin Oncol* 10(7):400–410.
6. Yan R, Hallam A, Stockley PG, Boyes J (2014) Oncogene dependency and the potential of targeted RNAi-based anti-cancer therapy. *Biochem J* 461(1):1–13.
7. Borodavka A, Tuma R, Stockley PG (2012) Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc Natl Acad Sci USA* 109(39):15769–15774.
8. Borodavka A, Tuma R, Stockley PG (2013) A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA Biol* 10(4):481–489.
9. Dykeman EC, Stockley PG, Twarock R (2014) Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proc Natl Acad Sci USA* 111(14): 5361–5366.
10. Ling CM, Hung PP, Overby LR (1969) Specificity in self-assembly of bacteriophages Q beta and MS2. *Biochemistry* 8(11):4464–4469.
11. Stockley PG, et al. (2007) A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J Mol Biol* 369(2):541–552.
12. Dykeman EC, Stockley PG, Twarock R (2013) Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *J Mol Biol* 425(17):3235–3249.
13. Dent KC, et al. (2013) The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release. *Structure* 21(7):1225–1234.
14. Bunka DH, et al. (2011) Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus: Implications for the assembly of a T=1 capsid. *J Mol Biol* 413(1):51–65.
15. Ford RJ, et al. (2013) Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *J Mol Biol* 425(6):1050–1064.
16. van Lipzig R, et al. (2002) The 5′ and 3′ extremities of the satellite tobacco necrosis virus translational enhancer domain contribute differentially to stimulation of translation. *RNA* 8(2):229–236.
17. Bringloe DH, Gultyaev AP, Pelpel M, Pleij CW, Coutts RH (1998) The nucleotide sequence of satellite tobacco necrosis virus strain C and helper-assisted replication of wild-type and mutant clones of the virus. *J Gen Virol* 79(Pt 6):1539–1546.
18. Bringloe DH, Pleij CW, Coutts RH (1999) Mutation analysis of cis-elements in the 3′- and 5′-untranslated regions of satellite tobacco necrosis virus strain C RNA. *Virology* 264(1):76–84.
19. Meulewaeter F, Danthinne X, Van Montagu M, Cornelissen M (1998) 5′- and 3′-sequences of satellite tobacco necrosis virus RNA promoting translation in tobacco. *Plant J* 14(2):169–176.
20. Rudnick J, Bruinsma R (2005) Icosahedral packing of RNA viral genomes. *Phys Rev Lett* 94(3):038101.
21. van der Schoot P, Bruinsma R (2005) Electrostatics and the assembly of an RNA virus. *Phys Rev E Stat Nonlin Soft Matter Phys* 71(6 Pt 1):061928.
22. Belyi VA, Muthukumar M (2006) Electrostatic origin of the genome packing in viruses. *Proc Natl Acad Sci USA* 103(46):17174–17178.
23. Balint R, Cohen SS (1985) The incorporation of radiolabeled polyamines and methionine into turnip yellow mosaic virus in protoplasts from infected plants. *Virology* 144(1):181–193.
24. Bruinsma RF (2006) Physics of RNA and viral assembly. *Eur Phys J E Soft Matter* 19(3): 303–310.
25. Yoffe AM, et al. (2008) Predicting the sizes of large RNA molecules. *Proc Natl Acad Sci USA* 105(42):16153–16158.
26. Garmann RF, et al. (2014) Role of electrostatics in the assembly pathway of a single-stranded RNA virus. *J Virol* 88(18):10472–10479.
27. Kaempfer R, van Emmelo J, Fiers W (1981) Specific binding of eukaryotic initiation factor 2 to satellite tobacco necrosis virus RNA at a 5′-terminal sequence comprising the ribosome binding site. *Proc Natl Acad Sci USA* 78(3):1542–1546.
28. Routh A, Domitrovic T, Johnson JE (2012) Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc Natl Acad Sci USA* 109(6): 1907–1912.
29. Routh A, Domitrovic T, Johnson JE (2012) Packaging host RNAs in small RNA viruses: An inevitable consequence of an error-prone polymerase? *Cell Cycle* 11(20): 3713–3714.
30. Polyansky AA, Zagrovic B (2013) Evidence of direct complementary interactions between messenger RNAs and their cognate proteins. *Nucleic Acids Res* 41(18):8434–8443.
31. Polyansky AA, Hlevnjak M, Zagrovic B (2013) Proteome-wide analysis reveals clues of complementary interactions between mRNAs and their cognate proteins as the physicochemical foundation of the genetic code. *RNA Biol* 10(8):1248–1254.
32. Hlevnjak M, Polyansky AA, Zagrovic B (2012) Sequence signatures of direct complementarity between mRNAs and cognate proteins on multiple levels. *Nucleic Acids Res* 40(18):8874–8882.
33. Namba K, Pattanayek R, Stubbs G (1989) Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 A resolution by X-ray fiber diffraction. *J Mol Biol* 208(2):307–325.
34. Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* 371(6498):623–626.
35. Fukuda M, Meshi T, Okada Y, Otsuki Y, Takebe I (1981) Correlation between particle multiplicity and location on virion RNA of the assembly initiation site for viruses of the tobacco mosaic virus group. *Proc Natl Acad Sci USA* 78(7):4231–4235.
36. Guilley H, Jonard G, Richards KE, Hirth L (1975) Sequence of a specifically encapsidated RNA fragment originating from the tobacco-mosaic-virus coat-protein cistron. *Eur J Biochem* 54(1):135–144.
37. Takamatsu N, Ohno T, Meshi T, Okada Y (1983) Molecular cloning and nucleotide sequence of the 30K and the coat protein cistron of TMV (tomato strain) genome. *Nucleic Acids Res* 11(11):3767–3778.
38. Rao AL (2006) Genome packaging by spherical plant RNA viruses. *Annu Rev Phytopathol* 44:61–87.
39. Lane SW, et al. (2011) Construction and crystal structure of recombinant STNV capsids. *J Mol Biol* 413(1):41–50.
40. Sorger PK, Stockley PG, Harrison SC (1986) Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly in vitro. *J Mol Biol* 191(4):639–658.
41. Stockley PG, et al. (2013) Packaging signals in single-stranded RNA viruses: Nature's alternative to a purely electrostatic assembly mechanism. *J Biol Phys* 39(2):277–287.
42. Duss O, et al. (2014) Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* 509(7502):588–592.

Biophysical Journal
Article

Biophysical Society

Sizes of Long RNA Molecules Are Determined by the Branching Patterns of Their Secondary Structures

Alexander Borodavka,[1] Surendra W. Singaram,[2,3] Peter G. Stockley,[1] William M. Gelbart,[2] Avinoam Ben-Shaul,[3] and Roman Tuma[1,*]

[1]Faculty of Biological Sciences, Astbury Center for Structural Molecular Biology, University of Leeds, Leeds, United Kingdom; [2]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California; and [3]The Institute of Chemistry and Fritz Haber Research Center, The Hebrew University of Jerusalem, Jerusalem, Israel

ABSTRACT Long RNA molecules are at the core of gene regulation across all kingdoms of life, while also serving as genomes in RNA viruses. Few studies have addressed the basic physical properties of long single-stranded RNAs. Long RNAs with nonrepeating sequences usually adopt highly ramified secondary structures and are better described as branched polymers. To test whether a branched polymer model can estimate the overall sizes of large RNAs, we employed fluorescence correlation spectroscopy to examine the hydrodynamic radii of a broad spectrum of biologically important RNAs, ranging from viral genomes to long noncoding regulatory RNAs. The relative sizes of long RNAs measured at low ionic strength correspond well to those predicted by two theoretical approaches that treat the effective branching associated with secondary structure formation—one employing the Kramers theorem for calculating radii of gyration, and the other featuring the metric of maximum ladder distance. Upon addition of multivalent cations, most RNAs are found to be compacted as compared with their original, low ionic-strength sizes. These results suggest that sizes of long RNA molecules are determined by the branching pattern of their secondary structures. We also experimentally validate the proposed computational approaches for estimating hydrodynamic radii of single-stranded RNAs, which use generic RNA structure prediction tools and thus can be universally applied to a wide range of long RNAs.

INTRODUCTION

The discovery of ribozymes, RNA interference, and riboswitches brought RNA to the forefront of molecular biology by demonstrating that these molecules are ubiquitously involved in a wide range of cellular processes (1–4). Genome sequencing and high-throughput expression profiling have recently revealed novel long noncoding (lnc) RNAs, some of which are thousands of nucleotides long and are known to play important regulatory functions (5,6). For example, Xist lncRNA is a 17 kb-long transcript responsible for silencing one of the homologous pair of X chromosomes during mammalian development (7,8). Others, such as HOTAIR and NRON, are important regulators of gene expression (9–11) linked to diverse human diseases (12). Furthermore, a vast number of important pathogenic viruses including HIV, SARS coronavirus, poliovirus, Dengue fever virus, and many others utilize long RNAs as genetic material, which also play structural roles during virus assembly and genome packaging (13–20). Previous studies have established the importance of local secondary and three-dimensional structure in the biological function of RNA (21,22). However, the effects of the secondary structure on the large-scale properties (e.g., size) of long RNAs remain poorly understood, even while its importance for virus assembly has been demonstrated (13–20).

Several models have been developed to describe properties of double-stranded (ds) and single-stranded (ss) homopolymeric nucleic acids, both of which behave as linear polymers. Coarse-grained properties of long dsRNAs are well described by semiflexible polymer models such as the wormlike chain (23), which only take into account the overall contour length and average persistence length, the latter being weakly dependent on sequence or base composition. Similarly, the freely jointed chain model describes the conformational behavior of the more flexible single-stranded homopolymers (24,25). These models yield simple scaling laws, which relate the contour length (l) or a degree of polymerization (N, number of nucleotides) to the overall

Submitted May 26, 2016, and accepted for publication October 11, 2016.

*Correspondence: r.tuma@leeds.ac.uk

Editor: Tamar Schlick.

http://dx.doi.org/10.1016/j.bpj.2016.10.014

size, e.g., radius of gyration (R_g) or hydrodynamic radius (R_h):

$$R_g \sim R_h \sim b^{(1-\nu)} \times l^{\nu} \sim N^{\nu}. \quad (1)$$

Here ν is a scaling exponent that depends on the polymer chain model (e.g., $\nu = 0.5$ for an ideal Gaussian chain, $\nu = 0.59$ for a self-avoiding chain, and $\nu \sim 1$ for a stiff polyelectrolyte at low ionic strength), and b represents an effective segment length that is related to the persistence length (l_p) and describes polymer flexibility. Highly structured RNAs are described by a collapsed polymer chain model with ν close to 0.33, also applicable to other compact biopolymers such as globular proteins (26,27).

In contrast, due to extensive intramolecular basepairing arising from Watson-Crick complementarity of nucleotides separated by long distances along the chain contour, long ssRNAs fold into effectively branched structures with short duplex regions emanating from single-stranded loops (28) (Fig. 1). Furthermore, given the plethora of possible basepairing scenarios, thermally equilibrated long RNAs are expected to display a large number of secondary structures in solution. Notable exceptions are RNAs in large ribonucleoprotein complexes such as in ribosomes (29) or virus capsids (15,30–35). This view is supported by recent experiments confirming that protein-free viral genomic RNAs adopt an ensemble of branched conformations (28), which are further compacted upon viral assembly (14,20,36,37). Hence, selecting out a unique (native) or representative conformation is less appropriate and useful than averaging over a statistical (thermal) ensemble of secondary structures, for obtaining a reasonable estimate of the overall RNA size.

Here we examine the sizes (hydrodynamic radii, R_h) of a wide range of biologically relevant long RNA molecules at low nanomolar concentration using fluorescence correlation spectroscopy (FCS). The sizes compare well with those predicted by two ensemble averaging methods that take into account the sequence-dependent effective branching of long RNAs. Furthermore, this correlation holds even in the presence of polyvalent cations that enhance tertiary interactions and result in measurable compaction of RNAs, suggesting that these polymer theory-based methods can successfully predict sizes of long RNA molecules under a variety of conditions. Both methods are based on generic RNA structure prediction algorithms and, accordingly, would be widely applicable to other long RNAs with known sequences.

MATERIALS AND METHODS

DNA constructs used for transcribing long RNAs

MS2 phage RNA as well as the 3′ and 5′-end fragments of MS2 phage RNAs were transcribed as described in Borodavka et al. (36). The template for transcription of RpoB RNA was produced by cloning part of the open reading frame of *Escherichia coli* RNA polymerase B subunit gene (*rpoB*), as described in Borodavka et al. (14). The *Xenopus laevis* mRNA was produced by transcribing a plasmid pTRI-Xef, containing the 1.89-kbp elongation factor 1-α gene from *X. laevis* (Ambion/Thermo Fisher Scientific, Carlsbad, CA).

The TCV_pSMART_HC^{Amp} construct (Table S2 in the Supporting Material) was produced by PCR amplifying the full-length TCV cDNA using primers TCV_F1 and TCV_R1 R2 (Table S1) and a pBIN61-based vector, encompassing the full-length TCV cDNA, as a template. pBIN61-TCV plasmid was a gift from Professor George Lomonossoff (John Innes Centre, Norwich, UK). The resulting PCR product was then amplified using 5′-phosphorylated primers TCV_F2 and TCV_R2 (Table S1) to add a T7 promoter sequence to the 5′-end and an *Xho*I restriction site to the 3′-end. Further PCR product purification and cloning into pSMART HC^{Amp} vector were performed the same way as described above for the other DNA templates. Templates for transcription of 16S rRNA and 23S rRNA (16SrRNA_ pSMART_HC^{Amp} and 23SrRNA_ pSMART_HC^{Amp}) were produced by cloning the corresponding genes using genomic DNA extracted from *E. coli* BL21 cells. The primer pairs 16S_F1/16S_R1 and 23S_F1/23S_R1 (Table S1) were designed to amplify region 483879-485408 (16S ribosomal RNA, GenBank: CP001665.1) and region 228583-231490 (23S ribosomal RNA, GenBank: AM946981.2) of the BL21 DE3 *E. coli* genome. The resulting PCR products corresponding to the 16S and 23R rRNA-coding regions were used as templates for a second round of PCR amplification using 5′-phosphorylated primers 16S_F2/16S_R2 and 23S_F2/23S_R2 (Table S1), respectively. This amplification resulted in incorporation of T7 promoter sequences at the 5′-ends of both PCR products and *Dra*I (16S rRNA DNA) and *Hin*dIII (23S rRNA DNA) restriction sites at their respective 3′-ends. Further PCR product purification via agarose gel electrophoresis and subsequent cloning into a pSMART HC^{Amp} vector were performed as described above for other DNA templates. The resulting DNA constructs for in vitro transcription of 16S and 23S rRNAs are 16SrRNA_pSMART HC^{Amp} and 23SrRNA_pSMART HC^{Amp} (Table S2).

DNA template LZRS-HOTAIR (12) encompassing a 2146-nt long human HOTAIR lncRNA sequence (deposited by Professor Howard Chang, Howard Hughes Medical Institute, Stanford University, Stanford, CA) was obtained from the AddGene depository. Primers HotAir_F1 and HotAir_R1 (Table S1) were used to amplify a DNA region, corresponding to the human HOTAIR lncRNA using Q5 high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA), as described above. The resulting PCR product was used as a template in a second amplification with 5′-phosphorylated primers HotAir_F2 and HotAir_R2 (Table S1), which resulted in addition of T7 promoter sequence at the 5′-end and an *Eco*RV restriction site at the 3′-end. The obtained PCR product was agarose gel-purified and used for a subsequent ligation into a pSMART HC^{Amp} vector as described above, following the manufacturer protocols. The XL1 Blue competent cells (Agilent Technologies, Santa Clara, CA) were used for transformation with the ligated products, the resulting transformants were PCR-screened, and the positive clones were verified by DNA sequencing. The resulting construct HOTAIR_pSMART HC^{Amp} (Table S2) was used for in vitro transcription of the human HOTAIR lncRNA.

cDNA for lncRNA NRON was produced by reverse-transcribing phenol-chloroform extracted total RNA from HEK 293 cells using Superscript III Reverse Transcriptase and random hexamer oligonucleotide primers (Invitrogen, Carlsbad, CA), following the manufacturer's protocol. Primers NRON_F and NRON_R (Table S1) were used to amplify the resulting cDNA using Q5 high-fidelity DNA polymerase. The resulting PCR product was agarose gel-purified and used for a subsequent ligation into a pJET1.2 vector using a CloneJET PCR Cloning Kit (Thermo Fisher Scientific, formerly Fermentas), following the manufacturer protocol. The XL1 Blue-competent cells were transformed with the resulting ligated products. The transformants were csPCR-screened and the positive plasmid clones were verified by DNA sequencing. The resulting DNA construct NRON_pJET1.2^{Amp} (Table S2) was used for in vitro transcription of the human NRON lncRNA.

Several DNA constructs for in vitro transcription were generously provided upon request by various research groups. The DNA template for

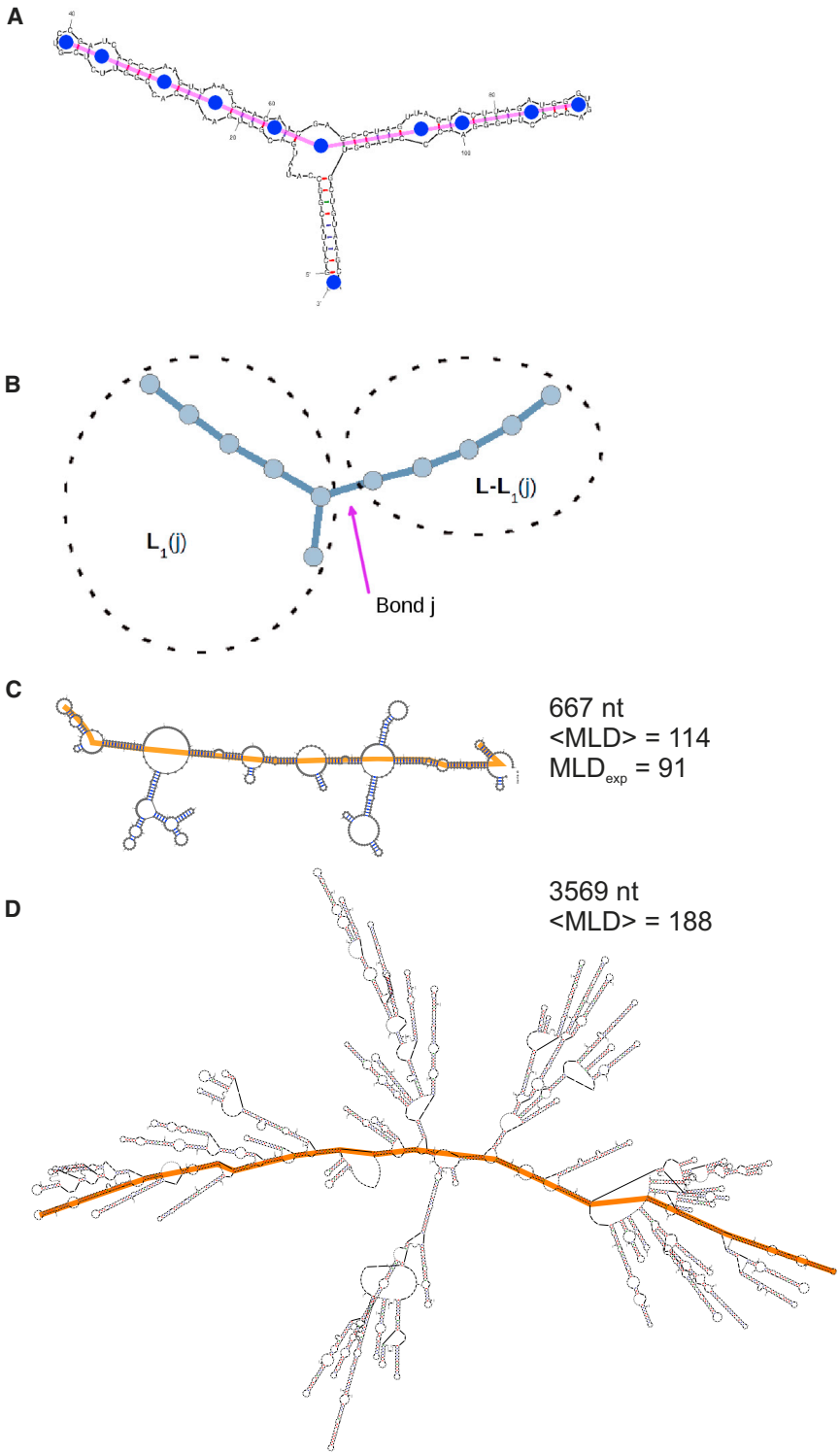


FIGURE 1 Schematics of an RNA molecule as a branched polymer. (*A*) Minimum free energy secondary structure with the maximum ladder path highlighted in magenta and flexible joints or branch points as blue dots. (*B*) Tree graph representation of the secondary structure in (*A*), with illustration of the partitioning into two halves ($L_1(j)$ and $L\text{-}L_1(j)$) at bond j for R_g computation using the Kramers theorem (see Materials and Methods). (*C*) An experimentally determined secondary structure of segment 11 (60) with maximum ladder path highlighted, and experimental MLD_{exp} and predicted $\langle MLD \rangle$ compared. (*D*) A representative secondary structure prediction for MS2 genomic RNA and predicted $\langle MLD \rangle$. To see this figure in color, go online.

production of the STNV-C genomic RNA was a gift from Dr. Robert Coutts (38). DNA construct HCV JFH1/Luc SGR was a gift from Professor Mark Harris (University of Leeds, Leeds, UK). DNA constructs pUC19T7RFs1 and pUC19T7RFs11 were a gift from Dr. Ulrich Desselberger (39) (University of Cambridge, Cambridge, UK). DNA constructs pF2100 and P2BS WT were donated by Professor Anette Schneemann (The Scripps Research Institute, La Jolla, CA). DNA constructs pT7riboBUN-S and BUNVL were a gift from Dr. John Barr (University of Leeds). All DNA constructs with their respective linearization restriction enzymes, used for in vitro transcription of long RNAs, are summarized in Table S2. The scrambled s11 sequence was synthesized as a gene block DNA and inserted into a pUC19 vector under control of a T7 promoter. Sequences and base compositions are summarized in the Supporting Material.

Transcription and fluorescent labeling of ssRNAs

In vitro transcription reactions were carried out using a T7 RNA transcription kit (HiScribe T7 or T3 High Yield; New England Biolabs) following the manufacturer's protocol. RNAs were purified using RNeasy mini kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol, except for the fluorescently labeled RNAs. In those samples the RNA-loaded column was washed four times with 80% (v/v) ethanol before elution with 30 μL of sterile nuclease-free water. MS2-derived RNAs were 3′-end labeled while all others were 5′-end amine-modified RNAs produced by incorporation of amino-GMP and fluorescently labeled as described in Borodavka et al. (14). All RNA samples were routinely examined on denaturing formaldehyde agarose gels to ensure their integrity. Every precaution was taken to avoid contamination with RNases, and RNA samples were kept as 10 μL aliquots at −80°C to minimize degradation.

FCS data collection and analysis

FCS measurements were performed on a custom-built FCS confocal setup. The excitation laser (Sapphire CW blue laser, 488 nm; Coherent, Bloomfield, CT) power was set to 65 μW. The immersion oil objective (63× magnification, numerical aperture of 1.4; Carl Zeiss, Jena, Germany) was used together with low autofluorescence immersion oil (refractive index 1.515, type DF; Cargille-Sacher Laboratories, Cedar Grove, NJ). The focus position was adjusted to 20 μm from the coverslip inner surface and precisely maintained by a piezoelectric feedback loop (Piezosystem, Jena, Germany). The photon count was recorded and analyzed by an ALV-5000 multiple tau digital correlator (www.alvgmbh.de) used in a single channel mode. Multiple runs of up to 100 autocorrelation functions with acquisition scan time of 30 s each were recorded for each of the samples using ALV-correlator software (ALV-5000/E/EPP, Ver. 3.0). Calibration of the confocal volume was performed by measuring the diffusion time of AF488-SDP dye (1 nM in RNA measurement buffer) before each data set collection. FCS data were analyzed by nonlinear least-squares fitting with a single-component diffusion model autocorrelation function corrected for the triplet state (14) using MATLAB (The MathWorks, Natick, MA). Calculation of R_h was based on the measured diffusion time value for AF488 dye and the established diffusion coefficient for a free dye using the Einstein-Stokes relationship.

RNA measurements were performed with 0.5–2 nM RNA in RNase-free 20 mM 3-(*N*-morpholino)-propanesulfonic acid (MOPS), 10 mM KOH buffer, pH 7.0 with 1 mM dithiothreitol at 25°C. RNA condensation experiments were performed in the presence of divalent (10 mM $MgCl_2$, Mg^{2+}) and trivalent (1 mM spermidine chloride, Sp^{3+}) cations, added to the 0.5–2 nM RNA samples before FCS measurements.

Theory

To account for the conformational statistics associated with an ensemble of secondary structures, it is useful (28,40–43) to represent the RNA secondary structure as a tree graph (44), i.e., a collection of points (vertices) each of which is connected by a line (bond) to at least one other point, without any closed paths. Fig. 1 illustrates this mapping for a simple case: here duplexes are treated as rigid bonds of the same length—tree edges, and single-stranded flexible loops are treated as tree vertices. Hairpin loops are vertices of order one; loops, including bulges, connecting two duplexes are twofold vertices; and loops from which three or more duplexes emanate are branched vertices (see Fig. 1 *A*). To calculate the size of the resulting branched polymer (Fig. 1 *B*), two approaches can be used. The first method makes use of the Kramers theorem (41,45,46) to directly calculate R_g from the tree topology. In the second method, the size is determined by identifying the longest chain of edges found within the tree—defined as the maximum ladder distance (MLD, Fig. 1 *A*) (42,43)—and the branched tree is replaced by a linear chain with effective contour length (N_{eff}) proportional to the MLD. Treating the resulting linear polymer as an ideal chain then gives

$$R_g = \left(b^2 N_{\mathrm{eff}}/6\right)^{1/2}. \quad (2)$$

Here the segment length b corresponds to the average length of a duplex ($\approx$5 bp) (17,28,47) and N_{eff} is the number of duplexes along the MLD, which is $N_{eff} = MLD/b$. Thus,

$$R_g = \left(b^2 MLD/6b\right)^{1/2} \sim (MLD)^{1/2} \quad (3)$$

in bp units (42,46). The MLD is estimated from RNA secondary structure predictions and can be further refined using structure probing experiments (21). Because there is heterogeneity among the many structures whose energies lie within a thermally available range (k_BT), we use the Boltzmann-averaged MLD (denoted $\langle MLD\rangle$), derived from an ensemble of RNA structures generated by prediction algorithms implemented in RNAfold (48). Earlier theoretical analyses have shown that while even the most sophisticated and accurate basepairing programs begin to fail for long RNAs like those treated here, the relative values of their $\langle MLD\rangle$ and R_g can still be meaningfully estimated (41,42).

Size computations

Average $\langle MLD\rangle$ values were computed from the 100 lowest-energy secondary structures (42) generated using the Vienna package (48). Relative values of R_g were estimated using (see Eq. 3) the relationship $R_g \sim (\langle MLD\rangle)^{1/2}$. Each tree graph representation was derived from a dot-bracket representation of the secondary structure (see the Vienna RNA web server manual at http://rna.tbi.univie.ac.at/help.html). The R_g was calculated from the tree graph by treating the vertices as perfectly flexible joints and the edges as rigid phantom bonds (i.e., as an ideal branched polymer), and using the Kramers theorem (46). More explicitly, the R_g of a branched polymer (tree graph) was calculated by

$$\overline{R_g^2} = \left(b^2/L^2\right)\Sigma_j L_1(j)[L - L_1(j)], \quad (4)$$

where the overbar denotes an average over all conformations of the ideal branched polymer. The sum in Eq. 4 is evaluated by summing over all L bonds the product of $L_1(j)$ and L-$L_1(j)$, the numbers of vertices on either side of the jth bond (see Fig. 1 *B*). The square root of Eq. 4 yields the radius of gyration of the tree graph (i.e., $\widehat{R}_g = \sqrt{R_g^2}$). We then averaged $\widehat{R}_g$ over the tree graphs we generated from the secondary structures, which for simplicity we refer to as the R_g (i.e., $R_g \equiv \langle \widehat{R}_g \rangle$). The predicted R_g values are reported in units of the average duplex length b.

RESULTS AND DISCUSSION

Due to their large sizes and high conformational flexibility, little is known about the structural organization and physical properties of long RNAs. Some of them, such as viral positive-sense ssRNA genomes, adopt compact conformations as part of their function and facilitate packaging into the confined space of icosahedral viral capsids (17). Likewise, several lncRNAs, including HOTAIR and SRA, assume well-defined conformations with separate domains, capable of folding into compact structures upon addition of divalent cations (49,50). These independent domains interact with their binding partners via evolutionarily conserved protein-binding motifs (49). To better understand

the architecture of long RNA molecules, e.g., their overall compactness or extendedness, we explore the relation between predicted sizes, using either the MLD or R_g obtained from Kramers theorem, respectively, and the experimentally determined hydrodynamic radii (R_h) for long RNAs, ranging from 600 to >9000 nucleotides in length.

We have examined a wide range of biologically relevant RNAs, including messenger, long noncoding, viral, and ribosomal RNAs. To minimize nonspecific intermolecular interactions between RNA molecules, we employ extremely dilute solutions (low nanomolar concentrations) and low ionic strength (i.e., good solvent conditions for charged polymers), and measure sizes of RNA molecules by FCS. In contrast to other ensemble solution techniques (small-angle x-ray and light scattering, and analytical centrifugation), the dilute conditions minimize aggregation due to intermolecular basepairing, which has previously been shown to result in an overestimation of sizes (51). Furthermore, we have also used FCS to examine compaction of individual RNA molecules in response to biologically relevant divalent (Mg^{2+}) and trivalent (spermidine, Sp^{3+}) cations. The latter conditions promote formation of tertiary structures (52,53).

Table 1 summarizes calculated $\langle MLD \rangle$ values and measured hydrodynamic radii for a range of long RNA molecules examined by FCS. Due to the low RNA concentrations and ionic strength conditions used here (notably nonphysiological, by design), aggregation and tertiary structure formation are unlikely, so that the effects of branching due to secondary structure can be accentuated and be probed directly under close-to-isolated molecule (infinite-dilution) conditions. We note that while the measured R_h broadly increases with the length, the rise significantly deviates from the monotonic behavior expected for the simple scaling laws (Eq. 1, Fig. 2 *A*). This result suggests that linear polymer scaling laws (Eq. 1) are not appropriate to describe long ssRNA, which is an effectively branched polymer. Instead, essential coarse-grained features of their sequences need to be taken into account.

To account for sequence variations, basepairing, and the resulting branching, we estimate branching patterns using the output of secondary structure algorithms (RNAfold)

TABLE 1 Hydrodynamic Radii Measured by FCS and Average Computed MLDs

Number	RNA[a]	Class[b]	Length (kb)	% BasePaired[c]	R_h Low Salt[d,e] (nm)	R_h Mg^{2+} (nm)[e,f]	R_h Sp^{3+} (nm)[e,g]	⟨MLD⟩ (rungs)[h]	R_g (a.u.)[i]
1	RV s11	ds	0.67	58	8.2 ± 1.1	11.2 ± 3.5 (9.6 ± 2)	7 ± 1.6 quenching[g]	114 ± 6	2.10
2	RV s11 scrambled	ds	0.67	56	6.5 ± 1.4	—	—	83 ± 6	2.1
3	BunVS	ss	0.96	65	10.0 ± 1.6	7.2 ± 2.1	8 ± 3.3	134 ± 11	2.23
4	STNV	ss	1.2	62	11.7 ± 1.0	8.5 ± 1.7	9 ± 2	154 ± 7	2.39
5	FHV2	ss	1.4	62	11.9 ± 2.0	9.4 ± 2.6	8.3 ± 2	176 ± 24	2.76
6	Ef2	m	1.8	60	8.8 ± 1.4	9.4 ± 1.6	9.7 ± 1.6	184 ± 14	3.12
7	16S rRNA	r	1.55	64	17.5 ± 4.0	14 ± 4.8	quenching[g]	149 ± 26	2.56
8	HOTAIR	lnc	2.4	61	16.2 ± 2.0	12.5 ± 2.4	13.4 ± 4.7	264 ± 19	3.39
9	5′-MS2	ss	2.5	69	10.7 ± 1.2	9.8 ± 0.6	10.3 ± 1.7	167 ± 17	2.74
10	3′-MS2	ss	2.6	69	13.8 ± 1.3	10.8 ± 0.8	10.5 ± 1	159 ± 9	2.68
11	NRON	lnc	2.6	58	17.6 ± 2.7	15.3 ± 3	13.7 ± 2.6	212 ± 11	3.14
12	23S rRNA	r	2.9	63	14.2 ± 2.5	11.3 ± 2.2	quenching[g]	252 ± 24	3.25
13	FHV 1	ss	3.1	62	15.6 ± 2.0	11.7 ± 4.3	9.6 ± 3.4	224 ± 13	3.12
14	RV s1	ds	3.3	58	18.4 ± 3.4	15.3 ± 2.2	18.1 ± 9 aggregation[g]	319 ± 24	3.66
15	MS2	ss	3.6	69	12.3 ± 0.6	11.3 ± 1.7	9.2 ± 1	188 ± 18	2.92
16	RpoB	m	3.6	64	18.3 ± 2.7	12 ± 1.2	10.6 ± 2	289 ± 20	3.69
17	TCV	ss	4.5	63	16.5 ± 1.7	14.7 ± 4.5	12.4 ± 4.5	341 ± 21	3.85
18	BunV L	ss	6.9	59	14.7 ± 2.4	11.7 ± 1.8	12.5 ± 2	375 ± 17	4.03
19	HCV	ss	8.9	64	33.1 ± 5.3	20.1 ± 2.6	18.8 ± 2.8	567 ± 43	4.81

[a]RV s1 and s11, human Rotavirus segment 1 and 11 precursors (single-stranded); BunVS and BunVL-Bunyamwera virus, small and large segment precursors, respectively (single-stranded); STNV, Satellite Tobacco Necrosis Virus genomic RNA; FHV1 and FHV2, Flock House Virus RNA1 and 2; Ef2 mRNA, *X. laevis* Ef2 gene transcript; 5′-MS2- 5′ end of MS2 phage genomic RNA (nucleotides 1–2469); 3′-MS2- 3′ end of MS2 phage genomic RNA (nucleotides 992–3569); TCV, Turnip Crinkle Virus genomic RNA; and HCV, Hepatitis C Virus genomic RNA.

[b]ds, single-stranded precursors of dsRNA viral genomes; ss, genomes of ssRNA viruses; m, cellular mRNA; r, ribosomal RNA; and lnc, long noncoding RNA.

[c]Percentage of basepairing averaged over 100 predictions.

[d]Measured in 20 mM MOPS-K^+, pH 7.

[e]The values are reported as average ± SD computed from at least 10 measurements. Long RNA molecules were transcribed and 5′ (or 3′; see Materials and Methods) end-labeled with Alexa Fluor 488 (Thermo Fisher Scientific), purified and subsequently checked for integrity by denaturing agarose gel electrophoresis (Fig. S1). In a few cases, quenching or aggregation affected or prevented determination of the diffusion correlation time.

[f]Measured in 10 mM $MgCl_2$ in 20 mM MOPS-K^+, pH 7.

[g]Measured in 1 mM spermidine in 20 mM MOPS-K^+, pH 7.

[h]Computed by averaging over 100 predictions (± SD).

[i]Computed using Kramers theorem.

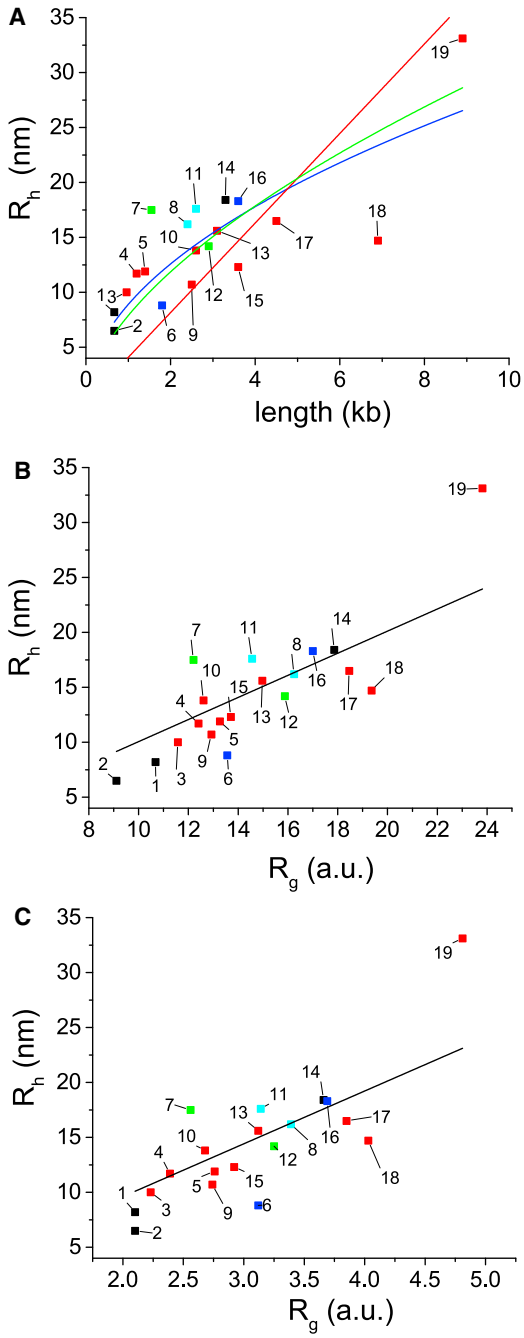


FIGURE 2 (A) Measured R_h as a function of nucleotide length (in kb). Numbering of RNAs is according to their increasing length (Table 1) and coloring is according to the class (*black*, single-stranded precursors of dsRNA viral genomes; *red*, genomes of ssRNA viruses; *blue*, cellular mRNAs; *green*, ribosomal RNA; and *cyan*, long noncoding RNAs. Lines and curves represent best fits to different linear polymer models: charged (*red*, Eq. 1, $\nu = 1$, reduced $\chi^2 = 35.85$), simple Gaussian coil (*blue*, Eq. 1, $\nu = 0.5$, reduced $\chi^2 = 13.37$), and a self-avoiding coil (*green*, Eq. 1, $\nu = 0.59$, reduced $\chi^2 = 14.85$). (*B*) Correlation between R_h and R_g predicted from $\langle MLD \rangle$ (in bp units); solid line is the best fit with reduced $\chi^2 = 11.28$. (*C*) Correlation between R_h and R_g predicted from Kramers theorem (in units of the average segment length, a.u.); solid line is the best fit with reduced $\chi^2 = 12.74$. RNA color coding and numbering is the same as in (*A*). Error bars were omitted for clarity; see Table 1 for standard deviations. To provide directly comparable reduced χ^2 values, all fitting was performed using the same nonlinear Levenberg-Marquardt algorithm in OriginPro (OriginLab, Northampton, MA). To see this figure in color, go online.

and ensemble average over the low energy structures. The measured hydrodynamic radii are in good agreement with the theoretical estimates of the R_g values based either on the MLD (Fig. 2 *B*, Eq. 3) or the Kramers theorem (Fig. 2 *C*). This result is consistent with most RNA molecules adopting branched structures in which the MLD largely determines the overall size (41,42). This is illustrated by comparing the maximum ladder path of rotavirus segment 11 precursor (s11, Fig. 1 *C*) with that of MS2 phage genomic RNA (Fig. 1 *D*). The experimentally determined secondary structure pattern of s11 is significantly less branched than that of the typical prediction for MS2. Furthermore, this is reflected in the relatively large MLD and hydrodynamic size of s11, comparable to that of Ef2 mRNA, which is three times the length. This demonstrates that the relatively simple MLD description can capture the essence of coarse-grained RNA structure, and yields quantitative predictions based on the RNA sequence alone.

Further compaction of RNA molecules and the formation of tertiary structure require di- and polyvalent cations (Mg^{2+}, spermidine^{3+}, spermine^{4+}) and/or association with RNA-binding proteins (54). As seen in Table 1, upon addition of divalent (10 mM Mg^{2+}) or trivalent cations (1 mM spermidine, Sp^{3+}), the measured R_h decreases for most RNAs, consistent with compaction driven by electrostatic screening and neutralization. Fig. 3 compares R_h before and after the addition of multivalent cations. The R_h values cluster along the line with the slope between 0.7 and 0.8, indicating that on average the RNAs undergo a 20–25% size compaction compared to their original R_h. As a consequence, the proportionality between R_h and predicted size holds for most of the RNAs even after addition of polyvalent cations. However, there are few RNAs that either fail to further compact (RV s11 No. 1 and Ef2 mRNA No. 6 in Fig. 3, where the R_h change is insignificant at confidence level 90%) or the compaction is more prominent in comparison with other RNAs examined (HCV, No. 19 in Fig. 3, where the R_h differs significantly from the expected value at confidence level 99%).

The quantitative relation between the experimental R_h and R_g predicted either from $\langle MLD \rangle$ or the Kramers theorem indicates that modeling the RNA as an ideal branched polymer constitutes a good starting point for predicting the overall size of long RNAs. However, there are several notable discrepancies between the predicted and measured sizes. One limitation of our approach is that computational predictions may yield an incorrect structure and hence an MLD that differs from that of the experimentally determined secondary structure, as in the case of STMV RNA (21,55). Such failures of the computational approach are more likely

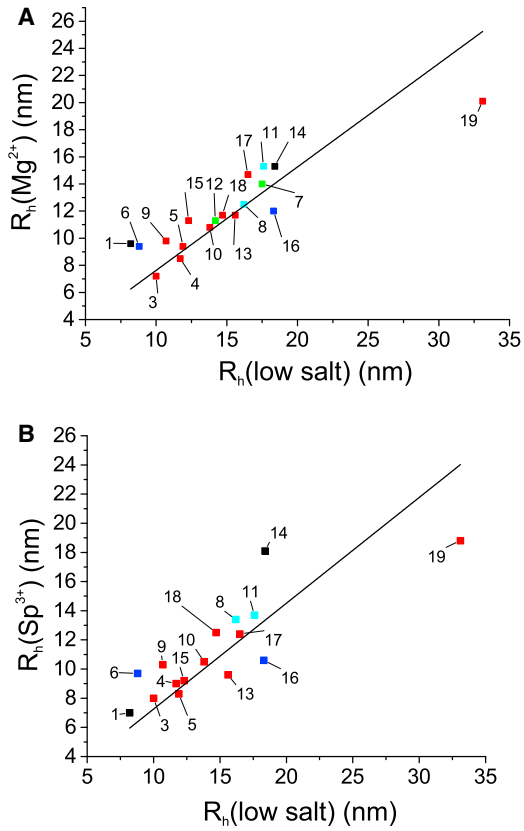


FIGURE 3 Hydrodynamic size reduction in the presence of Mg^{2+} (*A*) or spermidine Sp^{3+} (*B*). Coloring and numbering scheme is as in Fig. 2. R_h values that were compromised by either quenching or possible aggregation (RV s1 and s11 in Table 1) in the presence of multivalent cations were omitted from the plot. Linear regression lines with slopes 0.77 ± 0.03, Pearson's $r = 0.89$ for Mg^{2+} and 0.73 ± 0.04, and Pearson's $r = 0.87$ for Sp^{3+}, respectively, are shown. To see this figure in color, go online.

to occur when long RNA sequences are analyzed, thus explaining the largest deviations observed for HCV and BunV L (>5 kb), the longest RNAs examined here (Fig. 2 *B*; Table 1) (17,28). This situation can be remedied by estimating MLD from structure-probing data, which should improve the accuracy of RNA size calculations. This is demonstrated for s11 RNA for which the probing-derived MLD_{exp} is slightly lower than the computed average $\langle MLD \rangle$ (Fig. 1 *C*, compare $\langle MLD \rangle$ and MLD_{exp}), yielding a R_g ~ 9.3 value that agrees better with the experimental R_h (i.e., point No. 1 would be closer to the trend line in Fig. 2 *B*). On the other hand, when secondary structure determination (or prediction) is ambiguous, experimental size measurements by FCS can be used for selecting those structures with MLDs compatible with the experimentally determined hydrodynamic radii. In addition to the MLD prediction limitations, HCV RNA is compacted twofold in the presence of multivalent ions, i.e., to a much higher degree than other RNAs examined here (20–25% reduction), highlighting the importance of repulsive electrostatic interactions at low salt and formation of tertiary contacts stabilized by multivalent cations, not accounted for in our approach.

Another case of underestimating the size is 16S rRNA, which is predicted to be more compact than experimentally observed (Fig. 2, *B* and *C*, No. 7, R_h ~ 17.8 nm). This discrepancy likely reflects the presence of distinct domains within 16S rRNA that make the protein-free 16S rRNA relatively large (measured R_g ~11.4 nm). Only upon binding of multiple ribosomal proteins does it undergoes gradual compaction to its fully folded functional state (R_g ~ 7 nm) (56,57).

Ef2 mRNA is an example of overestimated size (Fig. 2, No. 6) and in this case tertiary contacts involving long-range interactions—not accounted for in our analysis—may play important roles in maintaining its compactness. The observed lack of further compaction of Ef2 mRNA in the presence of multivalent cations is consistent with preformed stable intramolecular contacts present in this RNA (Fig. 3, No. 6; Table 1).

To further test the observed correlation between $\langle MLD \rangle$ and the experimental hydrodynamic radius, we generated a scrambled s11 RNA sequence, as described in Materials and Methods. This disrupted as much as 25% of the original base pairings in the experimentally probed secondary structure of s11 (see sequence in the Supporting Material), while maintaining a similar level of basepairing (only 2% reduction of overall base pairing, Table 1). Analysis of the scrambled RNA sequence yields a reduction of $\langle MLD \rangle$ and is reflected in a concomitant decrease (significant at 99% confidence level) of experimentally measured R_h (Table 1; Fig. 2, RNA No. 2). In this case the native, original fold of s11 RNA is an extended conformation while the scrambled sequence produces an ensemble of more branched and hence more compact species, further demonstrating the predictive power of the MLD approach. However, the Kramers theorem approach fails to predict this reduction, most likely due to assuming the same average phantom bond length between vertices (compare to Fig. 1, *A* and *B*).

Overall, the observed differences in compactness and extendedness of RNA molecules may reflect various biological functions they perform. While MS2 phage genomic and subgenomic ssRNAs (Nos. 9, 10, and 15 in Fig. 2 *A*) are comparable in length to lncRNAs (Nos. 8 and 11) and the protein-free ribosomal RNA (No. 12), they appear to be smaller in size (see 2–4 kb region in Fig. 2 *A*). However, the Ef2 mRNA transcript (No. 6) is yet even more compact than the comparatively short viral RNAs (RNAs No. 4, 5, and 13 in Fig. 2 *A*), suggesting that although there might be evolutionary pressure on genomes of ssRNA viruses to fold into more compact structures (17), there is a number of exceptions, including more extended viral RNAs (21) and compact mRNAs. Moreover, relative size of viral RNAs may also reflect replication strategies and genome packaging mechanisms employed by viruses. For example, viruses with segmented RNA genomes may preferentially

utilize extended, less branched RNA conformations for their segment precursors (s11 in Fig. 1 *C*) to minimize the formation of nonspecific intersegment RNA-RNA contacts, while enabling formation of specific interactions facilitated by the viral RNA chaperones (58).

Remarkably, despite significant differences in the architectures of various long RNAs, we find that their sizes (hydrodynamic radii) can be estimated using coarse-grained theoretical predictions, even in the presence of multivalent ions stabilizing tertiary contacts. Because the theoretical approaches used here treat exclusively the branching patterns associated with the RNA secondary structures, our results provide experimental evidence that the overall sizes of long RNAs are determined predominantly by their secondary structure branching patterns (17). The effects of di- and polyvalent cations are more prominent for smaller RNAs, such as riboswitches and ribozymes, which adopt compact and unique tertiary structures in the presence of Mg^{2+} (59) via formation of specific tertiary contacts. Due to the heterogeneity of secondary structures in long RNAs, such specific contacts would be harder to achieve, while also explaining why long RNAs often require auxiliary proteins to guide their folding into a unique structure. This feature of RNA is likely to be the result of a limited repertoire of interactions offered by the four nucleobases and points to a fundamental limitation of RNA as a complex biopolymer when compared to proteins. We find that even relatively simple theoretical calculations based on ensembles of predicted secondary structures and MLD averaging correlate well with the experimental measurements for a diverse set of long RNA molecules, allowing our approach to account for the sizes and compactness of broad classes of ssRNAs.

SUPPORTING MATERIAL

One figure and two tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)30941-9.

AUTHOR CONTRIBUTIONS

A.B., S.W.S., P.G.S., W.M.G., A.B.-S., and R.T. designed research; A.B. and S.W.S. performed research; A.B.-S. and R.T. contributed analytic tools; and A.B., S.W.S., P.G.S., W.M.G., A.B.-S., and R.T. analyzed data and wrote the article.

ACKNOWLEDGMENTS

We thank Professor Mark Harris and Dr. John Barr (University of Leeds, UK), Dr. Ulrich Desselberger (University of Cambridge, UK), and Professor Anette Schneemann (The Scripps Research Institute, La Jolla, CA), for kindly donating plasmids JFH1/Luc SGR; pT7riboBUN-S, and pT7riboBUN-L; pUC19T7RFs1 and pUC19T7RFs11; and pF2100 and P2BS WT, which were used as templates for transcription of some of the viral RNAs.

This work was supported by the Wellcome Trust (grant Nos. 089310/09/Z and 103068/Z/13/Z to A.B.) and the Biotechnology and Biological Sciences Research Council (BBSRC) (grant No. BB/J00667X/1 to P.G.S. and R.T.).

REFERENCES

1. Zaug, A. J., and T. R. Cech. 1986. The intervening sequence RNA of Tetrahymena is an enzyme. *Science.* 231:470–475.
2. Fire, A., S. Xu, …, C. C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature.* 391:806–811.
3. Winkler, W., A. Nahvi, and R. R. Breaker. 2002. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature.* 419:952–956.
4. Lee, R. C., R. L. Feinbaum, and V. Ambros. 1993. The *C. elegans* heterochromic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell.* 75:843–854.
5. Rinn, J. L., and H. Y. Chang. 2012. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81:145–166.
6. Necsulea, A., M. Soumillon, …, H. Kaessmann. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature.* 505:635–640.
7. Clemson, C. M., J. A. McNeil, …, J. B. Lawrence. 1996. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell Biol.* 132:259–275.
8. Vallot, C., and C. Rougeulle. 2013. Long non-coding RNAs and human X-chromosome regulation: a coat for the active X chromosome. *RNA Biol.* 10:1262–1265.
9. Rinn, J. L., M. Kertesz, …, H. Y. Chang. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 129:1311–1323.
10. Willingham, A. T., A. P. Orth, …, P. G. Schultz. 2005. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science.* 309:1570–1573.
11. Wapinski, O., and H. Y. Chang. 2011. Long noncoding RNAs and human disease. *Trends Cell Biol.* 21:354–361.
12. Gupta, R. A., N. Shah, …, H. Y. Chang. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature.* 464:1071–1076.
13. Singaram, S. W., R. F. Garmann, …, A. Ben-Shaul. 2015. Role of RNA branchedness in the competition for viral capsid proteins. *J. Phys. Chem. B.* 119:13991–14002.
14. Borodavka, A., R. Tuma, and P. G. Stockley. 2012. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc. Natl. Acad. Sci. USA.* 109:15769–15774.
15. Dykeman, E. C., P. G. Stockley, and R. Twarock. 2014. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proc. Natl. Acad. Sci. USA.* 111:5361–5366.
16. Harvey, S. C., Y. Zeng, and C. E. Heitsch. 2013. The icosahedral RNA virus as a grotto: organizing the genome into stalagmites and stalactites. *J. Biol. Phys.* 39:163–172.
17. Gopal, A., D. E. Egecioglu, …, W. M. Gelbart. 2014. Viral RNAs are unusually compact. *PLoS One.* 9:e105875.
18. Comas-Garcia, M., R. F. Garmann, …, W. M. Gelbart. 2014. Characterization of viral capsid protein self-assembly around short single-stranded RNA. *J. Phys. Chem. B.* 118:7510–7519.
19. Cadena-Nava, R. D., M. Comas-Garcia, …, W. M. Gelbart. 2012. Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. *J. Virol.* 86:3318–3326.
20. Patel, N., E. C. Dykeman, …, P. G. Stockley. 2015. Revealing the density of encoded functions in a viral RNA. *Proc. Natl. Acad. Sci. USA.* 112:2227–2232.

21. Athavale, S. S., J. J. Gossett, …, S. C. Harvey. 2013. In vitro secondary structure of the genomic RNA of satellite tobacco mosaic virus. *PLoS One.* 8:e54384.

22. Ding, Y., Y. Tang, …, S. M. Assmann. 2014. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature.* 505:696–700.

23. Bustamante, C., J. F. Marko, …, S. Smith. 1994. Entropic elasticity of λ-phage DNA. *Science.* 265:1599–1600.

24. Seol, Y., G. M. Skinner, and K. Visscher. 2004. Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys. Rev. Lett.* 93:118102.

25. Sim, A. Y. L., J. Lipfert, …, S. Doniach. 2012. Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 86:021901.

26. Hyeon, C., R. I. Dima, and D. Thirumalai. 2006. Size, shape, and flexibility of RNA structures. *J. Chem. Phys.* 125:194905.

27. Werner, A. 2011. Predicting translational diffusion of evolutionary conserved RNA structures by the nucleotide number. *Nucleic Acids Res.* 39:e17.

28. Gopal, A., Z. H. Zhou, …, W. M. Gelbart. 2012. Visualizing large RNA molecules in solution. *RNA.* 18:284–299.

29. Schluenzen, F., A. Tocilj, …, A. Yonath. 2000. Structure of functionally activated small ribosomal subunit at 3.3 Ångstroms resolution. *Cell.* 102:615–623.

30. Toropova, K., G. Basnak, …, N. A. Ranson. 2008. The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. *J. Mol. Biol.* 375:824–836.

31. Toropova, K., P. G. Stockley, and N. A. Ranson. 2011. Visualising a viral RNA genome poised for release from its receptor complex. *J. Mol. Biol.* 408:408–419.

32. Dent, K. C., R. Thompson, …, N. A. Ranson. 2013. The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release. *Structure.* 21:1225–1234.

33. Devkota, B., A. S. Petrov, …, S. C. Harvey. 2009. Structural and electrostatic characterization of Pariacoto virus: implications for viral assembly. *Biopolymers.* 91:530–538.

34. Johnson, J. M., D. A. Willits, …, A. Zlotnick. 2004. Interaction with capsid protein alters RNA structure and the pathway for in vitro assembly of cowpea chlorotic mottle virus. *J. Mol. Biol.* 335:455–464.

35. Johnson, K. N., L. Tang, …, L. A. Ball. 2004. Heterologous RNA encapsidated in Pariacoto virus-like particles forms a dodecahedral cage similar to genomic RNA in wild-type virions. *J. Virol.* 78:11371–11378.

36. Borodavka, A., R. Tuma, and P. G. Stockley. 2013. A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA Biol.* 10:481–489.

37. Perlmutter, J. D., C. Qiao, and M. F. Hagan. 2013. Viral genome structures are optimal for capsid assembly. *eLife.* 2:e00632.

38. Bringloe, D. H., A. P. Gultyaev, …, R. H. A. Coutts. 1998. The nucleotide sequence of satellite tobacco necrosis virus strain C and helper-assisted replication of wild-type and mutant clones of the virus. *J. Gen. Virol.* 79:1539–1546.

39. Richards, J. E., U. Desselberger, and A. M. Lever. 2013. Experimental pathways towards developing a rotavirus reverse genetics system: synthetic full length rotavirus ssRNAs are neither infectious nor translated in permissive cells. *PLoS One.* 8:e74328.

40. Muroga, Y., Y. Sano, …, S. Shimizu. 2007. Studies on the conformation of a polyelectrolyte in solution: local conformation of cucumber green mottle mosaic virus RNA compared with tobacco mosaic virus RNA. *J. Phys. Chem. B.* 111:8619–8625.

41. Fang, L. T., W. M. Gelbart, and A. Ben-Shaul. 2011. The size of RNA as an ideal branched polymer. *J. Chem. Phys.* 135:155105.

42. Yoffe, A. M., P. Prinsen, …, A. Ben-Shaul. 2008. Predicting the sizes of large RNA molecules. *Proc. Natl. Acad. Sci. USA.* 105:16153–16158.

43. Bundschuh, R., and T. Hwa. 2002. Statistical mechanics of secondary structures formed by random RNA sequences. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 65:031903.

44. Zahran, M., C. Sevim Bayrak, …, T. Schlick. 2015. RAG-3D: a search tool for RNA 3D substructures. *Nucleic Acids Res.* 43:9474–9488.

45. Kramers, H. A. 1946. The behavior of macromolecules in inhomogeneous flow. *J. Chem. Phys.* 14:415–424.

46. Rubinstein, M., and R. H. Colby. 2013. Polymer Physics. Oxford University Press, New York.

47. Fang, L. T., A. M. Yoffe, …, A. Ben-Shaul. 2011. A sequential folding model predicts length-independent secondary structure properties of long ssRNA. *J. Phys. Chem. B.* 115:3193–3199.

48. Lorenz, R., S. H. Bernhart, …, I. L. Hofacker. 2011. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26.

49. Somarowthu, S., M. Legiewicz, …, A. M. Pyle. 2015. HOTAIR forms an intricate and modular secondary structure. *Mol. Cell.* 58:353–361.

50. Novikova, I. V., S. P. Hennelly, and K. Y. Sanbonmatsu. 2012. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* 40:5034–5051.

51. Strauss, J. H., Jr., and R. L. Sinsheimer. 1963. Purification and properties of bacteriophage MS2 and of its ribonucleic acid. *J. Mol. Biol.* 7:43–54.

52. Grilley, D., A. M. Soto, and D. E. Draper. 2006. Mg^{2+}-RNA interaction free energies and their relationship to the folding of RNA tertiary structures. *Proc. Natl. Acad. Sci. USA.* 103:14003–14008.

53. Draper, D. E. 2004. A guide to ions and RNA structure. *RNA.* 10:335–343.

54. Woodson, S. A. 2010. Compact intermediates in RNA folding. *Annu. Rev. Biophys.* 39:61–77.

55. Garmann, R. F., A. Gopal, …, S. C. Harvey. 2015. Visualizing the global secondary structure of a viral RNA genome with cryo-electron microscopy. *RNA.* 21:877–886.

56. Tam, M. F., J. A. Dodd, and W. E. Hill. 1981. Physical characteristics of 16 S rRNA under reconstitution conditions. *J. Biol. Chem.* 256:6430–6434.

57. Mandiyan, V., S. J. Tumminia, …, M. Boublik. 1991. Assembly of the *Escherichia coli* 30S ribosomal subunit reveals protein-dependent folding of the 16S rRNA domains. *Proc. Natl. Acad. Sci. USA.* 88:8174–8178.

58. Borodavka, A., J. Ault, …, R. Tuma. 2015. Evidence that avian reovirus σNS is an RNA chaperone: implications for genome segment assortment. *Nucleic Acids Res.* 43:7044–7057.

59. Hammann, C., and D. M. J. Lilley. 2002. Folding and activity of the hammerhead ribozyme. *ChemBioChem.* 3:690–700.

60. Li, W., E. Manktelow, …, A. M. Lever. 2010. Genomic analysis of codon, sequence and structural conservation with selective biochemical-structure mapping reveals highly conserved and dynamic structures in rotavirus RNAs with potential *cis*-acting functions. *Nucleic Acids Res.* 38:7718–7735.

7044–7057 Nucleic Acids Research, 2015, Vol. 43, No. 14 Published online 24 June 2015
doi: 10.1093/nar/gkv639

Evidence that avian reovirus σNS is an RNA chaperone: implications for genome segment assortment

Alexander Borodavka*, James Ault, Peter G. Stockley and Roman Tuma*

School of Molecular and Cellular Biology & Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK

Received February 06, 2015; Revised June 06, 2015; Accepted June 09, 2015

ABSTRACT

Reoviruses are important human, animal and plant pathogens having 10–12 segments of double-stranded genomic RNA. The mechanisms controlling the assortment and packaging of genomic segments in these viruses, remain poorly understood. RNA–protein and RNA–RNA interactions between viral genomic segment precursors have been implicated in the process. While non-structural viral RNA-binding proteins, such as avian reovirus σNS, are essential for virus replication, the mechanism by which they assist packaging is unclear. Here we demonstrate that σNS assembles into stable elongated hexamers *in vitro*, which bind single-stranded nucleic acids with high affinity, but little sequence specificity. Using ensemble and single molecule fluorescence spectroscopy, we show that σNS also binds to a partially double-stranded RNA, resulting in gradual helix unwinding. The hexamer can bind multiple RNA molecules and exhibits strand-annealing activity, thus mediating conversion of metastable, intramolecular stem-loops into more stable heteroduplexes. We demonstrate that the ARV σNS acts as an RNA chaperone facilitating specific RNA–RNA interactions between genomic precursors during segment assortment and packaging.

INTRODUCTION

Avian reoviruses (ARVs) are commercially significant pathogens that cause considerable losses in the poultry industry worldwide (1). ARVs belong to the genus *Orthoreovirus* of the *Reoviridae* family, which encompasses numerous human and animal pathogens, including rotaviruses, bluetongue virus and coltiviruses. Reoviruses possess double-stranded (ds) RNA genomes, partitioned into 10–12 genomic segments (2). The ARV genome (23.5 kb) is composed of 10 dsRNA segments, encoding 8 structural and at least four nonstructural proteins (1). The genomic dsRNA segments are encased within two concentric protein shells, forming the outer capsid and the core. Apart from genomic segments the core contains multiple copies of the RNA-dependent RNA polymerase and the capping enzyme (2). Upon infection the outer capsid is proteolytically removed, resulting in the transcriptional activation of the core. The released core extrudes mRNAs into cytoplasm (3), where they are used for both translation and as genome segment precursors.

RNA replication and morphogenesis of reoviruses occurs exclusively within cytoplasmic inclusion bodies, also known as viral factories, or 'viroplasms' (4). ARV viroplasms are primarily made of the nonstructural protein μNS in association with another nonstructural protein σNS (3,5,6). The mechanism of segment assortment, by which exactly one of each of the multiple genomic precursors is selected and packaged into individual virions, is largely unknown, remaining one of the most significant questions in the biology of pathogenic dsRNA viruses. Previous studies of other members of the *Reoviridae* family suggest that multiple RNA–RNA interactions are involved in segment assortment (7–9). These interactions are believed to drive formation of a hypothetical 'assortment complex', which includes a full set of genome segment precursors, single-stranded (ss) RNAs, destined for encapsidation (10–12). In ARVs and mammalian reoviruses (MRVs) nonstructural RNA-binding proteins σNS appear to be important in this process, although their functions remain poorly understood (3,13,14).

While σNS proteins from ARVs and MRVs share very little sequence similarity, both proteins are known to form oligomers that bind ssRNA, as well as ssDNA, having very low affinity for double-stranded nucleic acids (6,13,14). Here, using a combination of various biophysical techniques, we demonstrate that the ARV σNS assembles into elongated hexamers, capable of binding strands or segments of ssRNA with nanomolar affinity. Hexamer binding causes

*To whom correspondence should be addressed. Tel: +44 113 34 32881; Fax: +44 1133 437897; Email: A.Borodavka@leeds.ac.uk
Correspondence may also be addressed to Roman Tuma. Tel: +44 113 34 33080; Email: R.Tuma@leeds.ac.uk

local destabilization (unwinding) of RNA secondary structure, resulting in formation of ribonucleoprotein complexes of variable stoichiometry. Upon helix unwinding, σNS promotes annealing of complementary strands, yielding more stable intermolecular duplexes with extended complementarities. Our results show that the ARV σNS is capable of accelerating RNA folding, thus functioning as an RNA chaperone (15), consistent with its role in genomic segment precursor selection by facilitating specific RNA–RNA interactions in viroplasms.

MATERIALS AND METHODS

Plasmid construction, protein expression and purification

Total RNA, extracted from chicken embryo fibroblast cells infected with the ARV strain 1733, was a gift from Prof. Javier Benavente (University of Santiago de Compostela). The RNA was reverse-transcribed using Superscript III Reverse Transcriptase and random hexamer oligonucleotide primers (Invitrogen). Oligonucleotide primers F_sns and R_sns (Supplementary Table S1) were used to PCR-amplify the resulting cDNA, with NdeI and XhoI restriction sites used for ligating the resulting double-digested σNS-coding DNA fragment into a linearized pET-15b vector (Novagen). The resulting pET-15b-σNS DNA construct was verified by sequencing.

pET-15b-σNS-transformed BL21(DE3)pLysS *Escherichia coli* cell cultures grown at 37°C were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) upon reaching optical density (600 nm) of 0.6–0.7, after which the protein expression was continued at 21°C overnight. Following a low-speed centrifugation harvesting, cells were resuspended and incubated for 30 min in lysis buffer (50 mM Tris–HCl, pH 8, 200 mM NaCl, 1% Tween 20, 0.2 mg/ml chicken egg lysozyme), complemented with a complete protease inhibitor cocktail tablet (Roche), and then treated with DNAseI (0.1 mg/ml) for 15 min at room temperature (RT) before cellular debris were removed by centrifugation at 20 000r pm for 30 min. The clarified cell lysate was loaded onto a pre-equilibrated Ni^{2+}-charged HisTrap FF column (GE Healthcare), washed with buffer A (1.2 M NaCl, 50 mM Tris–HCl pH 8, 40 mM imidazole and 0.5% Tween 20), and eluted using a linear gradient of buffer B (0.2 M NaCl, 50 mM Tris–HCl pH 8, 1 M imidazole). Collected protein-containing fractions were dialysed against buffer C (50 mM Tris–HCl pH 8, 1 mM EDTA, 200 mM NaCl) at 4°C, treated with human plasma thrombin (1 U/100 μg of recombinant protein) for 12 h at 22°C in order to remove His-tags. Protein samples were then loaded onto a pre-equilibrated HiTrap Q column, washed with buffer D (50 mM Tris–HCl pH 8, 1 mM EDTA, 50 mM NaCl), and the protein was eluted with buffer E (50 mM Tris–HCl pH 8, 1 M NaCl). Concentrated σNS samples were injected onto a Superdex 200 10 × 300 GL column (GE Healthcare) in 25 mM HEPES–Na, pH 7.5, 150 mM NaCl. After the final purification step, the A_{260}/A_{280} absorbance ratio of ∼0.57 suggested negligible contamination with nucleic acids.

For strand-annealing assays, rotaviral protein NSP2 was expressed and purified, as previously described in (16).

RNA and DNA oligonucleotides used for σNS binding assays

All oligonucleotides were obtained from Integrated DNA Technologies (IDT). A 120-nt long RNA sequence was designed with a fixed nucleotide composition distribution (A = 5%, C = 40%, G = 5%, T = 50%) in order to minimise any potential base-pairing. The resulting nucleotide sequence was examined using the mfold web server (17), and the RNA sequence with the lowest propensity to form intramolecular base pairs was chosen. A DNA ultramer (Supplementary Table S1) was used for transcribing a 120-nt long RNA. This transcript, as well as longer RNA transcripts (3569-nt long bacteriophage MS2 RNA) were fluorescently labelled with AlexaFluor 488 (AF488) dye (Life Technologies), as described in (18). A 30-nt and a 20-nt long RNAs, representing portions of a 120-nt long RNA, were chemically synthesized, 3′-end labelled (AF488) and PAGE-purified. A 46-nt long DNA oligonucleotide (Supplementary Table S1) was designed to form a hairpin structure and was synthesized with fluorescent labels at both the 5′ and 3′ ends (AF488 and Texas Red) (IDT).

Sedimentation velocity (SV) analytical ultracentrifugation

Sedimentation velocity data were collected from samples, loaded into 1.2-cm path length, 2-sector meniscus-matching epon centrepiece cells in an An50-Ti rotor using Optima XL-I Beckman Analytical Ultracentrifuge (Beckman Instruments, Palo Alto, CA, USA). SV analyses were conducted at 21°C with protein concentrations ranging between 5 and 25 μM and RNA concentrations of 50–60 nM. Changes in solute concentration were detected using absorbance scans at 260 nm for nucleic acids and 280 nm for proteins, as well as by interference. Each cell was scanned 200 times during the 12 h run, preceded with 1 h pre-run period in the rotor at a given speed (30 000 and 48 000 rpm, respectively). Buffer density and viscosity were calculated using Sednterp (19). Partial specific volume values for hydrated RNA were assumed 0.53–0.55 cm^3/g (20), and 0.7321 cm^3/g for σNS (calculated from its amino acid composition) with an estimated hydration level of 0.3733 g/g (19).

Radial absorbance profiles were fitted to a continuous distribution c(S) Lamm equation model using the Sedfit version 12.1b software (21). During the analysis of the sedimentation velocity profiles systematic time-invariant and radial-invariant noise components were calculated and taken into account. Models for single and multiple discrete species were used, and the location of the meniscus was treated as a floating parameter. The resulting meniscus position was then used in the c(S) distribution analysis. Maximum entropy regularization was used to calculate the size distribution within a confidence level of 0.68 of the best-fit distribution (22). Equivalent Stokes radii (R_h) were calculated using Einstein–Stokes relationship, implemented in Sedfit (21).

Sedimentation equilibrium (SE) analytical ultracentrifugation

Protein samples (6–25 μM) were examined by SE, which was attained at 8000 rpm for 24 h and 12 000 rpm for 12 h

at 20°C in an eight-hole An50 Ti rotor. Global analysis of several datasets of radial absorbance obtained at different loading concentrations and rotor speeds was performed in Sedphat (23), and SE data were fit to a single ideal species model (24) in order to estimate apparent molecular masses.

Negative staining transmission electron microscopy (TEM)

Five microliters of the σNS–RNA complex, used in SV experiments (see above), was applied onto carbon-coated grids (Agar Scientific) to absorb for 1 min, after which the excess of sample was removed by blotting with filter paper. Grids were stained with 2% (w/v) uranyl acetate, and examined using JEOL 1200EX transmission microscope operating at 80 kV at 30 000× and 40 000× magnifications.

Circular dichroism (CD) spectroscopy

Protein samples (4–32 μM in 100 mM NaF, pH 7.8) were analysed in a 0.5-mm path length cuvette using a Chirascan CD spectrometer (Applied Photophysics). CD spectra were acquired between 260 and 170 nm at 21°C. Two spectra were recorded and averaged for each concentration, followed by subtraction of the buffer spectrum. ProData Viewer software (Applied Photophysics) was used to view the recorded spectra and to assess the quality of the spectral data by examining HT voltage (absorbance). Measured ellipticities (mdeg) were expressed in terms of mean residue ellipticity (deg cm^2/dmol). CD spectra were deconvolved by applying CDSSTR algorithm (25) using SP175 reference database (26,27), implemented in CD analysis suite DICHROWEB (28).

CD spectra of RNA were acquired in the 240–320 nm spectral range using a 1 cm-long path cell, thermostated at 37°C, as described in (29). CD spectra were recorded for the MS2 phage RNA (∼3.6 kb, 0.2 mg/ml), before and after incubation with 5–20 μM σNS for 15 min. RNA secondary structure transitions upon thermal melting were also monitored at various temperatures up to 95°C.

Dynamic light scattering (DLS)

Apparent hydrodynamic radii of σNS species (10–60 μM) were measured by DLS using a PrecisionDeconvolve PDDLS/Batch platform, equipped with a PD2000 DLS detector. All measurements were performed in a buffer with an estimated viscosity, $\eta = 1.002$ mPa s, and density, $\rho = 1.00453$ g/cm^3 at 25°C. Hydrodynamic radius distributions were computed using Precision Deconvolve regularization-based software (Precision Detectors).

Non-covalent electrospray ionisation ion mobility-mass spectrometry (ESI-IMS-MS)

Mass spectra of protein samples (12–150 μM in 200 mM ammonium acetate, pH 7.8) were acquired in the positive-ion mode using a Synapt HDMS (Waters, UK), with quadrupole-orthogonal acceleration time-of-flight geometry and a built-in traveling wave ion mobility device. Protein mass spectra were also acquired using the LCT Premier (Waters, UK) specially modified for the analysis of non-covalently bound macromolecular complexes. The ESI conditions were optimized for the highest sensitivity detection of multimeric complexes in the gas phase (30). The drift times for the IMS data have been extracted using MassLynx software and the m/z spectra were plotted against the drift time using Driftscope software (Waters, UK). Calibration of the drift time cross-section was done by the combined analyses of denatured proteins (equine cytochrome *c*, β-lactoglobulin, avidin, alcohol dehydrogenase, pyruvate kinase). For each protein, the individual charge state ions were identified by m/z ratios and their measured drift times were plotted against collision cross-sections (Ω), taken from Prof. D. Clemmer's database (Indiana University Bloomington), as described previously (31,32).

Solution small angle X-ray scattering (SAXS)

Experimental SAXS datasets from solutions were collected at the X33 beam line (EMBL, DESY, Hamburg). SAXS data were acquired at 21°C using a 2D Photon counting Pilatus 1M-W pixel X-ray detector with a sample to detector distance set to 2.7 m. Protein samples (1.1 – 9.5 mg/ml in 50 mM Tris–HCl, pH 7.8, 100 mM NaCl) were examined, while serial dilutions (1–10 mg/ml) of bovine serum albumin (BSA) in the same buffer were used for calibrating I_0 values. Data were processed using ATSAS v. 2.5.1 (33,34). Background subtraction and data quality checks were performed in PRIMUS (35). Radii of gyration (R_g) were estimated using AUTORG (33). Scattering curves that were collected at several concentrations were scaled and merged using PRIMUS package. The resulting scattering curves served as input for *ab initio* low resolution shape reconstruction algorithms DAMMIN and DAMMIF (36). Multiple models with different assumed symmetries were generated (225 in total), and their ion collisional cross-sections were estimated using Leeds algorithm (37). Models with computed cross-sections within 10% (estimated experimental error) of the value measured by IMS-MS were selected (∼70% of all models generated), aligned, selected and superimposed with DAMSEL and DAMSUP and finally averaged with DAMAVER (38). The resulting model was iteratively filtered (DAMFILT) to the volume with a computed cross-sectional area, corresponding to the hexamer as measured by IMS-MS.

Fluorescence correlation spectroscopy (FCS) data collection and analysis

FCS measurements were performed on a custom-built FCS confocal setup (39), as described in (18). FCS data were analyzed by non-linear least-squares fitting with an autocorrelation function model accounting for a single diffusion component and the triplet state dynamics in Matlab (ver 7.11, MathWorks). R_h values were estimated based on measured diffusion time values for AF488 dye molecule as described before (18).

Fluorescence anisotropy (FA) and binding affinity determination

AF488-labelled 120-nt, 30-nt and 20-nt long RNAs were used for estimating binding affinities of σNS to ssRNAs. All measurements were performed at 21°C using a Fluorolog spectrofluorimeter (Horiba Jobin-Yvon). σNS was titrated (10 nM–10 μM final concentrations) into 1 nM of each of the RNA in 10 mM HEPES–NaOH, pH 7.5, 150 mM NaCl, allowing equilibration for 30 min prior FA data collection. Normalized anisotropy was plotted as a function of protein concentration and fitted to a single-site binding model using OriginPro 9.0 software.

Ensemble and single-molecule FRET helix-unwinding assays

Fluorescently labeled 46-nt long probe with a FRET donor (AF488) and acceptor (Texas Red) pair at the 5′ and 3′ ends, respectively, was designed (Supplementary Table S1). Since σNS was shown to bind ssDNA, as well as ssRNA (14), due to the high cost and low yield, a dual-labeled, 46-nt long ssDNA was synthesized and HPLC purified instead of ssRNA. Ensemble FRET efficiencies were measured for 1 nM probe in the absence of σNS, upon thermal unfolding at 80°C, and in presence of the increasing amounts of σNS (50 nM–10 μM) in the assay buffer (10 mM HEPES–NaOH, pH 7.5, 150 mM NaCl). Fluorescence intensities at 515 nm (donor) and 615 nm (acceptor) were recorded upon excitation with 488 nm, and FRET efficiencies were estimated as described in (40). For single-molecule FRET measurements a custom-made inverted confocal microscope setup equipped with Alternating Laser Excitation (ALEX) was used (41), as previously described in (42). Dual-labeled 46-nt long probe (10 pM) was measured in assay buffer with variable amounts of σNS (25 nM–20 μM) at 37°C. During data acquisition laser alternation period was set to 100 μs with 488 nm laser intensity set to 90 μW and the 594 nm laser intensity set to 60 μW. Burst selection was performed with a 10 kHz threshold, with the burst photon number set to 200 photons. Uncorrected ratiometric observables E and S were calculated after burst identification as described in (42).

RNA strand-annealing assay

Two 33-nt long oligonucleotides, 33A and 33B, with self-complementary stem-forming regions (Supplementary Table S1) were designed with the aid of mfold (17) and used in the assay. Each 33A and 33B (100 μM) were separately heat-annealed for 5 min at 85°C in 100 mM NaCl, 10 mM $MgCl_2$, 50 mM Tris–HCl, pH 7.9, slowly cooled and diluted in assay buffer (25 mM HEPES–Na, pH 7.5, 0.05% Tween 20, 150 mM NaCl, 2 mM DTT) to 5 ng/μl. Reactions were set up with equimolar amounts of 33A and 33B (200 nM total concentration) and variable amounts (1–20 μM) of the ARV σNS or acetylated BSA (negative control). Reactions were allowed to proceed at 37°C for 5–30 min before they were stopped by adding gel-loading buffer, after which they were kept on ice before resolving products on a 1× TBE 15% PAGE. Gels were stained with SYBR Gold (Invitrogen) and visualized by a fluorescence scanner with a 488 nm laser excitation.

Additional strand-annealing assays were performed with viral RNA fragments, encompassing the last 91 nucleotides of the 3′ end (1552–1643 nt) of segment s1 precursor and a similarly sized RNA fragment (422–513 nt) of segment s4 precursor (S1133 strain of ARV). These regions were identified using RactIP tool (43) with the minimum folding energies and structures of the respective RNA sequences computed using mfold (17). DNA Ultramers (IDT), incorporating T7 promoter sequences upstream of either s1 or s4 sequence, were designed, commercially synthesised and used for *in vitro* transcription using T7 polymerase. The resulting RNA products were purified, and each RNA strand was heat-annealed (as described above) prior strand-annealing reactions. For strand-annealing reactions, s1 and s4 RNAs (1.5 μM each) were incubated with 50–70 μM σNS for 5–15 min at 37°C. Similarly, strand-annealing reactions were set up with the rotaviral nonstructural protein NSP2. All reactions were stopped by adding 40 μg of proteinase K (NEB), and incubating for another 15 min. RNA samples were resolved on a native 12% 1× TBE PAGE and visualised by staining with SYBR Gold (Invitrogen).

RESULTS

ARV σNS forms stable elongated hexamers

A full-length recombinant ARV σNS was expressed and purified as described in 'Materials and Methods' section, and a homogeneous, intact protein sample was obtained by size-exclusion chromatography (Figure 1 A and Supplementary Figure S1). Its early elution from the size-exclusion column suggests that σNS oligomerizes, whilst circular dichroism spectroscopy confirms that the protein is folded, and composed of ∼35% β-strands and ∼29% α-helices (Supplementary Figure S1). Sedimentation equilibrium analysis of the protein at micromolar concentrations yields a mass of 259.3 ± 12.0 kDa, suggesting that under these conditions it is a hexamer (Supplementary Figure S2). Sedimentation velocity analysis (Figure 1 B), combined with the apparent hydrodynamic radius (R_h) of 5.0 ± 0.4 nm (estimated by dynamic light scattering, Supplementary Figure S3 A), reveals 11.3S oligomers with a corresponding mass of 237.6 ± 23.0 kDa. Taken together, these analyses demonstrate that at low micromolar concentrations the ARV σNS assembles into stable 11.3S hexamers.

Previously, smaller oligomeric species of the ARV σNS were reported (10). We therefore hypothesized that these smaller oligomers could be assembly intermediates of the hexamer. In order to characterize the assembly of hexamers, we employed native electrospray ionization ion-mobility spectrometry mass-spectrometry (ESI-IMS-MS). As expected, σNS hexamers are observed (Figure 1 C), while the mass-spectrum also contains monomers, dimers and tetramers. These smaller oligomers likely result from dissociation during electrospray ionization, since neither velocity nor equilibrium sedimentation data indicate mass heterogeneity. The prevalence of oligomers with even number of subunits suggests that a dimer is the building block of the hexamer. Since the lower molecular weight species are only detected by the ESI-MS, it is likely that assembly of σNS dimers into hexamers is driven by hydrophobic interactions, which are significantly weaker in the gas phase.

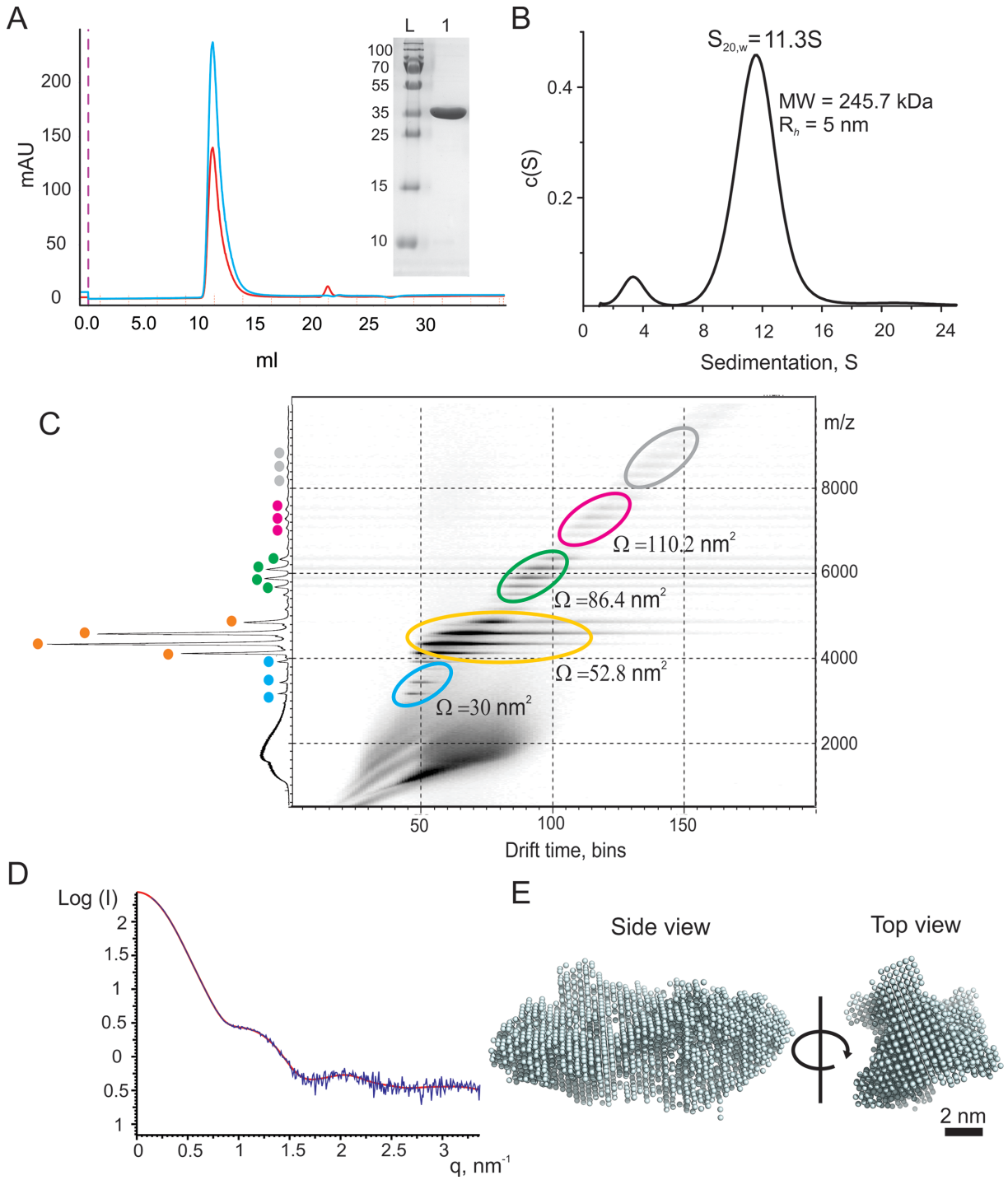


Figure 1. Quaternary structure of the ARV σNS. (**A**) Size exclusion chromatography (Superdex 200 10 × 300GL column) of σNS after His-tag removal and anion-exchange purification. 260 and 280 nm absorbance is shown in red and blue, respectively. Inset: SDS-PAGE of the purified protein (L – molecular weight marker with masses in Da shown on the right). (**B**) Sedimentation velocity analysis of the ARV σNS sedimenting at 11.3 ± 0.3S at 10 μM protein concentration. (**C**) ESI-IMS-MS Driftscope plot of the ARV σNS. Protein monomers and oligomers are highlighted in differently colored ovals (blue – monomers, orange – dimers, green – tetramers, magenta – hexamers, and gray – low intensity higher order oligomers) are separated by their drift times and m/z values. The cumulative m/z scan spectrum is shown on the right (y-axis) and the detected charge state ions are labeled according to the color scheme used in the Driftscope plot. Averaged collisional cross-sectional areas (Ω) values (in nm^2) for the lowest detected charge states are shown next to the corresponding protein oligomers. The m/z, molecular masses, drift time and cross-sectional area values are summarised in Supplementary Table S2. (**D**) Merged and buffer corrected experimental SAXS curves (blue), taken for a range of σNS concentrations (30–150 μM), with overlaid reconstructed fit (red). (**E**) A bead model of the ARV σNS hexamer reconstructed from the SAXS data. Multiple models compatible with SAXS data were produced and filtered according to their estimated collisional cross-sectional areas (Ω), as described in 'Materials and Methods' section.

Given the monodisperse nature of the hexamer, we characterized its shape using small-angle X-ray scattering (SAXS) (Figure 1D). The estimated mass and size of a σNS hexamer is in agreement with its hydrodynamic properties, with a radius of gyration (R_g) of 5.3 ± 0.3 nm, while the R_g/R_h ratio > 1 (44) suggests that its shape is elongated. We then performed *ab initio* shape reconstruction using the SAXS data (36), without symmetry (P1, Supplementary Figure S3B), or with P32 symmetry imposed, treating a hexamer as a trimer of dimers. Multiple reconstructed best-fit SAXS models were ranked using their estimated IMS-MS cross-section values (Supplementary Figure S4 and Supporting Table S2), selected, superimposed and averaged (see 'Materials and Methods' section for details). The resulting prolate bead model is shown in Figure 1E, consistent with a shape of a hexamer, which assembles as a trimer of dimers at low micromolar protein concentrations. A similar shape is obtained by reconstruction without any imposed symmetry (Supplementary Figure S3 B).

σNS hexamers bind ssRNAs with high affinity, forming ribonucleoproteins of variable stoichiometry

ARV σNS has been previously reported to bind ssRNAs, as well as single-stranded DNA *in vitro* (14). In order to characterize σNS binding to ssRNA quantitatively, we designed several defined ssRNA substrates of variable lengths (20–120 nt), with minimal propensity to base pair (Supplementary Table S1), dye-labeled at the 5′-end for use in fluorescence anisotropy (FA) measurements, as described in 'Materials and Methods' section. These substrates bind σNS with high affinity, with the estimated apparent dissociation constant (K_d) of 26.5 ± 4.2 nM for the 20-nt RNA (Figure 2A). Similar affinities are observed for equivalent 20-nt long ssDNAs (Supplementary Figure S5), suggesting the protein does not have any preference for ssRNA over ssDNA. The 120-nt ssRNA also binds σNS with similar affinities to those obtained for the 20-mer (Figure 2A).

Due to the oligomeric nature of the ARV σNS, it is possible that a single hexamer could bind several RNA strands simultaneously. At molar excess of the 20-mer RNA (Supplementary Table S3), native ESI-MS reveals 1:1 and 2:1 RNA:hexamer complexes, thus directly demonstrating binding of up to two RNAs per single hexamer. Other smaller protein oligomers were also detected, but only the hexamer bound RNA, suggesting that it is the functional RNA-binding oligomer. Since the stability of higher order species may be compromised in the gas phase, we also examined whether the 20-mer is capable of binding more than a single hexamer using fluorescence correlation spectroscopy (FCS). At saturating protein concentrations, the R_h of the 20-mer does not increase by more than the estimated hydrodynamic size of a single hexamer (Supplementary Figure S6), consistent with binding to a single hexamer.

Given that σNS hexamer binds ssRNAs as short as 20 nucleotides, we used sedimentation analysis to examine whether a longer RNA, the 120-mer, can bind multiple hexamers. SV of the protein-free 120-mer (60 nM) reveals a single 5.2 S species for the 41-kDa RNA (Figure 2B, red). In order to be able to detect the protein component in SV experiments, we used a molar excess of σNS (6 μM) over the 120-mer, which results in formation of 13 S and 18 S RNA-containing complexes, in addition to the 11.3S RNA-free hexamer (Figure 2B, blue). We then used FCS to characterize the apparent hydrodynamic radii (R_h) of these complexes. The R_h of the 120-mer (∼3.6 nm), incubated with 0.5–1 μM σNS, increases to 4.8 ± 0.8 nm (Supplementary Figure S7), corresponding to the size of a single hexamer, independently measured by DLS (see above). At higher protein concentrations (5–10 μM), the R_h further increases to 5.9 ± 1.0 nm, consistent with binding of a second hexamer to the RNA. Thus combining R_h and S values of ribonucleoproteins, the estimated masses of 13S and 18S species correspond only to complexes of 1:1 and 2:1 hexamer:RNA stoichiometry, respectively. A 2:2 stoichiometry is not compatible with the sedimentation data (Table 1). Thus longer ssRNAs can bind multiple hexamers of σNS simultaneously.

σNS hexamer binding results in RNA secondary structure destabilization

We then examined how nucleic acid secondary structure affects its affinity for σNS. Since σNS exhibits similar affinities for both ssRNA and ssDNA, we designed a 46-nt long dual end-labeled ssDNA probe (see 'Materials and Methods' section), comprising a 30-nt stem-loop and a 16-nt long single-stranded region (Figure 3A). The choice of ssDNA over ssRNA substrate was motivated by the technical difficulties in synthesis of the equivalent dual-labeled RNA with extensive secondary structure, whilst the apparent affinities for ssRNA and ssDNA are similar (Supplementary Figure S5). The apparent affinity of σNS for the partially double-stranded 46-nt long substrate was significantly lower (K_d = 333 ± 10 nM, Figure 3B, red), than for the unstructured 20-nt long ssDNA. Analysis of the ARV σNS binding to the partially double-stranded 46-nt probe yields Hill coefficient of 3.2 ± 0.23, suggesting binding cooperativity, in contrast to the behavior with fully single-stranded substrates.

We then examined whether σNS binding to the stem–loop results in destabilization of its base-paired region by monitoring changes in FRET upon protein binding. The probe alone exhibits high FRET efficiency, consistent with its predicted secondary structure (Figure 3A). When titrating the probe with molar excess of σNS, a gradual decrease in FRET efficiency is observed, that plateaus at ∼5 μM σNS (Figure 3B, black), suggesting helix unwinding. Fitting the FRET data to a single site binding model yields a K_d of 1.1 ± 0.22 μM, i.e. significantly higher than that measured by fluorescence anisotropy. This result agrees with a model in which binding of a hexamer captures transiently opened single-stranded regions, leading to gradual helix unwinding at higher protein concentrations. In this model binding effectively competes with the duplex stability, leading to the higher apparent, FRET-derived K_d for unwound species.

Since the binding model described above predicts the presence of multiple species with different degrees of unwinding, we employed a single-molecule FRET detection approach using the alternating laser excitation (ALEX) method (41) in order to obtain the distribution of the unwound species. A high FRET population (E = 0.89) is ob-

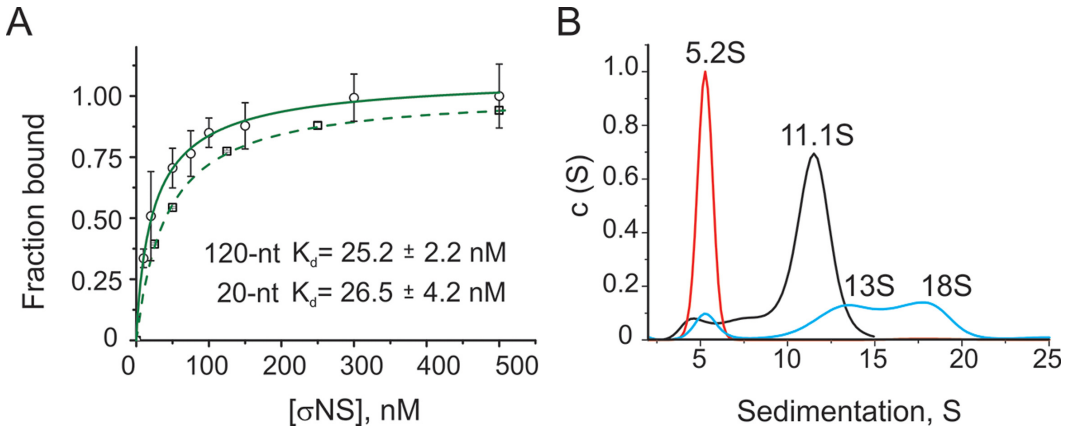


Figure 2. ARV σNS binding to unstructured ssRNAs. (**A**) The apparent affinities (K_d) of σNS for unstructured ssRNAs, measured by fluorescence anisotropy (FA). Normalized FA have been fitted to a one-site binding model for a 120-nt long RNA (solid line) and a 20-nt long RNA (dashed line), yielding K_d values of 25 ± 2.2 and 26.5 ± 4.2 nM for a 120-mer and a 20-mer, respectively. The scale of normalized FA data for the 20-mer substrate is offset for clarity. (**B**) σNS binding to a 120-nt long RNA, analysed by sedimentation velocity (SV). A c(S) distribution plot of the RNA alone (60 nM), sedimenting at 5.2S (red), and σNS hexamers (6 μM, black). σNS-RNA ribonucleoproteins are formed upon incubation of 60 nM RNA with 6 μM σNS (shown in blue), resolving into 13S and 18S species.

Table 1. Hydrodynamic characterization and molecular masses of a 120-nt long RNA:σNS complexes of variable stoichiometry

Sedimenting species, S	[σNS]:[RNA] stoichiometry	Expected MW (kDa)	Calculated from experimental *S* values MW (kDa)	δMW (%)	Partial specific volume (ml/g)
13	1:1	286	296	3.5	0.706
18	2:1	531	524	1.4	0.72
18	2:2	572	433	24.3	0.687

Partial specific volumes of [σNS]:[RNA] complexes were calculated based on partial specific volumes for the RNA alone (0.55 ml/g) and σNS alone (0.7321 ml/g), estimated as described in 'Materials and Methods' section. For 13S and 18S species partial specific volumes were calculated using the ratio of masses of RNA:protein for a given [σNS]:[RNA] stoichiometry (69). Combining the *S* values and the R_h values, measured by FCS, the apparent molecular masses (MW) of 13S and 18S species were calculated, as described in (70). The calculated MW of 13S and 18S species closely match those expected for 1:1 and 2:1 σNS:RNA complexes (δMW < 5%).

served for the probe alone (Figure 3C, blue), while unfolding by formamide resulted in complete conversion of high FRET molecules into a low FRET ($E = 0.15$) population (Figure 3C, red). Increasing the σNS concentration leads to a gradual shift from a high FRET population towards lower FRET values, with a significant broadening of the distribution at 1 μM protein concentration (Figure 3C). Further addition of σNS results in unwinding of more molecules (Figure 3D, and Supplementary Figures S9 and S10). The number of discernible populations does not change upon further increase in σNS (up to 20 μM), consistent with the presence of multiple folded and partially unfolded molecules at equilibrium.

Long ssRNAs expand upon σNS binding

Having examined σNS binding to short single-stranded nucleic acids with defined secondary structures, we then characterized its binding to longer ssRNAs. Given that multiple hexamers can bind ssRNAs ≥ 120 nts, it is possible that σNS binding results in RNA condensation, thus facilitating viral genome packaging (45,46). We used FCS to examine conformational behavior of the well-characterized 3.6 kb long MS2 phage genomic ssRNA (18,47,48). The apparent hydrodynamic radius of the RNA (1 nM, Figure 4A, red) does not change significantly upon addition of up to 125 nM σNS (Figure 4A, blue), while further increase in [σNS] up to 250 nM results in a marked increase in the R_h of RNA (Figure 4A, magenta). The apparent K_d for binding to long ssRNA is higher (~250 nM), than for unstructured shorter ssRNA, consistent with the high content of secondary structure in the 3.6 kb RNA (47–49). Given the high affinity of σNS for unstructured ssRNA (~25 nM), this result implies that despite the initial binding of σNS at low nanomolar concentrations, the examined ssRNA neither aggregated nor underwent significant conformational changes. However, further addition of σNS (up to 500 nM) results in an increase of the hydrodynamic size of the RNA, which reaches saturation above ~500 nM σNS (Figure 4A). These results suggest that the observed increase in R_h reflects expansion of the RNA and formation of large ribonucleoproteins ($R_h \sim 18$ nm), rather than aggregation of multiple ssRNA molecules. In order to characterize the shape and dimensions of these ribonucleoproteins, we performed SV analysis of long ssRNAs (60 nM), incubated with 6 μM σNS. Multiple spherical 50–90 S complexes, presumably of variable RNA:σNS stoichiometry, are formed (Figure 4 B), consistent with the appearance of negative stain EM samples (Figure 4B, inset).

Taking into account the helix-unwinding activity of the ARV σNS, we hypothesized that the observed expansion of the long ssRNA is likely due to its conformational re-

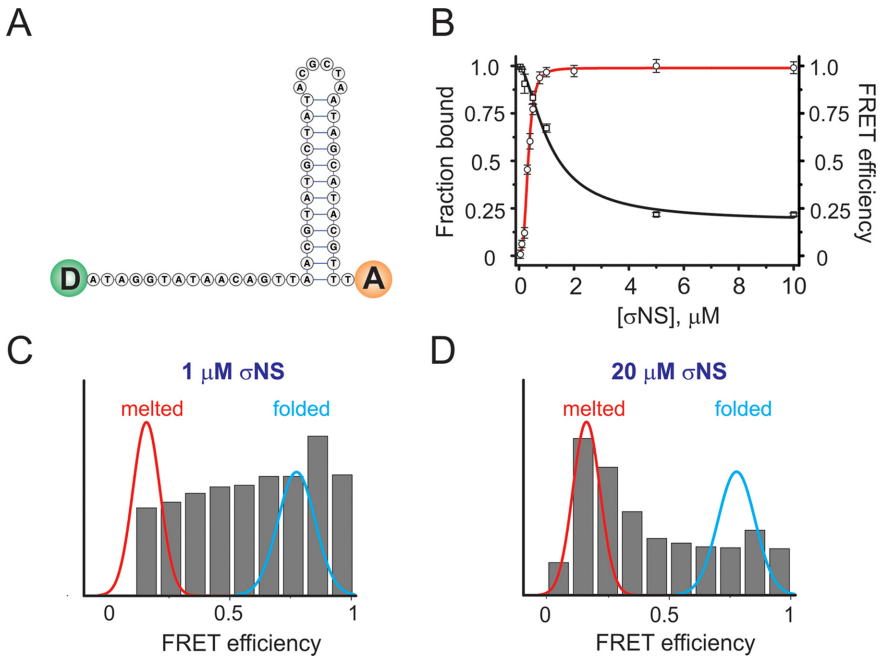


Figure 3. σNS binding to partially double-stranded nucleic acids results in helix-unwinding. (**A**) A minimum folding energy structure of a 46-nt long stem-loop used for helix unwinding assays. FRET donor and acceptor pair is schematically shown as green and orange circles at the 5′ and 3′ ends, respectively. (**B**) Binding of σNS to the stem-loop shown in (A), monitored by FA (red binding curve fit), with K_d of 333 ± 10 nM. The second fit (in black) shows changes in ensemble FRET upon σNS binding to the stem-loop due to helix unwinding (apparent K_d = 1.1 ± 0.2 μM). (**C**) Multiple unwound stem–loop intermediates, revealed by the wide distribution of the smFRET population histogram of the stem-loop probe (10 pM), incubated with 1 μM σNS. Gaussian fits of data for formamide-melted (red, low FRET) and heat-annealed (folded, cyan, high FRET) probes are overlaid for reference. (**D**) smFRET population histogram of the 46-nt stem-loop (as shown in C), incubated with 20 μM σNS.

arrangement, caused by destabilization of its secondary structure. We therefore examined RNA secondary structure changes upon σNS binding using circular dichroism (CD) spectroscopy. A positive peak at 263 nm was observed for RNA alone (Figure 4C, black), consistent with the largely base-paired RNA in an A-duplex conformation (48). Thermal unfolding of the RNA results in a shift of the 263 nm peak to 271 nm (Figure 4C, red), as expected for an RNA in a single-stranded form (29). Similar changes occur upon incubation of the RNA with 15 μM σNS (Figure 4C, dashed blue line), suggesting that binding of the ARV σNS induces melting of local secondary structures in large folded ssRNAs, resulting in their expansion.

σNS facilitates RNA annealing upon helix unwinding

Since helix-unwinding activity is often associated with a strand exchange between preformed helices (15,50,51), we then examined the strand-annealing activity of the ARV σNS using two partially complementary RNA oligonucleotides. Two self-complementary stem-loops with 11 bp-long helical stems and 11 nt-long loops were designed (Figure 5A). Incubation of both strands (100 nM each) with 1 μM σNS at 37°C for 5 min results in a small amount of double-stranded (DS) product, while a large proportion of the single-stranded form remains intact (Figure 5B, lane 4). A significant increase in the amount of dsRNA is observed when using higher molar excess of σNS (5 μM or more, Figure 5B, lanes 5–7). These results suggest that strand annealing requires more than a stoichiometric amount of σNS to unwind preformed helices. Since σNS has very low affinity for dsRNA, any RNA-bound hexamers are expected to dissociate from the newly formed duplexes upon strand annealing, thus being able to bind new ssRNAs. However, prolonged incubation with σNS does not yield more DS product (Figure 5B, lane 4 cf. lane 8). As mentioned earlier, both anisotropy and FCS measurements of the 20-mer ssRNA indicate that σNS hexamers do not associate into higher order species, while native mass spectrometry suggests multiple ssRNAs can bind each hexamer. Collectively, these results suggest that strand-annealing occurs between two complementary strands bound to a single hexamer, while large ribonucleoproteins observed on a native PAGE (Figure 5B, black rectangle at the top of the gel) indicate that a fraction of the ssRNA stays bound to a hexamer.

Having established that σNS promotes strand-annealing between short complementary RNA stem-loops, we then examined whether it exerts similar effects on longer viral RNAs, e.g. ARV segment precursors. In ARVs, segment assortment and packaging signals are believed to involve terminal untranslated regions (UTRs), extending into the protein-coding regions of each segment (52,53). We therefore examined the 3′-terminal 200-nt long sequence of the ARV segment s1 to identify a possible complementary sequence in other genomic segments. A section of segment s4, capable of forming stable (>10 bp long) intersegmental helix with the s1 RNA was identified, as described in 'Materials and Methods' section. The resulting 91 nt long RNAs, representing partially complementary fragments of the ARV genomic segment precursors s1 (1552–

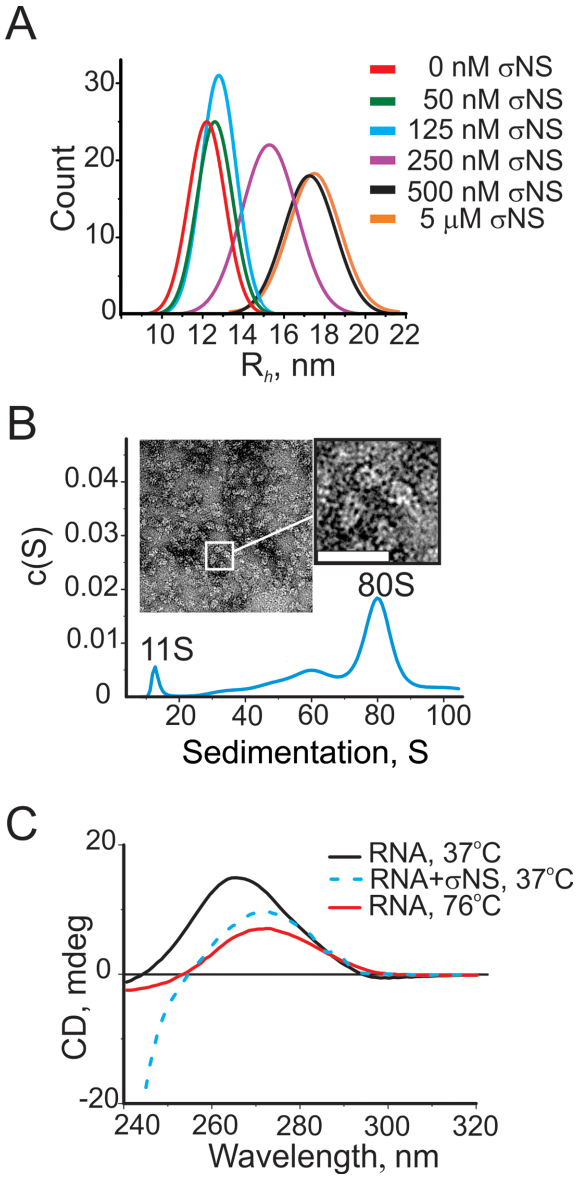


Figure 4. Assembly of σNS ribonucleoproteins with long ssRNAs. (**A**) Size (R_h) distributions of a 3.6 kb-long ssRNA alone (1 nM, red), and upon addition of σNS, measured by FCS. (**B**) Sedimentation velocity of the long ssRNA (25 nM), incubated with σNS (10 μM). A large 80S complex with fewer smaller species is formed. Inset – negative stain EM micrograph of the σNS, bound to long ssRNAs, revealing multiple spherical ribonucleoproteins. Bar = 25 nm. (**C**) Circular dichroism (CD) spectra of the long ssRNA (as in panels A and B), reveal RNA secondary structure destabilization upon incubation with the ARV σNS. CD spectra for folded RNA alone at 37°C (15 nM, black) and thermally unfolded RNA at 76°C (red) are shown along with the CD spectrum of the RNA, incubated with σNS for 5 min (15 μM, blue dashed line). Note the significantly lower intensity of the 263 nm peak in the latter, characteristic for an A-form double-stranded helix, with its shift toward 271 nm, typical for single-stranded RNA. Below 240 nm, the CD signal from RNA is mainly dominated by the contribution of the protein.

1643 nt) and s4 (422–513 nt), respectively (Figure 5 C), were used in strand-annealing assays. Similarly to shorter complementary RNA stem–loops, incubation of s1 and s4 RNA fragments in the presence of saturating amounts of σNS resulted in strand annealing, albeit with significantly lower efficiency (Figure 5 D inset, lane 4, cf. 5 B, lanes 5–7). No detectable annealed products were observed in reactions without σNS (Figure 5 D, lane 2), or when only one of the two RNA strands was incubated with σNS (Figure 5D, lane 3). Interestingly, when the ARV σNS was substituted with its functional rotaviral homologue NSP2, the annealed RNA products were also formed (Figure 5D, lane 5). These results suggest that both proteins may perform similar functions in these related viruses, and that the specificity is encoded by RNA–RNA interactions, while the protein acts in sequence-independent fashion.

DISCUSSION

Comparison with other non-structural RNA binding proteins within *Reoviridae* family

Non-structural ssRNA-binding protein σNS is one of the major components of viroplasms in avian reovirus-infected cells. It has been suggested that the protein is involved in the viral genome packaging (1,3). Although various roles of the ARV σNS have been proposed (14,54), including RNA-polymerase activity (55) and segment assortment (54), its mode of interaction with viral RNAs has remained poorly understood. In order to gain insight into the mechanistic details of the σNS–RNA interactions, we first determined the quaternary organization of the functional oligomeric form of the ARV σNS. Previously, the oligomeric nature of the recombinant ARV σNS was characterized by sucrose density fractionation (14). Here, using a combination of multiple techniques, we demonstrate that the protein readily assembles from σNS dimers into hexamers at low micromolar concentrations. Native ESI-MS data suggest that assembly of functional hexamers is likely to be driven by hydrophobic interactions between multiple protein dimers and that only hexamers bind RNA. Given the high protein concentration in viroplasms during late infection (4), the concentration of σNS is expected to be above low micromolar, suggesting the protein functions as a hexamer *in vivo*.

While both avian and mammalian reovirus σNS, as well as other viroplasm-forming non-structural proteins from *Reoviridae* have many common features, these proteins share very little, if any sequence homology. Non-structural ssRNA-binding proteins of most non-turreted *Reoviridae,* including bluetongue virus (BTV), rotaviruses, and some phytoreoviruses assemble into large oligomers, having a ring-like appearance (16,56–58). Interestingly, the secondary structure composition of the ARV σNS is rather similar to that of rotavirus non-structural protein NSP2, which assembles into octamers, with an estimated 35% of β-sheet. This likely reflects functional similarities between the two proteins, since β-sheet rich surfaces are often involved in nucleic acid binding (59). Likewise, viroplasm-forming non-structural proteins of some turreted reoviruses assemble into RNA-binding octamers (60), while in mammalian reoviruses (MRV) σNS forms 7–9S species without obvious ring-like architecture (61), suggesting that the ring-like organization of this class of proteins is not unique amongst *Reoviridae.* It is unclear whether the viroplasm-building blocks of most turreted *Reoviridae,* e.g., ARV and MRV, are different in their quaternary organization from the non-structural viroplasm-forming proteins of non-turreted

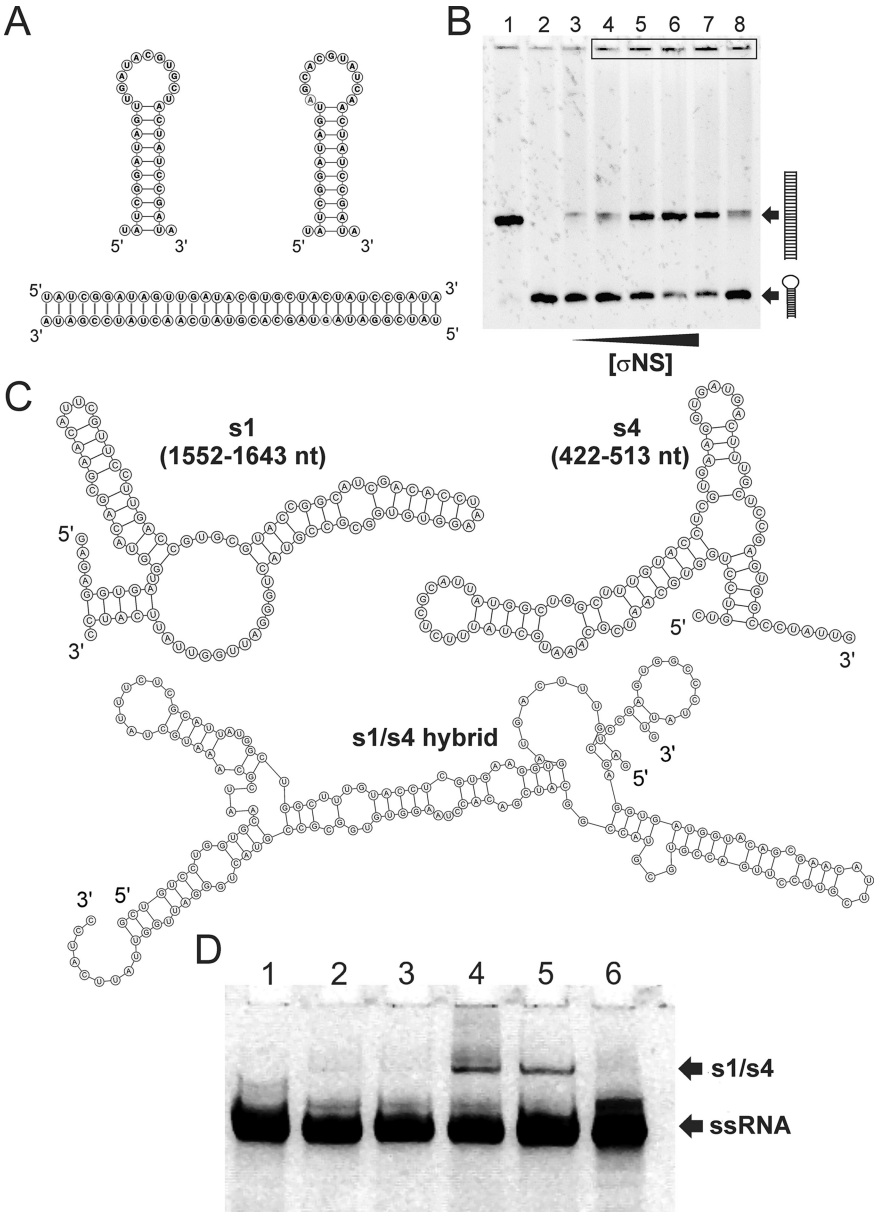


Figure 5. Helix destabilization facilitates strand-annealing activity of σNS. (**A**) Complementary ssRNA stem–loops used in the strand-annealing assay, as described in 'Materials and Methods' section (top), with the resulting re-annealed dsRNA form shown below. (**B**) RNA stem–loops (100 nM each, shown in A) were incubated with σNS at 37°C for 5 min, as described in 'Materials and Methods' section. The resulting products were resolved on a 10% (w/v) native PAGE, and visualised by staining with SYBR Gold. DsRNA and ssRNA stem–loops were loaded into lanes 1 and 2, respectively, while lane 3 contains the annealing reaction incubated with 20 μM BSA protein. Lane 4 – 1 μM σNS, lane 5 – 5 μM σNS, lane 6 – 10 μM σNS, lane 7 – 20 μM σNS. Note a fraction of the RNA in the presence of σNS fails to enter the gel (top, black rectangle). Prolonged incubation with σNS up to 30 min does not yield higher amounts of the double-stranded product (cf. lanes 4 and 8; 5 min and 30 min, respectively). (**C**) MFE structures of partially complementary regions of the ARV genomic precursors of segments s1 and s4 (Top). Both RNA structures are characterised by relatively high stabilities (dG > –20 kcal/mol) preventing spontaneous hybridisation *in vitro.* Below – the MFE structure of annealed s1 and s4 strands ('s1/s4 hybrid') (D) ARV σNS facilitates strand-annealing of viral RNA segment precursor sequences *in vitro.* 91 nt long RNAs (shown in C) were incubated alone, or in presence of the ARV σNS at 37°C for 15 min. Strand-annealing reactions were stopped and RNA products were analyzed on a native PAGE and visualized, as described in 'Materials and Methods' section. Lane 1 – s4 RNA, lane 2 – s1 + s4, incubated with 20 μM BSA protein, lane 3 – s4 + σNS, lane 4 – s1 + s4, incubated with σNS, lane 5 – s1 + s4, incubated with the rotaviral non-structural protein NSP2, lane 6 – s1 RNA. Positions of S1/S4 hybrids and free ssRNAs (S1 or S4) are indicated by arrows.

viruses. However, given a small fraction of octameric σNS species, observed in native ESI-MS spectrum at high protein concentration, it is possible that different functional oligomeric states of these proteins may be important for protein-RNA and protein-protein interactions during formation of viroplasms and subsequent virus assembly.

Both mammalian and avian reovirus σNS readily bind ssRNA, forming large ribonucleoproteins of variable stoichiometries (13,14,62). Interestingly, RNAse A treatment of large ribonucleoprotein complexes formed by MRV σNS releases 13S–19S particles (61), with similar sedimentation properties to the 13S–18S ribonucleoproteins observed when the ARV σNS hexamers were bound to 120-mer RNAs. Although MRV σNS has been shown to bind RNA with positive cooperativity and some preference for ssRNA over ssDNA (13), our results suggest ARV σNS binds RNAs lacking secondary structure without apparent cooperativity, with similar affinities for both ssRNA and ssDNA.

Previously, helix-unwinding activity has been reported for a number of non-structural proteins from *Reoviridae* family, including rotavirus NSP2 and MRV σNS (13,63). However, no strand-annealing activity has been demonstrated for these proteins so far. Both reovirus σNS and rotavirus NSP2 appear to be non-specific ssRNA-binding proteins, while BTV NS2 has been reported to preferentially bind BTV segment precursors (64). Detailed structural characterisation of both σNS and NS2 is needed to see whether these functionally related proteins play similar roles in RNA segment assortment and packaging.

Biological implications of the RNA chaperone-like activity

Our results suggest that σNS hexamer appears to be the functional RNA-binding species. Previously, σNS has been proposed to act as a condensing agent for long viral ssRNAs in viroplasms (1). Here, we provide a direct experimental evidence that its effect on long ssRNA contrasts with charge neutralization-driven RNA compaction, or sequence-specific ssRNA collapse, previously demonstrated for some icosahedral ssRNA viruses (18). These results suggest that although the ARV σNS is capable of recruiting ssRNAs into viroplasms with high nanomolar affinity, it does not condense it, but rather causes the expansion of bound ssRNAs. This expansion reflects its helix-unwinding activity, demonstrated for both long ssRNA and ssDNA substrates.

Since hexameric σNS does not exhibit ATPase activity and ring-like shape of typical hexameric helicases (65), the observed helix-unwinding activity is likely driven by its binding affinity. The ARV σNS reportedly does not bind nucleic acid duplexes, therefore helix destabilization must involve binding of hexamers to stretches of single-stranded RNA, resulting in further recruitment of neighbouring, spontaneously melted single-stranded regions. This model is compatible with the observed broad population of partially unfolded substrates that are formed during σNS binding to the hairpin substrate.

Given its non-specific ssRNA-binding and ATP-independent helix-unwinding activity, ARV σNS exhibits typical characteristics of an RNA chaperone (15). While it appears to be indispensable for viral replication, it is not present inside virions. Therefore, σNS must dissociate from the pre-assorted RNAs prior or during encapsidation. Likewise, RNA chaperones do not require ATP for their helix-unwinding activity, and once the correct RNA structure is folded, their presence is no longer needed (66). Binding of RNA chaperones often relies on multiple electrostatic interactions with their targets (15). Similarly, RNA binding by σNS is mediated by multiple electrostatic contacts, since σNS-RNA complexes readily dissociate under high ionic strength conditions, while several arginine residues have been shown to be important for binding to RNA (14). While many RNA chaperones have intrinsically disordered regions that undergo folding when bound to RNA (15), CD and SAXS indicate that the bulk of the protein is well folded. However, *in silico* secondary structure and disorder sequence analysis using Phyre2 (67) predicts an intrinsically disordered region encompassing the first seven residues at the N-terminus, as well as few regions within the C-terminus (residues 329–338 and 361–367). Previously, it was shown that the deletion of the first 11 residues of the ARV σNS, as well as Arg365Leu substitution abrogates RNA binding (14). Interestingly, while in mammalian reovirus σNS the first N-terminal residues are also essential for RNA binding, they are predicted to form an amphipathic α-helix (61), contrasting with the secondary structure organisation of the ARV σNS.

The strand-annealing activity of σNS appears to result from helix unwinding, followed by strand hybridization, presumably mediated by the proximity of two complementary strands bound to a single hexamer (Figure 6). This model is further supported by native mass spectrometry and sedimentation analysis, which demonstrate that σNS can bind at least two RNAs, and that the assembly of a stable hexamer is required for RNA binding. Furthermore, using fragments of the ARV genomic precursors s1 and s4, we demonstrate that σNS can promote specific RNA–RNA interactions between two segments *in vitro*. The reaction appears to be more efficient with shorter, less stable RNA stem–loops, likely as the result of higher thermodynamic stabilities of longer viral RNA fragments with multiple and more stable RNA helices. This agrees well with previous observations that mammalian reovirus σNS fails to unwind a 17-bp long helix (13), while the rotaviral NSP2 was inefficient in strand-displacement assays with duplexes over 10 bp (63). Interestingly, both ARV σNS and its functional homologue rotaviral NSP2 mediate similar strand-annealing reactions, suggesting these non-structural proteins may play similar roles in segment assortment in reoviruses and rotaviruses despite lack of apparent sequence homology. Given that rotavirus NSP2 promotes strand-annealing between the ARV genomic sequences, it appears that the assortment specificity in these viruses is primarily controlled by specific RNA–RNA interactions, facilitated by the virus-encoded, but sequence-independent, RNA chaperones. Recent studies of BTV segment assortment and packaging strongly support this model, suggesting that protein-free RNA segment precursors can specifically interact *in vitro.* However, in infected cells this process requires expression of NS2 and formation of viroplasms prerequisite for RNA replication and packaging (9).

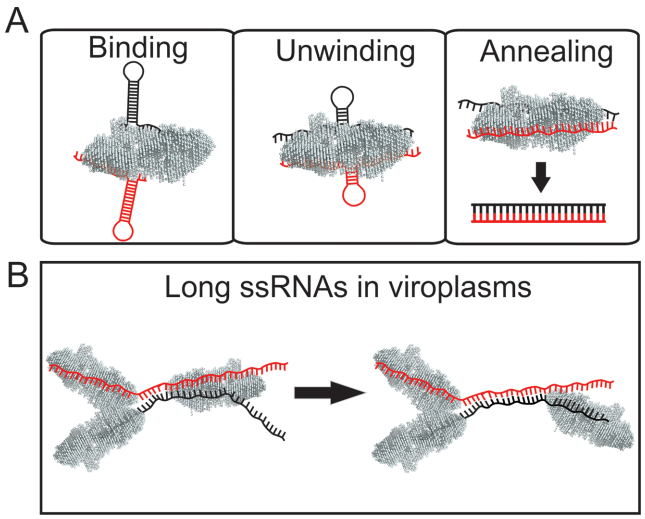


Figure 6. ARV σNS acts as an RNA chaperone. (**A**) *Binding mode*. σNS hexamers bind to ssRNA, including stem–loops with partial sequence complementarity (shown in red and black). Longer single-stranded stretches (>10 nts) are required for σNS binding, while shorter loops are too small to accommodate a σNS hexamer (see Supplementary Figure S8D). Multiple RNA–σNS complexes of variable stoichiometry are formed, depending on the length of the RNA substrate. *Unwinding mode.* At low micromolar and above concentrations, σNS binding to a partially double-stranded RNA results in its gradual helix unwinding. *Strand-annealing mode.* Binding of two complementary strands to a hexamer accelerates strand annealing, likely due to the molecular crowding effect. For short RNA substrates (20–40 nt), annealing occurs when two strands are bound to a single hexamer. After annealing the resulting dsRNA is released from the σNS hexamer. (**B**) Proposed roles of σNS in ARV-infected cells. The ARV σNS accumulates in cytoplasmic inclusion bodies ('viroplasms') in infected cells, likely by interacting with another viroplasm-forming protein μNS (1). Multiple ssRNAs (for clarity, only two strands are shown in red and black) are bound to σNS hexamers, which mediate partial unwinding of the RNAs. Two partially complementary strands bound to a hexamer can readily re-anneal, forming a more stable duplex between the two RNA strands, thus facilitating specific RNA–RNA interactions between genome segment precursors. This results in a displacement of the hexamer from the re-annealed dsRNA region and its subsequent binding to another ssRNA region elsewhere in the viroplasms.

In ARVs, strand-annealing reaction leads to a stable helix formation between the two strands, resulting in displacement of the σNS hexamer from the duplex (Figure 6 A), or its subsequent binding to a neighbouring single-stranded region on longer ssRNAs (Figure 6 B). The annealing reaction appears to be thermodynamically controlled, since prolonged incubation of complementary strands with σNS does not result in the increased yield of heteroduplexes. This implies that σNS accelerates spontaneous strand hybridization events, while the extent of the annealing reaction is controlled by equilibrium between heteroduplex and intramolecular secondary structure.

While the ARV σNS has high affinity for unstructured ssRNA, the amount of protein required for its helix-unwinding and strand-annealing activities is significantly above its K_d for ssRNA. It should be noted that early morphogenesis of ARVs occurs exclusively within the viroplasm (Figure 6B), into which σNS is selectively recruited via protein-protein interactions with the major nonstructural protein μNS (6). Thus, the effective local concentration of σNS inside viroplasms is expected to be substantially higher than in cytosol. During the early virus infection, multiple (+) ssRNAs are translated in cytoplasm, and σNS may concentrate the viral ssRNAs in viroplasms and destabilize their local secondary structures, thus assisting formation of specific RNA–RNA interactions between strands during segment assortment (Figure 6 B), consistent with helix-unwinding and strand-annealing properties of the protein.

In *Reoviridae*, RNA–RNA interactions are likely to be involved in genome assortment and packaging (10). For avian and mammalian reoviruses, RNA sequences responsible for genome packaging were shown to be located close to segmental termini, encompassing untranslated regions (UTRs), and overlapping with neighbouring ORF sequences (52,53,68). Here, we demonstrate that ARV nonstructural protein σNS displays RNA chaperone activity, which augments specific RNA–RNA interactions between different segment precursors and facilitates efficient selection and assortment of multiple genome segments, destined for encapsidation.

ACKNOWLEDGEMENTS

We thank Prof. Javier Benavente (University of Santiago de Compostella) for the generous gift of viral RNA. The authors would like to thank Dr John T. Patton (NIAID, NIH, Bethesda) and Dr Ulrich Desselberger (University of Cambridge) for valuable comments on the manuscript. We thank X33 beamline/BioSAXS support team (Hamburg, Germany) for technical assistance with SAXS data collection, in particular Dr D. Svergun (EMBL, Hamburg). We thank Dr Amit Sharma (University of Leeds) for assistance with ALEX data acquisition and processing, and Amy Barker for technical assistance with analytical ultracentrifugation experiments.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Wellcome Trust [089310/09/Z and 103068/Z/13/Z to A.B.]; University of Leeds for supporting the Single Molecule Facility and the Wellcome Trust for funding Joint Infrastructure Facility [062164 and 090932/Z/09/Z]; BBRSC and the University of Leeds SRIF4 for supporting the mass spectrometry facility at the University of Leeds. Funding for open access charge is provided by the Wellcome Trust.
Conflict of interest statement. None declared.

REFERENCES

1. Benavente,J. and Martínez-Costas,J. (2007) Avian reovirus: Structure and biology. *Virus Res.*, **123**, 105–119.
2. Chappell,J., Duncan,R., Mertens,P. and Dermody,T. (2005) In: Fauquet,CM, Mayo,MA, Maniloff,J, Desselberger,U and Ball,LA (eds). *Virus Taxonomy. Classification and Nomenclature of Viruses. Eighth Report of the International Committee on the Taxonomy of Viruses*. Elsevier Academic Press, Oxford, p. 455.
3. Benavente,J. and Martínez-Costas,J. (2006) Early Steps in Avian Reovirus Morphogenesis. In: Roy,P (ed). *Reoviruses: Entry, Assembly and Morphogenesis.* Springer, Berlin Heidelberg, Vol. **309**, pp. 67–85.

4. Patton,J.T., Silvestri,L.S., Tortorici,M.A., Vasquez-Del Carpio,R. and Taraporewala,Z.F. (2006) Rotavirus genome replication and morphogenesis: role of the viroplasm. In: Roy,P (ed). *Reoviruses: Entry, Assembly and Morphogenesis*. Springer, Berlin Heidelberg, Vol. **309**, pp. 169–187.
5. Tourís-Otero,F., Martínez-Costas,J., Vakharia,V.N. and Benavente,J. (2004) Avian reovirus nonstructural protein μNS forms viroplasm-like inclusions and recruits protein σNS to these structures. *Virology*, **319**, 94–106.
6. Tourís-Otero,F., Martínez-Costas,J., Vakharia,V. and Benavente,J. (2004) Avian reovirus morphogenesis occurs within viral factories and begins with the selective recruitment of sigmaNS and lambdaA to microNS inclusions. *J. Mol. Biol.*, **341**, 361–374.
7. Anzola,V.J., Xu,Z., Tetsuya,A. and Nuss,L.D. (1987) Segment-specific inverted repeats found adjacent to conserved terminal sequences in wound tumor virus genome and defective interfering RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 8301–8305.
8. Lourenco,S. and Roy,P. (2011) In vitro reconstitution of Bluetongue virus infectious cores. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 13746–13751.
9. Sung,P.-Y. and Roy,P. (2014) Sequential packaging of RNA genomic segments during the assembly of Bluetongue virus. *Nucleic Acids Res.*, **42**, 13824–13838.
10. Trask,S.D., Mcdonald,S.M. and Patton,J.T. (2012) Structural insights into coupling of virion assembly and rotavirus replication. *Nat. Rev. Microbiol.*, **10**, 165–177.
11. Lu,X., McDonald,S.M., Tortorici,M.A., Tao,Y.J., Vasquez-Del Carpio,R., Nibert,M.L., Patton,J.T. and Harrison,S. (2008) Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. *Structure*, **16**, 1678–1688.
12. Mcdonald,S.M. and Patton,J.T. (2011) Rotavirus VP2 core shell regions critical for viral polymerase activation. *J. Virol.*, **85**, 3095–3105.
13. Gillian,A.L., Schmechel,S.C., Livny,J., Schiff,L.A. and Nibert,M.L. (2000) Reovirus protein σNS binds in multiple copies to single-stranded RNA and shares properties with single-stranded DNA binding proteins. *J. Virol.*, **74** 5939–5948.
14. Tourís-Otero,F., Martínez-Costas,J., Vakharia,V.N. and Benavente,J. (2005) Characterization of the nucleic acid-binding activity of the avian reovirus non-structural protein σNS. *J. Gen. Virol.*, **86**, 1159–1169.
15. Rajkowitsch,L., Chen,D., Stampfl,S., Semrad,K., Waldsich,C., Mayer,O., Jantsch,M., Konrat,R., Blasi,U. and Schroeder,R. (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.*, **4**, 118–130.
16. Schuck,P., Taraporewala,Z., McPhie,P. and Patton,J.T. (2001) Rotavirus nonstructural protein NSP2 self-assembles into octamers that undergo ligand-induced conformational changes. *J. Biol. Chem.*, **276**, 9679–9687.
17. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
18. Borodavka,A., Tuma,R. and Stockley,P.G. (2012) Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 15769–15774.
19. Haye,D., Laue,T. and Philo,J. (1995) *Program Sednterp: Sedimentation Interpretation Program version 1.09*. Alliance Protein Laboratories, Thousand Oaks, CA.
20. Voss,N.R. and Gerstein,M. (2005) Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J. Mol. Biol.*, **346**, 477–492.
21. Brown,P.H. and Schuck,P. (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys. J.*, **90**, 4651–4661.
22. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.
23. Vistica,J., Dam,J., Balbo,A., Yikilmaz,E., Mariuzza,R., Rouault,T. and Schuck,P. (2004) Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal. Biochem.*, **15**, 234–256.
24. Zhao,H., Brautigam,C.A., Ghirlando,R. and Schuck,P. (2013) Overview of current methods in sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation. *Curr. Protoc. Protein Sci.*, doi:10.1002/0471140864.ps2012s71.
25. Sreerama,N. and Woody,R.W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.*, **287**, 252–260.
26. Lees,J.G., Miles,A.J., Wien,F. and Wallace,B.A. (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics*, **22**, 1955–1962.
27. Janes,R.W. (2008) Reference datasets for protein circular dichroism and synchrotron radiation circular dichroism spectroscopic analyses. In: Wallace,BA and Janes,RW (eds). *Modern Techniques in Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy*, pp. 283–202.
28. Whitmore,L. and Wallace,B. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, **89**, 392–400.
29. Sosnick,T. (2001) Characterization of tertiary folding of RNA by circular dichroism and urea. *Curr. Protoc. Nucleic Acid Chem.*, doi:10.1002/0471142700.nc1105s04.
30. Ruotolo,B.T., Benesch,J.L.P., Sandercock,A.M., Hyung,S.-J. and Robinson,C.V. (2008) Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.*, **3**, 1139–1152.
31. Smith,D.P., Knapman,T.W., Campuzano,I., Malham,R.W., Berryman,J.T., Radford,S.E. and Ashcroft,A.E. (2009) Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur. J. Mass Spectrom.*, **15**, 113–130.
32. Bush,M., Hall,Z., Giles,K., Hoyes,J., Robinson,C. and Ruotolo,B. (2010) Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.*, **82**, 9557–9565.
33. Konarev,P.V., Petoukhov,M.V., Volkov,V.V. and Svergun,D.I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **39**, 277–286.
34. Petoukhov,M.V., Franke,D., Shkumatov,A.V., Tria,G., Kikhney,A.G., Gajda,M., Gorba,C., Mertens,H.D.T., Konarev,P.V. and Svergun,D.I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **45**, 342–350.
35. Konarev,P.V., Volkov,V.V., Sokolova,A.V., Koch,M.H.J. and Svergun,D.I. (2003) PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **36**, 1277–1282.
36. Franke,D. and Svergun,D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **42**, 342–346.
37. Knapman,T.W., Morton,V.L., Stonehouse,N.J., Stockley,P.G. and Ashcroft,A.E. (2010) Determining the topology of virus assembly intermediates using ion mobility spectrometry–mass spectrometry. *Rapid Commun. Mass Spectrom.*, **24**, 3033–3042.
38. Volkov,V.V. and Svergun,D.I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **36**, 860–864.
39. Gell,C., Brockwell,D. and Smith,A. (2006) *Handbook of Single Molecule Fluorescence Spectroscopy*. Oxford University Press, pp. 24–44.
40. Lakowicz,J.R. (2006) *Principles of Fluorescence Spectroscopy*. 3rd edn. Springer, NY.
41. Kapanidis,A.N., Laurence,T.A., Lee,N.K., Margeat,E., Kong,X. and Weiss,S. (2005) Alternating-laser excitation of single molecules. *Acc. Chem. Res.*, **38**, 523–533.
42. Sharma,A., Leach,R.N., Gell,C., Zhang,N., Burrows,P.C., Shepherd,D.A., Wigneshweraraj,S., Smith,D.A., Zhang,X., Buck,M. *et al.* (2014) Domain movements of the enhancer-dependent sigma factor drive DNA delivery into the RNA polymerase active site: insights from single molecule studies. *Nucleic Acids Res.*, **42**, 5177–5190.
43. Kato,Y., Sato,K., Hamada,M., Watanabe,Y., Asai,K. and Akutsu,T. (2010) RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming. *Bioinformatics*, **26**, 460–466.
44. Burchard,W. and Patterson,G.D. (1983) Light scattering from polymers. *Adv. Polym. Sci.*, 167.
45. Dermody,T.S., Parker,J.S.L. and Sherry,B. (2013) Orthoreoviruses. In: Knipe,DM and Howley,PM (eds). *Fields Virology*. 6th edn. Lippincott Williams & Wilkins, pp. 1304–1346.

46. Borodavka,A., Tuma,R. and Stockley,P.G. (2013) A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA Biol.*, **10**, 481–489.
47. Fiers,W., Contreras,R., Duerinck,F., Haegeman,G., Iserentant,D., Merregaert,J., Min Jou,W., Molemans,F., Raeymaekers,A., Van den Berghe,A. *et al.* (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*, **260**, 500–507.
48. Thomas,G.J. Jr, Prescott,B., McDonald-Ordzie,P.E. and Hartman,K.A. (1976) Studies of virus structure by Laser-Raman spectroscopy: II. MS2 phage, MS2 capsids and MS2 RNA in aqueous solutions. *J. Mol. Biol.*, **102**, 103–124.
49. Jacobson,A. (1976) Studies on secondary structure of single-stranded RNA from bacteriophage MS2 by electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **73**, 2–6.
50. Cristofari,G. and Darlix,J.-L. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acid Res. Mol. Biol.*, **72**, 223–268.
51. Woodson,S.A. (2010) Taming free energy landscapes with RNA chaperones. *RNA Biol.*, **7**, 677–686.
52. Ni,Y. and Kemp,MC (1994) Subgenomic S1 segments are packaged by avian reovirus defective interfering particles having an S1 segment deletion. *Virus Res.*, **32**, 329–342.
53. Roner,M.R. and Roehr,J. (2006) The 3′ sequences required for incorporation of an engineered ssRNA into the Reovirus genome. *Virology*, **3**, 1–11.
54. Stamatos,N.M. and Gomatos,P.J. (1982) Binding to selected regions of reovirus mRNAs by a nonstructural reovirus protein. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 3457–3461.
55. Gomatos,P.J., Stamatos,N.M. and Sarkar,N.H. (1980) Small reovirus-specific particle with polycytidylate-dependent RNA polymerase activity. *J. Virol.*, **36**, 556–565.
56. Mumtsidu,E., Makhov,A.M., Roessle,M., Bathke,A. and Tucker,P.A. (2007) Structural features of the Bluetongue virus NS2 protein. *J. Struct. Biol.*, **160**, 157–167.
57. Taraporewala,Z.F., Chen,D. and Patton,J.T. (2001) Multimers of the bluetongue virus nonstructural protein, NS2, possess nucleotidyl phosphatase activity: Similarities between NS2 and rotavirus NSP2. *Virology*, **280**, 221–231.
58. Akita,F., Miyazaki,N., Hibino,H., Shimizu,T., Higashiura,A., Uehara-Lchiki,T., Sasaya,T., Tsukihara,T., Nakagawa,A., Iwasaki,K. *et al.* (2011) Viroplasm matrix protein Pns9 from rice gall dwarf virus forms an octameric cylindrical structure. *J. Gen. Virol.*, **92**, 2214–2221.
59. Horvath,M.P. (2008) Single-stranded Nucleic Acid (SSNA)-binding Proteins. In: Rice,P and Correll,C (eds). *Protein-nucleic Acid Interactions: Structural Biology*. RSC Biomolecular Sciences, Cambridge, pp. 91–121.
60. Wu,J., Li,J., Mao,X., Wang,W., Cheng,Z., Zhou,Y., Zhou,X. and Tao,X. (2013) Viroplasm protein P9-1 of rice black-streaked dwarf virus preferentially binds to single-stranded RNA in its octamer form, and the central interior structure formed by this octamer constitutes the major RNA binding site. *J. Virol.*, **87**, 12885–12899.
61. Gillian,A.L. and Nibert,M.L. (1998) Amino terminus of reovirus nonstructural protein σNS is important for ssRNA binding and nucleoprotein complex formation. *Virology*, **240**, 1–11.
62. Gomatos,P.J., Prakash,O.M. and Stamatos,N.M. (1981) Small reovirus particles composed solely of sigma NS with specificity for binding different nucleic acids. *J. Virol.*, **39**, 115–124.
63. Taraporewala,Z.F. and Patton, J.T. (2001) Identification and Characterization of the Helix-Destabilizing Activity of Rotavirus Nonstructural Protein NSP2. *J.Virol.*, **75**, 4519–4527.
64. Lymperopoulos,K., Wirblich,C., Brierley,I. and Roy,P. (2003) Sequence specificity in the interaction of Bluetongue virus non-structural protein 2 (NS2) with viral RNA. *J. Biol. Chem.*, **278**, 31722–31730.
65. El Omari,K., Meier,C., Kainov,D., Sutton,G., Grimes,J.M., Poranen,M.M., Bamford,D.H., Tuma,R., Stuart,D.I. and Mancini,E.J. (2013) Tracking in atomic detail the functional specializations in viral RecA helicases that occur during evolution. *Nucleic Acids Res.*, **41**, 9396–9410.
66. Schroeder,R., Barta,A. and Semrad,K. (2004) Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell. Biol.*, **5**, 908–919.
67. Kelley,L.A. and Sternberg,M.J.E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.*, **4**, 363–371.
68. Roner,M.R. and Steele,B.G. (2006) Localizing the reovirus packaging signals using an engineered m1 and s2 ssRNA. *Virology*, **358**, 89–97.
69. Kar,S.R., Lebowitz,J., Blume,S., Taylor,K.B. and Hall,L.M. (2001) SmtB−DNA and protein−protein interactions in the formation of the cyanobacterial metallothionein repression complex: Zn^{2+} does not dissociate the protein−DNA complex in vitro. *Biochemistry*, **40**, 13378–13389.
70. Siegel,L. and Monty,K. (1966) Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim. Biophys. Acta*, **112**, 346–362.