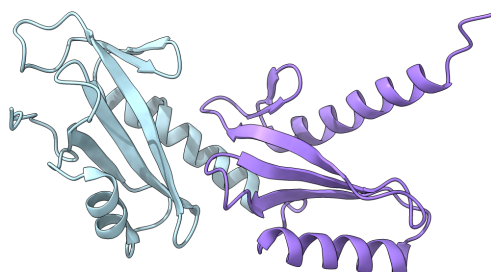# NMR-based Structural Biology

# for Studying Biomolecular Interactions

**Karel Kubíček**



**Habilitation Thesis**

**Faculty of Science**
**&**
**Central European Institute of Technology**
**Masaryk University**

**Brno 2019**

I declare hereby that this habilitation thesis is based on information taken from the listed references and from experimental work, which I performed during my research career.


Brno, December 20, 2019                                         ...................................

*To all kind people who will read this ...*

# Abstract

NMR spectroscopy is a powerful tool for studying molecules from various point of view. NMR can be used as analytical method as well as method for structure determination or for studying spin relaxation properties of the system(s) of interest.

Within this thesis, overview of methods for studying interaction of biomolecules - covering interaction of DNA with platinum anticancer drugs, protein-peptide interactions, protein-RNA interaction and interaction of large assembly using integrative approach of modern structural biology - is described to provide the reader with both practical hints and applications of the NMR spectroscopy.

Despite massive technological and methodological advances in electron microscopy and always strong X-ray crystallography, NMR has its invaluable position in studying wide range of interactions and in studies of disordered proteins.

# Acknowledgments

My scientific career would not be possible without many people I met both in private and professional life. And here I would like to thank some of them. I will not mention everybody and I apologize to those who will not be mentioned and am asking to forgive me that.

Definitely, I have to thank prof. Vladimír Sklenář and people with whom he build up the NMR lab at the Masaryk university. Amongst those people special thanks goes to prof. Michaela Wimmerová and assoc. prof. Lukáš Trantírek who were guiding my first scientific steps during my diploma thesis project and it was a successful work on preparation of bacteriophage as orienting medium in NMR spectroscopy thanks to their support.

Then I need to thank prof. Ivano Bertini and prof. Lucia Banci who were supervising me during my PhD in Florence and taught me independence, courage and helped me to mature during my stay in CERM/CIRMMP. I should not forget to mention Isabela Felli and Simone Ciofi-Baffoni and Fabio Arnesano who were really great inspiration for me.

Experience and knowledge gained during master and PhD enabled me to join Teresa Carlomagno and Christian Griesinger in Göttingen at the Max Planck Institute for Biopyhsical Chemistry and it was the best two years I ever had and am really grateful for those two years in "scientific paradise" where only sky and own imaginations are the limits. Stress and hard work is involved but many nice, smart and inspiring people. Thank you Teresa and Christian for all the discussions about everything (even the 2006 Championship in Soccer, btw: Italy gold, Germany bronze).

And when coming back home to my *alma mater* after 6 years, prof. Josef Humlíček, prof. Dominik Munzar and assoc. prof. Richard Štefl are those who I would like to thank for support, stimulations, patience and mentorship.

Special thanks goes to my friends and colleagues who were standing by my side at any time needed and were able to stand all my talks and discussions and are constantly challenging and humiliating me by their almost perfection. Namely, Erik Caha, Deli, Hana Křížová, Cristina Del Bianco, Gabrielle Cavalaro, Claudia Andreini, Massimo Lucci, Enrico Morelli, prof. Marie Krčmová, Nills Lakomek, Julien Orts, Christophe Fares, Jiří Chaloupka, Ondřej Caha, Juraj Rusnačko, Jiří Nováček, Robert Vácha, Pavel Plevka, Štěpánka Vaňáčová and others. Marek Šebesta is a great person and I am very happy for the opportunity to meet him and share office, pain and troubles as well as funny moments with him!

I would like to particularly thank my former student Magdaléna Krejčíková. Furthermore, Eliška Šmiřáková, Tomáš Hošek, Miroslav Jurásek, Olga Jasnovidová, Fruzsina Hobor, Pepa Pasulka, Veronika Janštová, Míša Twarógová, Kuba Macošek,Veronika Šubertová, Pavel Veverka, Tomáš Nováček, Martina Zánová and Matyáš Pinkas. They are / were great students who made me a better teacher, better supervisor/consultant and a better person.

Sure enough, I have forgotten some of very kind and for my life important people so I am asking them for apologies - I did not do it on purpose.

Last thanks has to go to my family for their patience, support and all they are doing for me. My wife, my kids, my mother and my brother with his family. Thank you very much and hope I will ever be able to pay back at least part of your kindness.

I would like to thank God for leaving the light at the end of the tunnel permanently ON!

# Contents

# List of Figures

# List of Tables

# 1   Introduction: *Structural Biology and NMR*

Structural biology, as defined by Nature publishing group (https://www.nature.com/subjects/structural-biology), is the study of the molecular structure and dynamics of biological macromolecules, particularly proteins and nucleic acids, and how alterations in their structures affect their function. Structural biology incorporates the principles of molecular biology, biochemistry and biophysics. This definition says it all though it could be expanded by the goal that once both structure and function of a given molecule or a complex is known, the mechanism of biological process can be understood at the atomic level resolution.

The best indicator of the evolution of structural biology is the PDB database (World-wide Protein Data Bank) http://www.rcsb.org [1]

The total number and yearly increase of the deposited structures is one of the important indicators to see, the other very important statistics is the molecular weight of the determined systems. The major part of structures deposited is of molecular mass bellow 60 kDa but it is notable more than 3% of structures is of size over 360 kDa (Fig. 1).
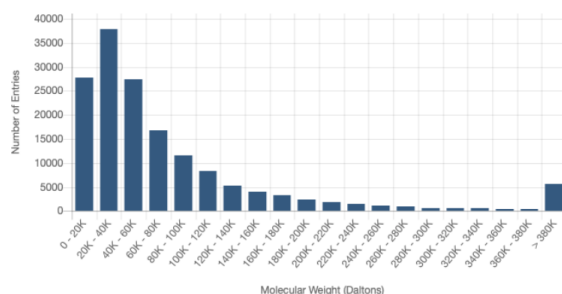


Figure 1: PDB Data Distribution by Molecular Weight - adapted from: http://www.rcsb.org/stats/distribution_molecular-weight-structure

---

[1]I started to use the database 20 years ago when it was Research Collaboratory for Structural Bioinformatics project and even nowadays prefer the mirror rcsb.org over the pdb.org since RCSB keeps the page as simple as possible. For clarity about PDB and RCSB: https://en.wikipedia.org/wiki/Protein_Data_Bank: In October 1998, the PDB was transferred to the Research Collaboratory for Structural Bioinformatics (RCSB); the transfer was completed in June 1999. The new director was Helen M. Berman of Rutgers University (one of the managing institutions of the RCSB, the other being the San Diego Supercomputer Center at UC San Diego). In 2003, with the formation of the wwPDB, the PDB became an international organization. The founding members are PDBe (Europe), RCSB (USA), and PDBj (Japan). The BMRB joined in 2006. Each of the four members of wwPDB can act as deposition, data processing and distribution centers for PDB data.) where one can retrieve, among others, information about number of structures resolved each year since 1976. While in the year of 2000 there has been only  13500 structures in total, since 2016, the number of deposited structures is nearly as high (ca. 11 thousands of structures) reaching more than 151 thousands structures today (May 2019).

The size distribution of the resolved structures is of course a clear consequence of the stability of larger biomolecules (but this does not implicate that small proteins it is always easy to express and purify in comparison with the larger ones). Technological and methodological challenges coupled with the production of large systems in amounts and purities suitable for structural biological studies is another bottleneck. However, there is no statistics provided by PDB that would show correlation of size of the studied systems with time but there are articles providing clear indications for this statement[1].

Statistics that completes the picture of structural biology realm and its evolution from the past till today is about the "Growth of Structures Released per Year" (Fig. 2). The statistics says us that 90% of deposited structures has been derived by X-ray crystallography, 8% by nuclear magnetic resonance (NMR) and 2% by electron microscopy. While crystallography is continuously delivering structures, NMR has provided only 4 thousands structures over past 10 years (out of 12.6 thousands in total) and its contributions are shrinking every year. In completely opposite situation is electron microscopy (EM), which has been registering only ca. 240 structures in 2009 while in last three and a half year there were about two thousands of structures solved by EM deposited to the PDB with total number ca. 3.1 thousands of structures.
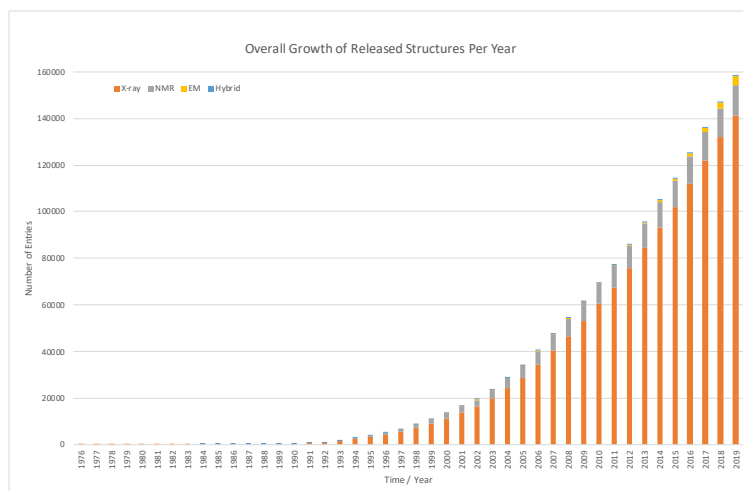


Figure 2: Overall Growth of Released Structures Per Year. Plotted are cumulative depositions of structures to PDB database as of November 2019, structures solved by X-ray (orange), NMR (grey), electron microscopy (yellow), and hybrid methods (blue) are resolved.

Last of the tables that is worth to check (and will be used later in the text) in PDB is summarizing "PDB Data Distribution by Experimental Method and Molecular Type" (Tab. 1) where one can see the NMR is competitive method only in case of determination of structures of nucleic acids but in all other fields (structures of proteins and protein/nucleic acid complexes) X-ray is the leader and EM is growing. This table clearly indicates that once the biological molecule of interest expresses well and purity and stability were optimized, X-ray and EM are the methods of choice and NMR should be used only in cases where these two methods fail.

| Expert. Method | Proteins | NA | Protein/NA Complex | Other | Total |
|---|---|---|---|---|---|
| X-Ray | 132463 | 2079 | 6819 | 8 | 141369 |
| NMR | 11264 | 1309 | 262 | 8 | 12843 |
| Electron Microscopy | 3039 | 35 | 1043 | 0 | 4117 |
| Other | 282 | 4 | 6 | 13 | 305 |
| Multi Method | 145 | 5 | 2 | 1 | 153 |

Table 1: PDB Data Distribution by Experimental Method and Molecular Type as of November 2019

Such a niche in structural biology is studying interactions of biomolecules. NMR can provide information about interaction at atomic level on a large scale of interaction strengths. Particularly strong is NMR in studying weak interactions while such complexes are difficult or impossible to crystallize and EM requires crosslinking that needs optimization steps[2].

And interaction studies with NMR is what all my work has been about during all my scientific career and what I will describe in this thesis. I used NMR for studying protein-ion interaction, protein-RNA interaction, determination of structure of cisplatinated DNA, and protein-peptide interaction. Also, I was using NMR for studying structure of a ligand that is bound to a protein which is in size exceeding limits of solution NMR.

NMR is a great method and in order to compete or complement EM or X-ray, it needs methodological and technological revolution. Since 2005/6 NMR is dominating as a method for studying disordered proteins [3, 4] and with recent huge interest in liquid-liquid phase separation[5], there is great future for NMR and valuable results will be discovered with NMR.

# 2  Systems

## 2.1  Tubulin

The first system I would like to introduce is tubulin. I will refer to a protein that is forming microtubules in eukaryotic cells and is undergoing polymerization and depolymerization in a dynamic manner. Tubulin forms a heterodimer between $\alpha$- and $\beta$-tubulin that are the building blocks (Fig. 3) of the polymerized forms (e.g. rings, sheets, microtubules). Each of the subunit is of molecular weight ca. 55 kDa and the heterodimer is formed in such a way that $\beta$-tubulin has accessible its positively charged surface while the $\alpha$-subunit is accessible on its negative surface. This obviously results in filaments with plus (+) end and minus (-) end. 13 filaments fold into a microtubule with an outer diameter of ca. 25 nm [6, 7].
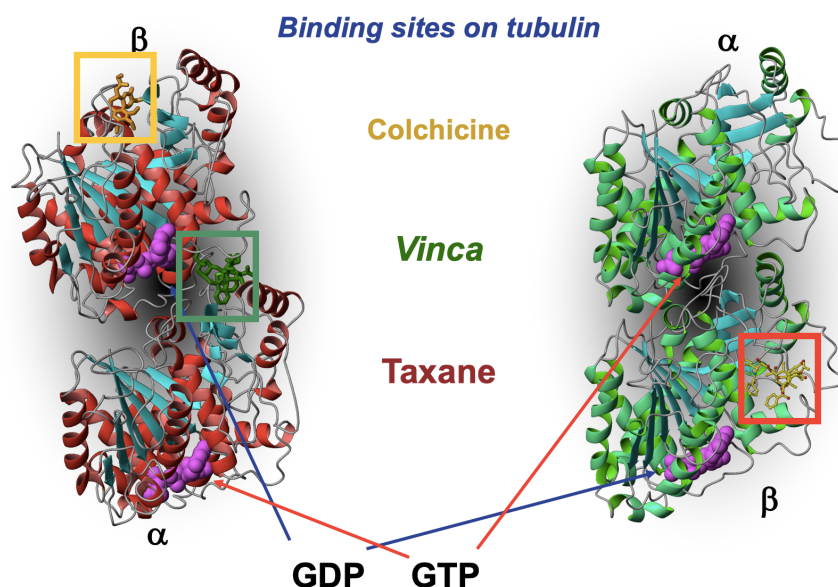


Figure 3: Ribbon representation of tubulin heterodimer ($\alpha$ and $\beta$ symbols underneath or above the ribbon structures indicate name of a given domain) with indicated binding sites for GDP, GTP and representative de/polymerizing agents. Adapted from Gigant et al., $Nature$ 2005, Nogales et al., $Nature$ 1998.

Important role and structural feature of tubulin is its dynamic de/polymerizing behaviour. Crucial role in this process plays GTP. The polymerization process occurs only in case when GTP is bound to the E binding-site of the $\beta$-tubulin. Further de/stabilization of the polymeric state of tubulin heterodimers is affected by several protein binding partners (e.g. stathmin)[8, 9, 10]. There are three binding sites named after the drugs (or alkaloids) binding to these binding

pockets – paclitaxel (Taxol®), colchicine, and vinblastine (*vinca*-site). All of these binding sites are present on the $\beta$-tubulin (Fig 3).

Tubulin has been identified as an important target for anti-cancer drugs. The tubulin-binders strongly affect the de/polymerizing equilibrium and therefore there are two groups – stabilizers (e.g. paclitaxel, epothilones) of the tubulin polymerized form, and destabilizers (e.g. tubulysin). One of the first tubulin-binding drugs isolated from Pacific yew in 1971 and approved for medical use in 1993. However, as there are severe side effects connected with the use of taxanes, several drugs sharing the efficacy of taxanes and ability to shift the polymerization equilibrium towards the microtubule form were tested. Among these natural compounds we can find e.g. discodermolides and epothilones. In the results-summarizing part of this theses articles describing interaction of tubulin with epothilone are listed.
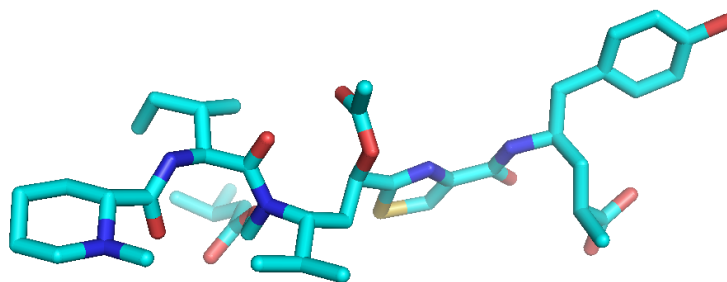


Figure 4: Stick representation of free form of tubulysin. Extended, free form of depolymerizing agent tubulysin is shown. Carbon atoms/bonds are shown in cyana, nitrogen in blue, oxygen atoms in red and sulphur atom in yellow.

Among the destabilizing agents originally isolated from myxobacterium that gained attention of pharmaceutical industry is tubulysini (Fig. 4). Tubulysin A is peptidic in nature, however most of the amino acids are atypical and there is a central thiazole heterocycle. Tubulysin A also contains a basic N-terminal tertiary amine and an aromatic and acidic C-terminal end[11]. Tubulysin A induces apoptosis in cancer cells but not in normal cells and shows significant potential antiangiogenic properties in several in vitro assays. We have succeeded in determination of the first solution NMR structure of tubulin bound tubulysin A and the methodology and detailed description of the interaction is attached in the results part of this theses[12].

## 2.2   RNA Polymerase II

RNA polymerase II (RNAPII)[13, 14] (Fig. 5) is one of the three eukaryotic polymerases that are found in nucleus. Its role is to synthetize pre-mRNA to be further translated into proteins. Besides pre-mRNA, RNAPII is transcribing also short non-coding RNAs (snRNA, snoRNA, microRNA), cryptic unstable

constructs (CUTs), stable unannotated transcripts (SUTs), and Xrn1- dependent unstable transcripts (XUTs). Since the RNAPII has to read this very diverse set of constructs and respond to a variety of environmental changes, transient interactions with various RNA, DNA, and protein factors have to control the timing and course of the RNAPII's activity.
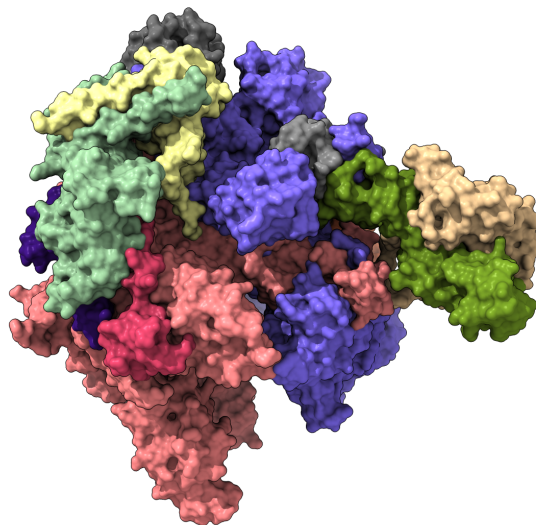


Figure 5: RNA polymerase II - surface model of RNA polymerase II with domains diferentiated by colours with the largest subunit - Rpb1 - coloured in purple. Created by UCSF Chimera from PDB ID: 5VVS

### 2.2.1   C-Terminal Domain

Platform for such transient interactions can be found on the Rpb1 subunit of the RNAPII – its C-terminal domain (CTD). CTD can not be found in any other RNA polymerase (I or III) and is therefore unique for RNAPII. CTD does not possess any well-defined structure in the free form.

The RNAPII's CTD is composed of three parts – i) N-terminal flexible linker which connects the CTD to the Rpb1 subunit, ii) central region with conserved heptad repeats with a sequence $(YSPTSPS)_n$ length of which varies through organisms (e.g. $n$ is 26 in yeast and 52 in humans) (Fig. 6), and iii) last part - divergent C-terminal sequence[15]. Deletion of the CTD is lethal as shown in studies on yeast and on mammalian cells, while removal of some repeats impairs growth, but is not lethal. Deleting nearly a half of the original repeats leads to lethality of yeasts and mammalian cells.

CTD is not only vital and exhibits structural plasticity, but it is also post-translationally modified by so called "CTD writers" or "CTD erasers"[16, 17]. Five out of the seven conserved residues that are forming repetitive heptad can be phosphorylated and the remaining two residues - prolines - can undergo isomerisation [18, 19, 20]. These modification lead then to a creation of a CTD (or its parts) with unique chemical and structural features which led Buratowski and colleagues to the proposition of the so-called "CTD-code" more than a decade ago. Later, Schwer and colleagues have shown that the CTD-code is read as a 10–11-letter word (i.e. not in single heptads but rather one and a half hepta-repeat) of sequence Y1a-S2a-P3a-T4a-S5a-P6a-S7a-Y1b-S2b-P3b-Thr4b[21]. All these features of the CTD provide it with the possibility to attract or repel a broad range of interacting factors. Proteins or interacting partners to RNAPII's CTD that are neither "eraser" nor "writers" are forming the third group called "CTD-readers" (of the CTD-code). Starting from the non-modified state of the RNAPII's CTD, there are possible $2 \times 2 \times 2 \times 3 \times 3 \times 2 \times 3 = 432$ combinations of post-translational modifications together with the two proline configurations in a single hepta-consensus repeat. In a 10–11 functional unit, these combinations multiply to $2 \times 2 \times 2 \times 3 \times 3 \times 2 \times 3 \times 2 \times 2 \times 2 \times 3 = 10368$[15]. However, it is still unclear how many of these combinations actually occur *in vivo*, and it can be expected that some modifications exclude each other and that the CTD-code is degenerate and/or less complex.
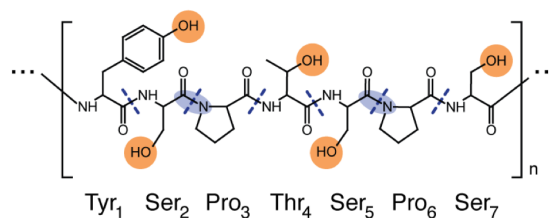


Figure 6: Consensus hepta-repeat of CTD. Groups prone to post-translational modifications (e.g. phosphorylation) are emphasized in orange, isomerizations of prolyl bonds are highlighted in blue. Primary sequence and numbering of aminoacids is shown. "n" indicates variable number of repetitions of CTD in various organisms, for details see text. O. Jasnovidova is acknowledged for the art-work - adapted from Jasnovidova & Stefl Wiley Interdiscip Rev RNA 2013

Distribution of phosphorylations throughout the whole CTD had been studied with the aim to tackle the question of number of phosphorylations per heptad. Suh and co-workers have shown that phosphorylations are uniformly distributed along the CTD with average density of ∼0.4 phospho-groups per repeat. Monophosphorylations per repeat are very abundant followed by double-phosphorylations. Triple-phosphorylated hepta-repeat of the CTD is very rare. Serine 5 followed by serine 2 are the two residues with hishest level of phosphorylations, whereas phosphorylations of the remaining three residues occur with

about 2 orders lower magnitude. Schüller *et al.*[22] hypothesize that abundance of double phosphorylations - the most common combination is Y1-S2 - are result of transition between trsncription cycle stages.

To complete the biological and structural picture of the CTD, chromatin immunoprecipitation (ChIP) experiments have been performed on yeasts and human cells with the aim to reveal phosphorylation profiles throughout the transcription cycle. These experiments have shown a surprising difference in profiles in yeasts and in human cells with the most significant difference in Tyr1, which peaks at the beginning of the transcription cycle in human cells while in yeast is mostly phosphorylated towards the end of the cycle[23, 24].

During my work on RNAPII, I was mostly involved in studies of RNAPII with "CTD-readers", namely CTD-interacting domains (CIDs) and with tSH2 domain of elongation factor Spt6. The following chapters hence will provide introduction to the structural and functional features of these CTD-binding factors.

## 2.3   CTD Interacting Domains

CTD-interacting domains is a family of protein domains that recognize specific phosphorylation patterns of the CTD-heptads. There are four multi-domain proteins containing CID domains found in budding yeast – Pcf11, Rtt103, Nrd1, and Ctk3. In further text, I will focus on Rtt103's and Nrd1's CID domains[25, 26, 27].

CID domains are typically 130-160 amino-acids long domains arranged in helical bundle. There are 8 helices[2] forming the bundle (Fig. 7). CTD interacting surface is formed by helices $\alpha 2$, $\alpha 4$, and $\alpha 7$. Despite high overal structural conservancy of the CID domains among the proteins, primary structure composition and secondary structure elements vary, forming so interactiong surface specific for given CID domain.

### 2.3.1   Structural & Functional Features of the CIDs

**2.3.1.1   $\beta$-turn Recognition**   The central residues of the CTD's consensus hepta-repeat, SPTS, are prone to form $\beta$-turn that is stabilized by intramolecular hydrogen bonds. The $\beta$-turn conformation of the CTD seems to be independent on serine 2 phosphorylation state[28].

**2.3.1.2   *cis/trans* Isomerization**   As mentioned above, besides post-translational modifications, the CTD-code is also defined by prolyl bond isomerization. Proline residue is being able to rather easily adopt both *cis* and *trans* isomers of its prolyl peptide bonds. Which conformation will the prolyl bond assume can be also affected by the type of amino acid preceding the prolyl peptide bond. The CID of Nrd1[29, 19] has amongst CID domains a unique, positively charged,

---

[2]the $6^{th}$ helix is composed of two smaller helices and often visualizing programs display these two helices as one long $\alpha$-helix hence careful reader may be misled while counting only 7 helices in the CID structure
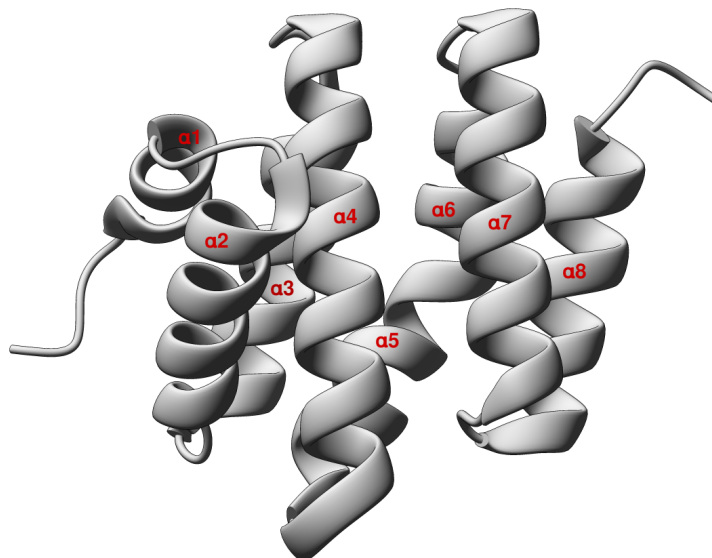
Figure 7: CTD Interacting Domain - CID from protein Rtt103 as a representative of CID domains, numbering of helices is indicated for clarity. Picture produced with UCSF Chimera using PDB ID: 5M9D

flexible region at the tip of helix $\alpha 2$. There are serine 25 and arginine 28 located in this region and recognizing phospho-mark of the CTD's serine 5. Because of this recognition, the phosphorylated serine 5-proline 6 peptidyl-prolyl bond can be found in an unusual *cis* configuration that maximizes the CTD's contact to the $\alpha 1$-$\alpha 2$ tip region.

**2.3.1.3  Electrostatic Recognition & Key Residues**  As we accept the concept of the CTD-code, there has to be "reader" protein to read the code or specific phosphorylation pattern of the CTD-stretch that is recognized. CIDs provide interaction surface that is able to accommodate 10-11 residues of the CTD, which is in agreement with the above mentioned length of the functional CTD-unit.

   The two readers that recognise phosphorylation marks at specific position are CIDs from proteins Rtt103 and Nrd1. Whereas Rtt103 recognizes phosphorylation at serine 2 (and as recently shown also threonine 4), the CID of Nrd1 protein recognizes CTD phosphorylated at position of serine 5. Rtt103 recognizes pS2 by forming an H-bond with the guanidinium group of Arg108. In the case of Nrd1's CID, the recognition of phosphomark at position of serine 5 is achieved by two residues at the top of $\alpha 2$ – serine 25 and arginine 28. Nrd1 CID reads out serine 2 phosphorylation of the CTD sequence that is placed

up-stream of the $\alpha$-turn while the second serine 5 within the $\beta$-turn, appears not to be specifically recognised at all.

### 2.3.1.4   Non-CTD binding CID domain & Non-CTD interactions

CID domains or their structural analogues that, however, do not recognize any CTD motif are also found is in some proteins. For example the Ctk3 subunit of the CTDK-I kinase which is phosphorylating CTD to promote the elongation step of the transcription cycle. The canonical interaction surface of Ctk3's CID formed by the helices $\alpha2$, $\alpha4$, and $\alpha7$ is negatively charged hence is not suitable to attract CTD but rather different substrates[30].

It is also important to note that the consensus hepta-repeats of the CTD are not the only binding partners to CIDs. As shown recently, CID may also interact with other substrates with rather high affinity ($K_D$ of $\sim 1~\mu$M), for example C-terminus of Trf4-protein. For its strong interaction with Nrd1's CID, the peptide of sequence DDDEDGYNPYTL was dubbed NIM – Nrd1 Interacting Motif.

## 2.4   Cisplatin & its derivatives

### 2.4.1   Cisplatin

Cisplatin (Fig.  8)is a square plannar inorganic compound containing platinum(II) as a central atom, and two chloride anions and two amine groups in *cis*-configuration, as ligands. Cisplatin has been successfully applied as anti-tumor therapeutics against testicular, bladder, neck, ovarian, and head cancer since it has been approved by the Food & Drug Administration (FDA) in the USA. *Trans* isomer of platinum(II) complex exists as well but its activity is much lower than the one of cisplatin hence it will not be further mentioned in the following text[31, 32].

Cisplatin, *cis*-diamminedichloroplatinum(II) enters the cell as a neutral molecule, which is activated in the *intra*-cellular environment where concentration of chloride ions drops from $\sim$100 mM to $\sim$4 mM concentration and one of the chlorides is replaced by water molecule to form mono-aqua $[\text{PtCl}(\text{H}_2\text{O})(\text{NH}_3)_2]^+$ cation, and more slowly the di-aquo $\text{Pt}(\text{H}_2\text{O})_2(\text{NH}_3)_2]^{2+}$ cation. This positively charged molecule crosses nuclear membrane and crosslinks DNA, but can also bind to RNA and sulfhydryl groups of amino-acid side-chains in proteins. Bifunctional cisplatin cation is forming first monofunctional adduct with the preference to guanine's N7 and subsequently major *intra*-strand cross-link to adjacent purine base.

While the 1,2-*intra*-strand d(GpG) cross-link between cisplatin and two adjacent guanines is the major product, 1,2-*intra*-strand d(ApG) can be also formed. Furthermore, 1,3-*intra*-strand d(GpNpG), *inter*-strand cross-link and monofunctional adduct are created as well.

These 1,2-*intra*-strand cisplatin cross-links bend and unwind DNA and are hence blocking DNA replication and transcription and lead to a programmed

cell death – apoptosis. Despite high efficacy of cisplatin in treatment of cancer, there are numerous side effects connected with the toxicity of cisplatin, e.g. nephrotoxicity, nausea, neurotoxicity and nausea and vomiting. Additionally, some types of tumour exhibit resistance against cisplatin treatment. To improve better cisplatin delivery in the cell and enlarge the spectrum of cisplatin treatable cancers, development of new drugs derived from cisplatin has been performed[33, 34].



Cisplatin

Carboplatin, 2$^{nd}$ generation drug
(lower toxicity)

Oxaliplatin, 3$^{rd}$ generation drug
(broader spectrum of activity)

Nedaplatin

Satraplatin

Lipoplatin
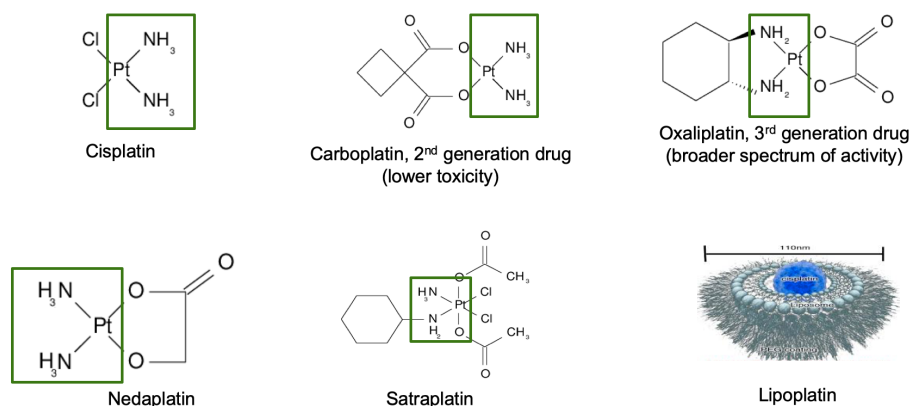
Figure 8: Cisplatin and its $2^{nd}$ and $3^{rd}$ generation analogues are shown. Green rectangle is emphasizing the core structure of cisplatin which is common for all cisplatin drugs and its derivatives. Adapted from: $JMedChem$, 55: 7182–7192, 2012

### 2.4.2  Cisplatin Derivatives

There are several representatives that should accomplish the goal of better deliverable and more potent successors of cisplatin I would like to mention. They are representatives of second and third generation (Fig. 8), carboplatin and oxaliplatin, respectively. And lipoplatin is derivative of cisplatin with the aim of better delivery of cisplatin molecule into the cell[35].

Carboplatin - $cis$-diammine-(1,1-cyclobutanedicarboxylato)platinum(II) – is a member of the so-called second generation of platinum drugs. Carboplatin interacts with DNA the same way as cisplatin, i.e. forming the same type of 1,2-(GpG) adducts with the difference that the conversion to the active complex is much slower in the case of carboplatin. Carboplatin is sold since 1986 when it was approved for medical use as Paraplatin and since then has been used against neuroblastoma, head and neck cancers as well as against ovarian and lung cancer. Nevertheless, carboplatin shows cross-resistance with cisplatin[36].

Development of cisplatin analogues is still ongoing. Lipoplatin, picoplatin, satraplatin and tetraplatin are some of the tested or even promising ones. Lipoplatin[37] is assembly of cisplatin encapsulated in a micelle with the aim to

improve the delivery of the drug into the cell, while picoplatin (*cis*-amminedichloro-
2-methylpyridineplatinum(II)) shows lower inactivation by non-DNA interac-
tions and higher activity in cisplatin resistant cell lines. Tetraplatin, tetrachloro-
(D,L-*trans*)1,2-diamminocyclohexaneplatinum(IV), is one of the candidates that
was tried in clinical test but did not show sufficient activity compared to cis-
platin and was not approved.

## 2.5   RNA binding proteins

In the previous chapters, wide range of interaction regimes and interacting part-
ners have been introduced (weak to strong binding or even covalent bond/cross-
link in the case of cisplatin). In these case proteins were interacting with pep-
tides or other proteins or nucleic acid with inorganic or organic compounds. In
this chapter, proteins interacting with nucleic acids or specifically RNA, will be
presented[38, 39].
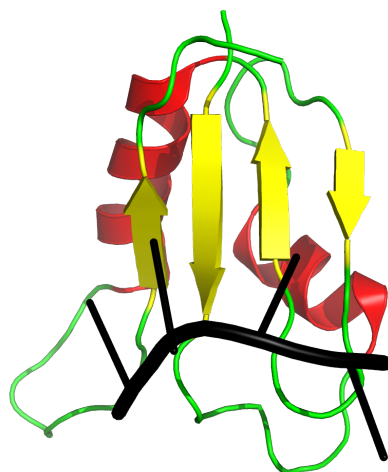


Figure 9: Cartoon representation of RRM domain of protein Nab3 (PDB ID:
2L41) with $\beta$-strands shown in yellow, $\alpha$-helices in red, and RNA oligomer
UCUU in black. $\beta$-sheet is of topology $\beta4\beta1\beta3\beta2$ (left to right) for details see
the text.

   There are several binding motives or protein domains that interact with RNA
to regulate many aspects of RNA processing, such as RNA splicing, polyadeny-
lation, capping, modification, transport, localization, translation, and stability,
called RNA metabolism. Out of the known RNA-binding proteins or motives
(RBPs), namely - double-stranded RNA binding motives (**drRBM**), Sterile
alpha motif (**SAM**), **PAZ** domain named after proteins Piwi, Argonaut and
Zwille, **zinc fingers**, K homology domains (**KH**), highly conserved motives

in RNA helicases - Asp-Glu-Ala-Asp (**DEAD/DEAH boxes**), Pumilio/FBF (*Caenorhabditis elegans* Pumilio-*fem*-3 binding factor, **PUF**) domains, pentatricopeptide-repeat domains (**PPR**), and RNA-recognition motifs (**RRM**s) further text will describe the RRM domains only.

The RNA recognition motif (**RRM**) or RNA-binding domain (**RBD**) or even ribonucleoprotein domain (**RNP**)[40] is a family of ∼90 amino acids domains that contain a central sequence of eight conserved residues that are mainly aromatic and positively charged. The canonical fold of RRM is an $\alpha\beta$ sandwich with a topology $\beta1\alpha1\beta2\beta3\alpha2\beta4$. The domain structure is composed of a $\beta$-sheet formed by four antiparallel $\beta$-strands in order (left to right) $\beta4\beta1\beta3\beta2$. The helices $\alpha1\alpha2$ are packed against the $\beta$-sheet (Fig. 9)[41].

Sequence of the central region of the RRM, termed RNP, is defined by consensus residues Arg/Lys-Gly-Phe/Tyr-Gly/Ala-Phe/Tyr-Val/Ile/Leu-X-Phe/Tyr, where X stands for any amino-acid. Since later, another motif with a sequence Ile/Val/Leu-Phe/Tyr-Ile/Val/Leu-X-Asn-Leu has been defined, the first motif is referred as RNP1 and the second as RNP2.[42] The binding of the RRM to RNA is through three aromatic side-chains from the central $\beta$-strands that stack to three RNA nucleotides. Both motives were shown to bind RNA with wide range of affinities and specificities, while the typical length of recognized RNA is four nucleotides. In addition to the protein-RNA interactions, RRM can be involved in protein-DNA interactions or even in protein-protein interactions.

# 3 Before We Start to Measure Interactions With NMR

Despite the fact that NMR is a great method for studying molecules and interaction and dynamic processes in solution, there are several important points that have to be taken into consideration prior of any NMR study is performed. There are five options one can face before studying complex formation between biomolecules: i) unknown structure of the protein of interest, ii) known X-ray structure of the protein of interest, iii) known NMR structure of the protein of interest, iv) known chemical shift assignment, v) unknown chemical shift assignment. Any combination of these situations is imaginable and the best for NMR studies is situation iii)+iv) since it means most of the tedious work has been done and one can focus on interaction studies only. In all other cases, substantial time and money investment has to be taken into account.

## 3.1 Sample Preparation

To have full profit of the NMR technology when studying interactions between biomolecules, one of the interacting partners has to be mostly $^{15}$N, and/or $^{13}$C enriched. Sometimes specific labelling is also beneficial but that is beyond the scope of this overview.

It is good to design well what is the goal of the study and chose labelling / enrichment accordingly. With $^{15}$N labelled sample, which is usually sufficient, most of the amino-acids can be studied in detail including side-chain residues. While histidine's imidazole ring can be measured through long-range $^{1}$H-$^{15}$N correlation in the range 160—240 ppm in $^{15}$N dimension, N-H pair in arginine's side-chain appears between 75-95 ppm in a standard $^{1}$H-$^{15}$N HSQC experiment and the same is true for -NH$_2$ groups in side-chains of asparagine and glutamine. Backbone N-H pairs are monitored through std. $^{1}$H-$^{15}$N HSQC with the obvious exception of prolines.

If aliphatic and aromatic side-chains are of interest, $^{13}$C enriched sample is needed and cleavage of His-tag (if used for purification purposes) is strongly recommended.

## 3.2 Chemical Shift Assignment

The bottleneck of NMR studies of biomolecules is chemical shift assignment. It is the most demanding step on wo/man-power, experimental time, processing and analysis time. There are several strategies to achieve the NMR assignment but still time to successful obtaining of the chemical shift assignment varies from weeks to months.[43, 44]

## 3.3 Structure Determination

Structure determination (of one or both interacting partners using NMR spectroscopy) is a step that is not necessarily needed when studying interaction of

biomolecules. If situations ii) or iii) apply (i.e. structure of one of the interacting partners is known from either X-ray or NMR or electron microscopy), with the use of bioinformatic tools and/or molecular dynamics, valuable results can be obtained. Otherwise, when structure is unknown, one has to follow the standard procedures relying on nuclear Overhauser effect (NOE), interligand NOE, transferred NOE, scalar couplings, residual dipolar couplings, paramagnetic shifts, paramagnetic shielding etc[44].

## 3.4  NMR Spin Relaxation Methods

Besides "static" information about chemical shift assignment or structure of a free or complexed protein, NMR provides also information about the dynamics of the protein(s) and change in spin relaxation upon interaction with binding partner. Relaxation measurement in NMR enable us monitoring dynamic properties over a wide range of time-scale covering large portion of considerable motions in biomolecules. The biggest drawback of dynamic measurement is that biological community does not appreciate these enough and only a small portion of dynamic data is recognized despite big effort of NMR spectroscopists and molecular dynamics experts[45, 46].

# 4  Studying Interactions by NMR

## 4.1  NMR as a Universal Tool For Studying Interactions in Solution

All biological processes are happening through the interactions between biomolecules and their interactions with organic or inorganic compounds. The course of the processes is strongly dependent on the kinetics and thermodynamics of the interactions. While for the thermodynamics the typical physico-chemical method of choice is isothermal titration calorimetry (ITC), there are several experimental methods that enable us to study the kinetics of the interactions. Amongst these methods, solution nuclear magnetic resonance (NMR) deserves special attention as it allows to monitor and identify many aspects of the interaction – identifying the key residues for the interaction to appear, revealing the interaction surface and also through relaxation measurement suggest the molecular weight of the reaction adduct (REF). When these information are combined with ITC or fluorescence anisotropy measurement, almost complete picture of the interaction process is revealed.

### 4.1.1  NMR Titration

The first and most crucial step in studies of interactions using NMR spectroscopy is preparation of the complex of interest which is achieved through an NMR titration. With NMR titration we understand a stepwise addition of one of the interacting partners (titrant or ligand) in small volumes and high concentrations to the solution of volume of several hundred of microliters (depending of NMR tube used – 3 mm, Shigemi, and 5 mm range from 180 microliters, through 280 up to 600 microliters, respectively) containing the interacting partner (protein or titrand). Addition of small volumes of titrant to high volumes of the titrand maintains the total volume of the titration nearly constant hence the concentration of the titrand is constant as well. An NMR spectrum is measured after each addition of the titrant and changes in the spectrum are analyzed. NMR titration can be performed through 1D as well as through multidimensional NMR spectra. In case of protein NMR spectroscopy, the most typical set up is that one of the partners is $^{15}$N and or $^{13}$C $/^{15}$N enriched and series of heteronuclear-single-quantum coherence spectra (HSQC) is measured throughout the titration.

### 4.1.2  Fast to Slow Exchange Regimes - Analysis of NMR Titration

Upon completion of the titration, both $K_d$ and interaction surface of the labeled titration partner can be identified supposing previous knowledge of backbone resonance assignment of the measured spectra.

Dissociation constant or $K_d$ is in case of NMR titration derived from changes in chemical shift changes throughout the titration. The first requirement for the analysis of such experiments is to establish which exchange condition applies.

For many complexes studied by NMR, the association rate constant will be diffusion–limited, with value in the range $10^7$–$10^8$ $M^{-1}s^{-1}$. To a first approximation, then, the exchange rate will depend on the equilibrium constant. If the complexes bind tightly (with dissociation constant $K_d < 10^{-7}$ $M$), the **slow** exchange condition usually applies, and **separate** signals for the complex and the free species are observed[47].

When $K_d > 10^{-3}$ $M$ the dissociation rate constant are larger, and exchange between the complex and the free species is often **fast** on the NMR timescale. Under these conditions **a single averaged resonance** is observed for nuclei at corresponding sites in the free and bound species, and this resonance **shifts progressively** as the relative concentrations of the two species are changed.

Plotting the chemical shift variation (in fast exchange mode) or the intensity increase (in the slow exchange one) against molar ratio of the proteins throughout the titration, a **binding curve** is obtained and its analysis yields both either the chemical shift or the intensity of the fully formed complex and the dissociation constant. The most commonly used approach for studies of protein–protein (also protein–ligand) interactions in NMR is acquisition of serious of $^1$H-X HSQC spectra (X is substituted either for $^{15}$N or for $^{13}$C) throughout the titration from which the intensity or chemical shift variations can be evaluated[48].

Between the two extremes of "fast" and "slow" exchange, complex with non-Lorentzian line–shapes from species in intermediate exchange is observed. Detailed analysis of such a spectrum would allow one to determine chemical shifts and lifetimes of the contributing species. This lineshape analysis is often difficult to perform in experiment, hence in practice, the assumption of fast exchange is usually made to facilitate the analysis, since there is no simple test to establish whether in fact the fast exchange condition is completely fulfilled.

Considering a protein $A$, that binds to a protein $B$, to form a 1:1 complex AB:

$$A + B \rightleftharpoons AB \tag{1}$$

The dissociation constant, $K_d$, is defined as follows:

$$K_d = \frac{[A][B]}{[AB]} = \frac{([A]^0 - [AB])([B]^0 - [AB])}{[AB]} = \frac{k_{off}}{k_{on}} \tag{2}$$

and the complex concentration is given by:

$$[AB] = \frac{1}{2}[([A]^0 + [B]^0 + K_d) - \sqrt{([A]^0 + [B]^0 + K_d)^2 - 4[A]^0[B]^0}] \tag{3}$$

where $[A]^0$ and $[B]^0$, in a case of an 1:1 complex formation, can be written:

$$[A]^0 = [A] + [AB]$$
$$[B]^0 = [B] + [AB]$$

The fast exchange condition is:

$$2\pi \mid \nu_F - \nu_B \mid \ll \left( \frac{1}{\tau_F} + \frac{1}{\tau_B} \right) \tag{4}$$

where $\nu_F$ and $\nu_B$ are the resonant frequencies (in Hz) of the nucleus in the free and bound states, respectively, and $\tau_F$ and $\tau_B$ are its lifetimes in those states (with $\tau_B = 1/k_{off}$). If this condition is fulfilled, we shall observe a single resonance at a frequency:

$$\nu_{obs} = P_F \nu_F + P_B \nu_B \tag{5}$$

where $P_F$ and $P_B$ are the fractional populations of the free and bound states,

$$P_B = \frac{[AB]}{[A]^0} = 1 - P_F \tag{6}$$

so that

$$\nu_{obs} = \left( 1 - \frac{[AB]}{[A]^0} \right) \nu_F + \frac{[AB]}{[A]_0} \nu_B \tag{7}$$

or

$$(\nu_{obs} - \nu_F) = \frac{[AB]}{[A]^0} (\nu_B - \nu_F) \tag{8}$$

Combining Eqs. (3) and (8) it can be seen that the variation $(\nu_{obs} - \nu_F)$ as a function of $[A]^0$ (with $[B]^0$ constant[3]) can be analyzed by nonlinear least-square fitting in terms of two unknowns, $(\nu_B - \nu_F)$ and $K_d$ [47, 50, 49, 51, 48].

The change in resonant frequencies induced by complexation is usually normalized to a reference value (*i.e.* $\Delta\nu = \nu_{obs} - \nu_f$). In addition, it is useful to define the maximum frequency ($\Delta\nu_{max}$) that would be observed in the limit of full formation the complex. In the above described system, this would be $\nu_B - \nu_F$. Hence, the final equation to be used for fitting the experimental data looks as follows:

$$\nu_{obs} = \Delta\nu_{max} \left\{ \frac{([A]^0 + [B]^0 + K_d) - \sqrt{([A]^0 + [B]^0 + K_d)^2 - 4[A]^0[B]^0}}{2[B]^0} \right\} \tag{9}$$

In the case of slow exchange, the binding constant can still be quantitated by measuring the intensities of the disappearing and/or appearing peaks as a function of the titration progression.

## 4.2   Mapping Protein Binding Surface

Qualitatively important information about interacting surface of the labeled partner can be obtained by monitoring the chemical shift perturbation per residue. The chemical shift of $^{15}$N and $^1$H are particularly sensitive to any

---

[3]for corrections on concentration, see [49]

change in environment; thus, perturbation of these shifts as a result of complex formation provides a highly sensitive tool for the mapping of binding sites on a protein. Method commonly used to combine perturbations observed in both dimensions of $^{15}$N-$^{1}$H HSQC throughout an NMR titration is described by following equation [52, 53]:

$$\Delta\delta_{av} = \sqrt{\frac{(\Delta\delta^{15}\text{N}/5)^2 + (\Delta\delta^1\text{H})^2}{2}} \tag{10}$$

where $\Delta\delta_{av}$ is average amide chemical shift perturbation, in which $\Delta\delta^{15}$N and $\Delta\delta^1$H represent the change in the chemical shift of amide nitrogen and proton, respectively.

The binding interface hence will be defined by the residues exhibiting the largest shifts upon the complex formation, which in other words means that the residues that do not exhibit any change or exhibit changes within the range of error are those not-localized at the binding interface.

The interaction surface and key residues can be of course also determined by solving the structure of the studied complexes but such approach is far from being quick and easy, where quick is also possible only when the resonance assignment is already available.

Another option for determination of interaction surface is performing hydrogen exchange experiments [28], however, these have not been used within any of the presented works and this approach will therefore not discussed.

# 5 Results

This section is formed from articles that were either submitted or already published during my scientific career. The publications were selected such to demonstrate the spectrum of my know-how both from the part of studied systems as well as from the point of view of techniques that have been used to achive the published results.

## 5.1 The Tubulin-Bound Structure of the Antimitotic Drug Tubulysin

The attached publication is describing the use of transferred nuclear Overhauser enhancement for deteremination of tubulin-bound structure of tubulysin - a tubulin-depolymerizing agent with promissing anti-cancer effect.

## Conformation Analysis

# The Tubulin-Bound Structure of the Antimitotic Drug Tubulysin

*Karel Kubicek, S. Kaspar Grimm, Julien Orts, Florenz Sasse, and Teresa Carlomagno\**

Tubulysin is a highly cytotoxic peptide isolated from the myxobacterial species *Archangium gephyra* and *Angiococcus disciformis*.[1] It consists of *N*-methylpipecolic acid (Mep), L-isoleucine (Ile), and the two unusual and novel amino acids tubuvaline (Tuv) and tubutyrosine (Tut)[2] (Scheme 1). Tubu-
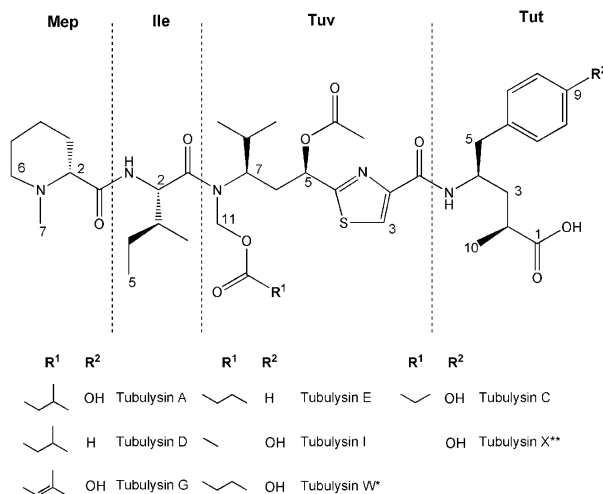


**Scheme 1.** Chemical structure of natural tubulysins; Mep, *N*-methylpipecolic acid; Ile, isoleucine; Tuv, tubuvaline; Tut, tubutyrosine (\* = deacetyl; \*\* = *N*,*O*-deacetyl).

lysin displays extremely potent cytotoxic activity in mammalian cells, including multidrug-resistant cell lines, with IC$_{50}$ values in the lower nanomolar range.[1,3] The cytotoxic activity of tubulysin is connected with its ability to interfere with microtubule (MT) dynamics and to inhibit tubulin polymerization both in vivo and in vitro.[2] With respect to its peptidic

nature and biological activity, tubulysin is closely related to dolastatin-10, an anticancer agent derived from the marine mollusc *Dolabella auricularia* (see Scheme S1 in the Supporting Information). Other tubulin ligands, such as vinblastine and colchicines, share the MT-destabilizing activity of tubulysin while showing a very different chemical structure (see Scheme S1 in the Supporting Information).

The binding sites of vinblastine and colchicines to tubulin have been elucidated by X-ray crystallography[4,5] and provided the basis to propose a mechanism for the cytotoxic activity of these two MT-destabilizing agents. Tubulysin A (TBS, Scheme 1) strongly inhibits the binding of vinblastine to tubulin; however, the experiments suggest a noncompetitive pattern of inhibition,[2] as previously observed for other antimitotic peptides, such as dolastatin-10 and phomopsin A.[6a–c] Instead, antimitotic polyketides, such as rhizoxin[6a] and disorazol,[2] have been shown to inhibit binding of vinblastine to tubulin in a competitive manner. This led to the proposal of two binding sites for MT-destabilizing agents on tubulin, a *Vinca* site where the *Vinca* alkaloids bind and a peptide site for phomopsin and TBS binding, both sites situated in a region of the β-tubulin called the *Vinca* domain.

In addition to the many classes of MT-destabilizing agents, other groups of compounds, such as the epothilones and paclitaxel, function as MT-polymerizing enhancers. In vitro studies showed that induction of polymerization of tubulin by these compounds is strongly inhibited in the presence of TBS,[2] and that TBS is able to dissolve MTs formed in the presence of paclitaxel or epothilone.[2]

Knowledge of the structure of the tubulin–TBS complex is essential to understand the mechanisms of action of TBS and the interplay between this MT-destabilizing agent and other tubulin ligands. Direct structural information on the interaction of natural products with tubulin includes complexes of tubulin with either taxol or epothilone A (EpoA) as the polymerizing agents,[7–9] and with vinblastine and colchicines as the depolymerizing agents[4,5] (see Figure S1 in the Supporting Information). However, in the case of TBS structural knowledge is limited to the unbound conformation of TBS in a methanol/water mixture, determined by X-ray diffraction, or in dimethyl sulfoxide (DMSO), studied by NMR spectroscopy.[3]

Herein, we describe the bioactive tubulin-bound conformation of TBS, as determined by NMR structural analysis in aqueous solution using transferred NOE (tr-NOE) data. We find that the bioactive conformation of TBS dramatically differs from that in the solid state determined by X-ray crystallography. The conformational differences between the unbound and tubulin-bound forms presented here allow the first rationalization of the biological data available on the tubulin depolymerization activity of tubulysins.

[*] Dr. J. Orts, Priv.-Doz. Dr. T. Carlomagno
EMBL, Structural and Computational Biology Unit
Meyerhofstrasse 1, 69117 Heidelberg (Germany)
E-mail: teresa.carlomagno@embl.de

Dr. K. Kubicek,[+] [++] Dr. S. K. Grimm[++]
Department of NMR-Based Structural Biology
Max Planck Institute for Biophysical Chemistry
Am Fassberg 11, 37077 Göttingen (Germany)

Dr. F. Sasse
Department of Chemical Biology and Medicinal Chemistry at the
Helmholtz-Zentrum für Infektionsforschung GmbH
Inhoffenstrasse 7, 38124 Braunschweig (Germany)

[+] Present address: Department of Condensed Matter Physics, Faculty
of Science, Brno (Czech Republic)

[++] These authors contributed equally to this work.

Supporting information for this article is available on the WWW
under http://dx.doi.org/10.1002/anie.200906828.

There are only weak cross-peaks observed in the NOESY spectra of tubulin-free samples of TBS, as expected for a small molecule tumbling rapidly in solution. In contrast, the NOESY spectra of a 500 μM solution of TBS in the presence of 10 μM tubulin show intense cross-peaks (tr-NOE), thus indicating that TBS binds to soluble tubulin. In the presence of tubulysin and under the conditions described in the Experimental Section, tubulin is either dimeric or forms rings (see Figure S2 in the Supporting Information).

A total of 197 nonredundant tr-NOE peaks were identified in the NOESY spectrum of a 500 μM aqueous solution of TBS in the presence of 10 μM tubulin, and were used as restraints in the structure calculation (see Figure S3 in the Supporting Information). The full relaxation matrix approach was used to calculate the structure of tubulin-bound TBS. This method minimizes the difference between the computed and the experimental 2D NOE intensities while accounting for spin-diffusion effects. The structure calculation converged to a unique family comprising the 12 lowest-energy conformers (Figure 1; heavy atoms, root mean square deviation 0.33 Å).
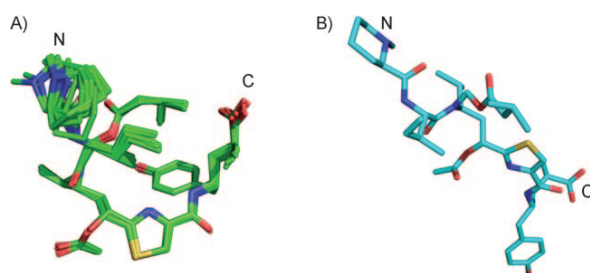


***Figure 1.*** Conformation of A) tubulin-bound (overlap of the best twelve NMR structures) and B) free TBS. Color code: blue N, red O, yellow S, green/cyan C.

The tubulin-bound conformation of TBS (Figure 1A, green) is compared to the free (unbound) conformation of TBS determined by X-ray crystallography (Figure 1B, cyan).[3] The compact structure of tubulin-bound TBS largely differs from the extended unbound conformation. In the bound state, the nitrogen atom of the thiazole ring of Tuv and the aromatic ring of the Tut form a basal platform at the bottom of the molecule upon which the *O*-acyl *N,O*-acetal side chain of Tuv packs to form a hydrophobic core. The Ile and the methyl group at the C-terminal end of the Tut also contribute to this core. The *O*-acetylmethyl group closes the hydrophobic pocket at the bottom. The remaining backbone towards the N-terminal *N*-pipecolic acid protrudes from this core region, as does the *O*-acetyl group, both of which are located on the same side of the basal platform. Comparison of the dihedral angles of the tubulin-bound and free conformations of TBS revealed several major differences, which are described in detail in the Supporting Information (Table S1).

The piperidine ring of the *N*-pipecolic acid shows only sparse interresidual tr-NOE data, whereas the expected number of intraresidual signals from the methylene groups is present in the NOESY spectra. The NOE data are not consistent with a single conformation of the ring but indicate a

mixture of twisted boat and chair conformations in the ratio 60:40 (Figure 1). Consistently with this conformational flexibility, the piperidine ring points away from the rest of the molecule and is likely not to be part of the pharmacophore. In contrast, in the tubulin-unbound structure the piperidine ring adopts a well-defined chair conformation.

Figure 2 shows a summary of the structure–activity relationship (SAR) data[10,11] mapped on the three-dimensional structure of tubulin-bound tubulysin. The tertiary
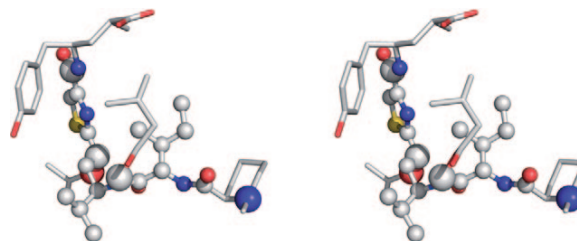


***Figure 2.*** Stereo view of the tubulin-bound three-dimensional structure of TBS mapping of the SAR data available from the literature. The radius of the spheres, which represent C (gray), N (blue), S (yellow), and O atoms (red), is proportional to the relevance of the site for cellular activity. No sphere indicates that the corresponding atoms can be eliminated without affecting the bioactivity; middle-size spheres represent the sites where modifications were not tested, while the large spheres represent the sites where modification of stereochemistry or suppression of the functionality negatively impacts activity.

amine group of the Mep residue at the N terminus is essential for activity, but very simple low-molecular-weight substituents are acceptable at this site. Neither the tyrosylethyl nor the γ-carboxy groups of the C-terminal Tut residue are relevant for cytotoxicity. Similarly, both labile Tuv C5-acetyl and *O*-acyl *N,O*-acetal groups can be eliminated without affecting cellular activity. On the other hand, a change in the stereochemistry at the Tuv C5 atom negatively impacts the ability of TBS to destabilize MTs.

The tubulin-bound structure of TBS allows rationalization of these SAR data. The piperidine ring of the N-terminal Mep, the tyrosylethyl and γ-carboxy groups of the C-terminal Tut, and the *O*-acyl *N,O*-acetal group are all situated on the same side of the molecule and are all dispensable for biological activity. On the other hand, the hydrophobic skeleton of the molecule, which assumes a T-shaped form described by the thiazole ring, the valine side chain of Tuv, and the Ile side chain, seems to be essential, as indicated by the relevance of the stereochemistry at the Tuv C5 atom. On the opposite side of the T-shaped core, the hydrophilic tail of the Mep residue is likely to be involved in a hydrogen bond with tubulin side chains, as indicated by the relevance of the tertiary amine group at the N terminus of TBS.

Tubulysin has been shown to inhibit the binding of vinblastine to tubulin in a noncompetitive manner.[2] Thus, the location of the tubulin-binding site of TBS remains unknown. To address this question, we applied the INPHARMA method[12,13] to two samples: 1) TBS (500 μM), EpoA (500 μM), and tubulin (10 μM); and 2) TBS (500 μM), vinblastine (500 μM), and tubulin (7 μM). In INPHARMA,

protein-mediated interligand NOEs can be observed between two ligands binding weakly and competitively to the same binding pocket of a macromolecular receptor. In agreement with the biological data, which suggest a noncompetitive inhibition of the tubulin binding of vinblastine by TBS, we did not observe any protein-mediated interligand NOEs between TBS and vinblastine. However, we cannot exclude that the failure in the observation of the interligand NOEs may result from the slow dissociation constant, $k_{off}$, of vinblastine rather than the presence of two different binding sites for the two drugs.

Surprisingly, we observed interligand NOEs between EpoA and TBS, which suggests that the two drugs share a common binding site on tubulin (see Table S2 and Figure S4 in the Supporting Information). Tubulin-mediated interligand NOEs are observed homogeneously for almost all protons of the two drugs, thus suggesting that they do not bind simultaneously to neighboring pockets. Furthermore, the piperidine ring of TBS does not show any tubulin-mediated interligand NOEs, which confirms that this ring is not in contact with tubulin. The existence of tubulin-mediated interligand NOEs between EpoA and TBS could imply that the taxane binding pocket, to date identified as an exclusive binding site for MT-stabilizing agents, accommodates MT-destabilizing agents as well. This in turn poses the fundamental question about which structural features make a tubulin ligand either a MT-stabilizing or a MT-destabilizing agent.

Another possible explanation for the observation of interligand NOEs between EpoA and TBS would be the competitive binding of the two drugs to a pocket different from the taxane one. Recent work by Diaz et al.[14] has identified a second binding site for MT-stabilizing agents, the cyclostreptin binding site, which is adjacent to the taxane binding site on the other side of the M-loop and includes residues F214, T220, T221, and P222 of the H6–H7 loop. This binding site was suggested to be partially present in dimeric tubulin as well, and is proposed to represent the entry gate for MT-stabilizing agents to the lumenal taxane binding pocket in MTs. Thus, it is conceivable that EpoA and TBS share the cyclostreptin binding pocket in nonpolymerized tubulin.

Recently, the binding mode to tubulin of soblidotin, a dolastatin-10 analogue lacking the thiazole ring, has been determined by X-ray crystallography using αβ-tubulin in complex with the RB3 stathmin-like domain.[15] An overlap of the tubulin-bound conformation of soblidotin and tubulysin is shown in Figure S5 in the Supporting Information. There is a remarkable overlap between the pharmacophore of the two drugs, both in the overall shape at the C termini of the peptides and in the position of both the aromatic and polar groups. In this study, the binding site of soblidotin to tubulin partially overlaps with that of vinblastine, thus contradicting previous reports on the noncompetitive nature of the inhibition of vinblastine binding by dolastatin-10.[6] Notwithstanding this contradiction, we note that the dolastatin-10 binding site identified in the crystallographic study[15] marginally overlaps with the cyclostreptin binding site,[14] especially in the involvement of the H6–H7 loop. Assuming that dolastatin-10 and tubulysin share a common binding site on tubulin, the results from our INPHARMA NOE experiments might indicate that EpoA and tubulysin both bind to soluble tubulin at a site close to the cyclostreptin binding pocket.

Further investigations are in progress to confirm or disprove the competitive binding of tubulysin and EpoA, either to the taxane or to the cyclostreptin binding pockets of tubulin.

### Experimental Section

Tubulin preparation: Bovine brain tubulin was purchased from Cytoskeleton Inc. (Denver, CO, USA; Product No. T238) and was prepared for the NMR measurements as described previously.[16]

Sample preparation: TBS stock solution was prepared by dissolving TBS in [$D_6$]DMSO (25 μL), and then mixed with tubulin solution to a final volume of 280 μL. The final sample contained 5 % [$D_6$]DMSO, 10 μM tubulin, and 500 μM TBS. The presence of DMSO increased the solubility of TBS. Tubulin-free NMR samples of TBS were prepared by diluting the [$D_6$]DMSO stock solution of TBS in the NMR buffer.

NMR spectroscopy: NMR experiments were performed on a Bruker 900 MHz spectrometer. Resonances of TBS were assigned from COSY, TOCSY, HSQC, and HMBC spectra. A series of NOESY experiments was recorded at 25 °C with mixing times of 40, 70, 100, and 150 ms on tubulin–TBS and tubulin-free TBS samples. NMR data were processed with NMRPipe[17] and analyzed with FELIX (Accelrys Software Inc., CA, USA); 221 cross-peaks were identified in the NOESY spectra. Build-up curves were extracted for each cross-peak and 197 nonredundant NOEs were used as restraints in the structure calculation.

Structure calculation: 120 structures were calculated with XPLOR-NIH 2.13[18] using restrained simulated annealing (SA) from a single extended starting template. NOEs were used in the full relaxation matrix approach. From the initial rates of NOE build-up curves of those proton pairs the distance of which was independent of the conformation, a value of approximately 9 ns was estimated for the effective correlation time ($\tau_{c,eff}$), which correlated well with the result of the XPLOR-NIH grid search routine. The SA protocol comprised one high-temperature phase (2000 K), two cooling phases (from 2000 to 1000 K in steps of 50 K and from 1000 to 100 K), and a final 200 steps of energy minimization. The same weight was applied to all experimental peak intensities. Since stereospecific assignment of the methylene groups was missing, the in-house XPLOR function SWAP was used for optimizing the energy. Protons in methyl groups were averaged as $\langle r^{-3} \rangle^{-1/3}$. The quality of the fit of the structures to the experimental NOE data was determined with the generalized $R$-factor $R^n$, with $n = 1/6$. To test the consistency of the experimental data, the NOE information from 10 and 15 % of the data was completely omitted from the experimental restraint list. Calculations run with the reduced sets of NOE data converged consistently to the same result. All our tests proved that the tubulin-bound conformation of TBS was exclusively determined by the experimental NMR data.

Structures have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu/) under accession number BMRB-20121.

Interligand tr-NOEs: Protein-mediated interligand NOEs between EpoA and TBS were observed in a NOESY spectrum acquired at 900 MHz by a sample containing tubulin (10 μM), EpoA (500 μM), and TBS (500 μM) in a $D_2O$/[$D_6$]DMSO (95:5) solution. The mixing times of the NOESY spectra were 450, 275, 200, 150, and 100 ms.

# Communications

[1] F. Sasse, H. Steinmetz, J. Heil, G. Höfle, *J. Antibiot.* **2000**, 53, 879–885.

[2] M. W. Khalil, F. Sasse, H. Lunsdorf, Y. A. Elnakady, *ChemBioChem* **2006**, 7, 678–683.

[3] H. Steinmetz, N. Glaser, E. Herdtweck, F. Sasse, H. Reichenbach, G. Höfle, *Angew. Chem.* **2004**, *116*, 4996–5000; *Angew. Chem. Int. Ed.* **2004**, *43*, 4888–4892.

[4] R. B. Ravelli, B. Gigant, P. A. Curmi, I. Jourdain, S. Lachkar, A. Sobel, M. Knossow, *Nature* **2004**, *428*, 198–202.

[5] B. Gigant, C. Wang, R. B. Ravelli, F. Roussi, M. O. Steinmetz, P. A. Curmi, A. Sobel, M. Knossow, *Nature* **2005**, *435*, 519–522.

[6] a) R. Bai, G. R. Pettit, E. Hamel, *J. Biol. Chem.* **1990**, *265*, 17141–17149; b) R. Bai, M. C. Roach, S. K. Jayaram, J. Barkoczy, G. R. Pettit, R. F. Luduena, E. Hamel, *Biochem. Pharmacol.* **1993**, *45*, 1503–1515; c) R. Bai, N. A. Durso, D. L. Sackett, E. Hamel, *Biochemistry* **1999**, *38*, 14302–14310.

[7] E. Nogales, S. G. Wolf, K. H. Downing, *Nature* **1998**, *393*, 191.

[8] J. H. Nettles, H. L. Li, B. Cornett, J. M. Krahn, J. P. Snyder, K. H. Downing, *Science* **2004**, *305*, 866–869.

[9] T. Carlomagno, M. J. J. Blommers, J. Meiler, W. Jahnke, T. Schupp, F. Petersen, D. Schinzer, K. H. Altmann, C. Griesinger, *Angew. Chem.* **2003**, *115*, 2619–2621; *Angew. Chem. Int. Ed.* **2003**, *42*, 2515–2517.

[10] Z. Wang, P. A. McPherson, B. S. Raccor, R. Balachandran, G. Zhu, B. W. Day, A. Vogt, P. Wipf, *Chem. Biol. Drug Des.* **2007**, *70*, 75–86.

[11] A. W. Patterson, H. M. Peltier, F. Sasse, J. A. Ellman, *Chem. Eur. J.* **2007**, *13*, 9534–9541.

[12] V. M. Sánchez-Pedregal, M. Reese, J. Meiler, M. J. J. Blommers, C. Griesinger, T. Carlomagno, *Angew. Chem.* **2005**, *117*, 4244–4247; *Angew. Chem. Int. Ed.* **2005**, *44*, 4172–4175.

[13] M. Reese, V. M. Sánchez-Pedregal, K. Kubicek, J. Meiler, M. J. J. Blommers, C. Griesinger, T. Carlomagno, *Angew. Chem.* **2007**, *119*, 1896–1900; *Angew. Chem. Int. Ed.* **2007**, *46*, 1864–1868.

[14] a) R. M. Buey, E. Calvo, I. Barasoain, O. Pineda, M. C. Edler, R. Matesanz, G. Cerezo, C. D. Vanderwal, B. W. Day, E. J. Sorensen, J. A. López, J. M. Andreu, E. Hamel, J. F. Díaz, *Nat. Chem. Biol.* **2007**, *3*, 117–125; b) J. F. Diaz, I. Barasoain, A. A. Souto, F. Amat-Guerri, J. M. Andreu, *J. Biol. Chem.* **2005**, *280*, 3928–3937.

[15] A. Cormier, M. Marchand, R. B. G. Ravelli, M. Knossow, B. Gigant, *EMBO Rep.* **2008**, *9*, 1101–1106.

[16] V. M. Sánchez-Pedregal, K. Kubicek, J. Meiler, I. Lyothier, I. Paterson, T. Carlomagno, *Angew. Chem.* **2006**, *118*, 7548–7554; *Angew. Chem. Int. Ed.* **2006**, *45*, 7388–7394.

[17] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277–293.

[18] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *J. Magn. Reson.* **2003**, *160*, 65–73.

## 5.2   Unusual Interstrand Pt(S,S-diaminocyclohexane)-GG Crosslink

Following publication summarizes our results obtained during the study of an unusual interstrand Pt–GG crosslink that is formed within a DNA duplex by spontaneous isomerization of a classical intrastrand crosslink bearing two adjacent guanines coordinated to $Pt(S,S\text{-DACH})^{2+}$

# COMMUNICATION

# Unusual Interstrand Pt(*S*,*S*-diaminocyclohexane)-GG Crosslink Formed by Rearrangement of a Classical Intrastrand Crosslink Within a DNA Duplex

**Karel Kubíček,[a] Jordan Monnet,[a, b] Simone Scintilla,[c] Jana Kopečná,[d] Fabio Arnesano,[c] Lukáš Trantírek,[d] Claude Chopard,[b] Giovanni Natile,[c] and Jiří Kozelka\*[a, b]**

Currently used platinum antitumor drugs, such as cisplatin ($cis$-[PtCl$_2$(NH$_3$)$_2$]) and oxaliplatin ([Pt(oxalate)($R,R$-DACH)]) (DACH = 1,2-diaminocyclohexane), are believed to trigger apoptotic cell death by a cascade of cellular events induced by an initial reaction with DNA.[1,2] This reaction yields predominantly covalent crosslinks between adjacent guanine bases, which produce a bend in the DNA double-helix.[3] The crosslinks are generally inert and thus persistent; nevertheless some examples have been identified where an initially formed crosslink rearranges to another lesion.[4] We report here a remarkable case of such a rearrangement where the initial duplex d(C$_1$C$_2$T$_3$T$_4$G*$_5$G*$_6$T$_7$C$_8$T$_9$C$_{10}$)-d(G$_{11}$A$_{12}$G$_{13}$A$_{14}$C$_{15}$C$_{16}$A$_{17}$A$_{18}$G$_{19}$G$_{20}$) (**1**) (G* bases crosslinked by Pt($S,S$-DACH)$^{2+}$, the enantiomer of Pt($R,R$-DACH)$^{2+}$ occurring in oxaliplatin) rearranges to the interstrand-crosslinked species d(C$_1$C$_2$T$_3$T$_4$G$_5$G*$_6$T$_7$C$_8$T$_9$C$_{10}$)-d(G*$_{11}$A$_{12}$G$_{13}$A$_{14}$C$_{15}$C$_{16}$A$_{17}$A$_{18}$G$_{19}$G$_{20}$) (**2**). **2** has an unprece-

[a] Dr. K. Kubíček, J. Monnet, Prof. J. Kozelka
Masaryk University
Department of Condensed Matter Physics
Kotlářská 2, 611 37 Brno (Czech Republic)

[b] J. Monnet, Dr. C. Chopard, Prof. J. Kozelka
Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques
UMR 8601 CNRS, Université Paris Descartes
45, rue des Saints-Pères
75270 Paris 06 (France)
Fax: (+331) 42-86-83-87
E-mail: kozelka.jiri@gmail.com

[c] Dr. S. Scintilla, Dr. F. Arnesano, Prof. G. Natile
Dipartimento Farmaco-Chimico
Università di Bari «Aldo Moro»
via E. Orabona 4
70125 Bari (Italy)

[d] J. Kopečná, Dr. L. Trantírek
Biology centre AS CR
v.v.i. & University of South Bohemia
Branišovská 31. 370 05 České Budějovice (Czech Republic)

💻 Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/asia.200900655.

dented structure with a head-to-tail orientation of the platinated bases G*$_6$ and G*$_{11}$ and a side-on interaction between G*$_{11}$ and the methyl group of T$_7$. This intrastrand-to-interstrand rearrangement is a new demonstration of the kinetic instability of Pt−N bonds in strained structures.

The initial aim of our work was to evaluate the impact of the ligand chirality {Pt($R,R$-DACH)$^{2+}$ versus Pt($S,S$-DACH)$^{2+}$} on the structure of the intrastrand-crosslinked duplex **1**. The $S,S$ enantiomer, as compared to the $R,R$ enantiomer, has lower antitumor and much greater mutagenic activities.[5] **1** was obtained by reaction of [Pt(NO$_3$)$_2$($S,S$-DACH)] with the d(C$_1$C$_2$T$_3$T$_4$G$_5$G$_6$T$_7$C$_8$T$_9$C$_{10}$) single-strand and hybridization of the purified G*$_5$G*$_6$-platinated single-strand with the complementary strand. In the course of the NMR investigation, we observed the appearance of a second set of NMR signals belonging to a new species (**2**) which formed at the expense of **1** and achieved, within one week, a concentration comparable to that of **1**, as judged from the cytosine H5–H6 crosspeak volumes. Electrophoretic studies (see Experimental Section) indicated that the mixture contained an interstrand-crosslinked species, and this conclusion was supported by the observation that, upon heating to 40 °C, the peaks of **1** disappear completely (as expected for an intrastrand-crosslinked duplex of this size),[6] whereas those of **2** persist indicating that it is a duplex stabilized by covalent bonds. In the course of one year, the fraction of **2** increased to virtually 100 %.

In spite of the structural perturbation caused to the DNA double-helix by this unusual interstrand crosslink (vide infra), the sequential connectivities among sugar and base protons, typical of B-DNA, are largely conserved in **2**, enabling an assignment based on standard methods (Table S1 in the Supporting Information).[7] A 500 ms NOESY spectrum of an approximately 1:1 mixture of **1** and **2** is shown in Figure 1. The spectrum of the intrastrand-crosslinked species **1** is characterized by the typical low-field signal of the H8 proton of the platinum-bound 5′-guanine, G*$_5$, around 8.7 ppm, which has a crosspeak with the H8 of the other

Figure 1. 500 ms 2D $^1$H,$^1$H NOESY spectrum of an approx. 1:1 mixture of **1** and **2**, total concentration ~2.6 mM in 20 mM phosphate buffer pH 7.2 in D$_2$O at 288 K. NOE crosspeaks of the most upfield signal (T$_7$CH$_3$ of **2**) are labeled in black, those of the most downfield signal (G$^*_5$H8 of **1**) are labeled in red.

platinated guanine, G$^*_6$ (Figure 1). The most striking NMR-spectroscopic feature of **2** is the extremely upfield-shifted methyl signal of T$_7$ (0.5 ppm), which shows NOE crosspeaks with a number of other protons, including the H8 protons of G$^*_6$ and G$^*_{11}$ (Figure 1, top). The G$^*_6$-T$_7$ and T$_7$-G$^*_{11}$ contacts indicate that G$^*_6$ and G$^*_{11}$ are close in space and T$_7$ inserts between them. NOE crosspeaks connecting the DACH residue with both G$^*_6$ and G$^*_{11}$ strongly support the proximity between bases G$^*_6$ and G$^*_{11}$ arising from a platinum crosslink. A relatively intense intranucleotide H1′-H8 crosspeak indicates that G$^*_{11}$ has a *syn* conformation. The absence of an H8–H8 NOE between G$^*_6$ and G$^*_{11}$ excludes a head-to-head orientation of the two bases. For the base-pairs G$_5$/C$_{16}$ and G$^*_6$/C$_{15}$, we observed relatively strong intra-base-pair H1–H41 NOEs, and also rather unusual sequential G$_5$H8-G$^*_6$H8, G$_5$H1′-G$^*_6$H1′, C$_{15}$H6-C$_{16}$H6, and C$_{15}$H2″-C$_{16}$H2′ NOEs. Together, these observations indicate that these two base-pairs lie in adjacent parallel planes, have conserved Watson–Crick (WC) base-pairing, and the step is considerably unwound.

GH1–CH41 imino–amino contacts and TH3–AH2 contacts were observed across all base-pairs except for the terminal base-pairs and T9A12, suggesting that, within the inner part of the duplex, the complementary bases are kept at short distances, even though their WC base-pairing may be perturbed.

We have used NOE volumes extracted from 50 ms NOESY spectra as distance constraints for a molecular dynamics simulation performed on the solvated duplex **2**. The final structure is in excellent agreement with the NOESY-derived distances and also explains the atypical chemical shifts of several protons. The global view of the refined structure is shown in Figure 2a. Figure 2b highlights the position of the T$_7$ base inserted between G$^*_6$ and G$^*_{11}$ and shows the protons that the NOESY spectra indicate to have contacts with the T$_7$ methyl group. The base pairs from C$_1$/G$_{20}$ to T$_7$/A$_{14}$ are largely unperturbed; the base-pair C$_8$/G$_{13}$ opens up, whereas the terminal base-pairs T$_9$/A$_{12}$ and C$_{10}$/G$^*_{11}$ are completely disrupted. This is caused by the platinum-coordination of G$^*_{11}$ and its side-on orientation with respect to T$_7$. This conformation appears to be stabilized by van der Waals contacts between the methyl group of T$_7$ and the six-membered ring of G$^*_{11}$. This interaction apparently "pushes" the T$_7$ base above the ring of G$^*_6$ with both bases (G$^*_{11}$ and G$^*_6$) strongly shielding the T$_7$ methyl group. As a result, the G$^*_6$-T$_7$ step is overwound (twist angle of 40°), as also suggested by the clear G$^*_6$H1–T$_7$H3 NOE.

Figure 2a makes it clear that the observed interstrand crosslink requires the initially unplatinated strand to make a U-turn at the 5′-terminus (bases G$^*_{11}$-A$_{14}$). The formation of such a loop is obviously possible only since G$^*_{11}$ is the 5′-terminal base of the initially unplatinated strand. Thus, a 5′-guanine in one strand of a DNA duplex, and an interior guanine within the other strand at an appropriate distance, are apparently the requirements for the formation of the type of interstrand crosslink seen in **2**.

The platinum coordination of G$^*_6$ does not perturb significantly the WC pairing with C$_{15}$. In contrast, the crosslinking of G$^*_{11}$ with G$^*_6$ disrupts the base-pairing of G$^*_{11}$ with C$_{10}$ and also that of A$_{12}$ with T$_9$. The A$_{12}$ base inserts its amino group between T$_9$ and C$_{10}$ so establishing short contacts with T$_9$O2, C$_{10}$O2, and C$_{10}$O4′ atoms. The positions of T$_9$ and C$_{10}$ nucleotides at the 3′-terminus of the top strand must be relatively disordered, as indicated by the absence of the H3-methyl crosspeak for T$_9$ and of the H41–H42 crosspeak for C$_{10}$ in the NOESY spectrum in 90% H$_2$O at 15°C. The C$_1$–C$_{10}$ strand has a fairly conserved single-helical B-DNA form, and the nH1′-(n+1)H6/8, nH2″-(n+1)H6/8, and nH3′-(n+1)H6/8 connectivities are seen in the 350 ms NOESY spectrum (used for assignment) for the whole chain except for the T$_9$–C$_{10}$ step for which only the nH3′-(n+1)H6 peak is present. Because of the unwinding of the G$_5$–G$^*_6$ step, the G$^*_6$H8 proton is located exactly above the five-membered ring of G$_5$ (Figure 2b) which explains its upfield resonance (7.5 ppm) in spite of the deshielding effect arising from platinum coordination.

The formation of **2** is highly informative. First, the isomerization from **1** to **2** underlines the thermal mobility of the terminal base-pairs in DNA duplexes, allowing G$^*_{11}$ to loop back on the duplex and to carry out a nucleophilic attack on the Pt atom. Second, the slow rearrangement of **1** to **2** shows that Pt−N(purine) bonds are not inert, and the initially formed Pt–DNA crosslink can rearrange to a thermody-
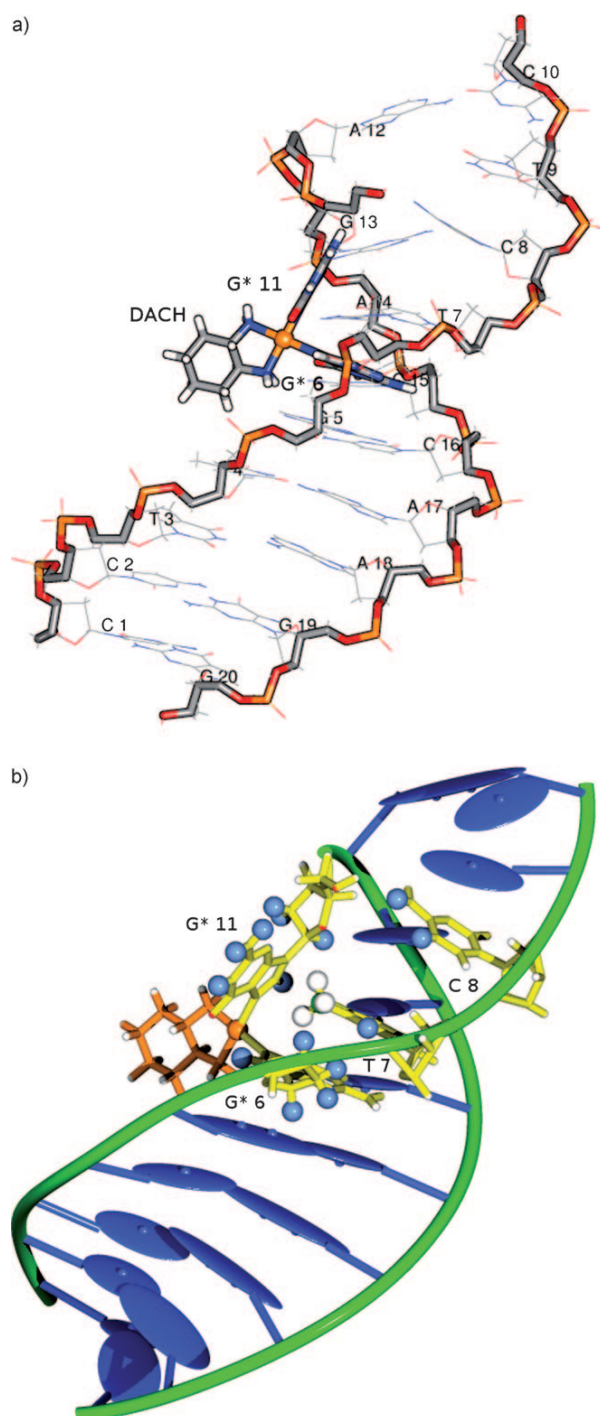
Figure 2. Molecular model of **2** selected from a cluster analysis of an MD trajectory under NOESY-derived distance constraints. a) Global view in stick representation, with the $[Pt(S,S\text{-}DACH)(Gua)_2]^{2+}$ complex and the backbone of each strand highlighted. b) Model representation highlighting the position of the $T_7$ methyl group (H atoms as white spheres) between the platinum-bound guanines $G^*_6$ and $G^*_{11}$, and showing the protons displaying NOEs with the $T_7$ methyl group (blue spheres).

namically more stable product.[4] Third, the apparent greater stability of **2** is interesting. Although we have not yet performed a modeling study on **1**, its NMR features are indicative of a classical conformation with head-to-head arrangement of the platinated guanines and bending towards the major groove.[8] In this latter form, all WC base-pairs are conserved and the major energetic loss arises from the destacking at the platination site. The geometry of **2** offers an interesting alternative, with a head-to-tail conformation (which is intrinsically more stable than the head-to-head conformation of the $[Pt(S,S\text{-}DACH)(Gua)_2]^{2+}$ complex[9]) and a more compact structure of the duplex, where the loss of some WC and stacking interactions is compensated by non-canonical interactions such as the side-on interaction between $G^*_{11}$ and $T_7$. The insertion of the methyl group of $T_7$ between the two platinated bases (Figure 2b) is reminiscent of the intercalation of a hydrophobic peptide residue between the two destacked bases seen in adducts between a recognition protein and a bent DNA.[10] Overall, the structure of **2** provides an impressive demonstration of the plethora of possibilities DNA has for stabilizing non-canonical structures.

Malina et al. reported differential scanning calorimetry experiments on DNA duplexes bearing GG–Pt intrastrand crosslinks with $cis$-$Pt(NH_3)_2^{2+}$, $Pt(R,R\text{-}DACH)^{2+}$, and Pt-$(S,S\text{-}DACH)^{2+}$ residues and found that, within a TG*G*T sequence, the GG–$Pt(S,S\text{-}DACH)^{2+}$ crosslink destabilizes the DNA duplex considerably more than the GG–$Pt(R,R\text{-}DACH)^{2+}$ and GG-$cis$-$Pt(NH_3)_2^{2+}$ crosslinks.[5] It is therefore possible that the $S,S$-chirality of the $Pt(DACH)^{2+}$ residue drives the isomerization of **1** into **2**. Experiments are underway to investigate whether an intrastrand-to-interstrand conversion occurs also with the clinically used Pt-$(R,R\text{-}DACH)^{2+}$ enantiomer.

Yang et al. already reported an isomerization of the octamer $d(C_1C_2T_3G^*_4G^*_5T_6C_7C_8)$-$d(G_9G_{10}A_{11}C_{12}C_{13}A_{14}G_{15}G_{16})$, containing two adjacent guanines (G*) crosslinked with $cis$-$Pt(NH_3)_2^{2+}$, to yield the interstrand-crosslinked species $d(C_1C_2T_3G^*_4G_5T_6C_7C_8)$-$d(G^*_9G_{10}A_{11}C_{12}C_{13}A_{14}G_{15}G_{16})$ **(3)**. The crosslinking of $G^*_4$ with $G^*_9$ was deduced from a sequential assignment featuring coincident and extremely upfield-shifted (6.83 ppm) $G^*_4H8$ and $G_5H8$ resonances. The authors proposed that the $G^*_5(3')$ coordination is more labile than the $G^*_4(5')$ coordination which appears more flexible in molecular models.[11] Moreover it was concluded that the $G^*_4G^*_9$ $cis$-$[Pt(NH_3)_2(Gua)_2]^{2+}$ moiety has head-to-head conformation of the two guanines, but no structural model was proposed. In contrast, **2** is formed upon release of the $G^*_5(5')$ guanine and our NMR-derived structural model is compatible only with a head-to-tail orientation of the platinated bases (vide supra).

In conclusion, we report here an unusual interstrand Pt–GG crosslink (**2**) that forms within a DNA duplex by spontaneous isomerization of a classical intrastrand crosslink bearing two adjacent guanines coordinated to Pt($S,S$-DACH)$^{2+}$ (**1**). In the observed rearrangement, the Pt-N7(5'-G) bond is broken and replaced by a Pt−N7 bond with the

5′-terminal guanine of the complementary strand. The possibility that an intrastrand Pt–GG crosslink in a DNA duplex is attacked by a 5′-terminal guanine of the unplatinated strand should be born in mind in studies of platinum adducts with oligonucleotides. The thermodynamically more stable **2** has a head-to-tail orientation of the cross-linked guanines, a more compact structure, and the loss of some base-pairing is compensated by non-canonical interactions such as a side-on interaction between guanine and thymine.

## Experimental Section

Preparation of **1**: Equimolar quantities (6 μmol) of [Pt(NO$_3$)$_2$($S,S$-DACH)] and d(CCTTGGTCTC) were stirred at 25 °C for 40 h; after 20 h another 1.2 μmol of the platinum complex was added. The single-stranded GG-adduct was purified by HPLC using a reversed-phase semi-preparative column. Hybridization with the complementary strand was followed by NMR, monitoring the disappearance of the 9.35 ppm peak characteristic of the 3′-G* of the platinated single-strand. The platinated duplex was desalted using a 3 K Centricon filter, lyophilized, and finally dissolved in 20 mM phosphate buffer at an estimated final concentration of 2.6 mM, pH 7.2, in D$_2$O, or in a 90:10 H$_2$O/D$_2$O solution for the NMR lock.

Molecular dynamics (MD) simulations: An initial structure of **2** was constructed using HGS molecular models (Maruzen International Co.), crosslinking G*$_6$ and G*$_{11}$ with Pt($S,S$-DACH)$^{2+}$ and modeling a cavity around the T$_7$ methyl group according to the NOESY-derived H–H distances. An AMBER topology file was constructed using the LEAP routine of the AMBER suite, and a pdb file with coordinates corresponding to the desk model was obtained from a B-DNA model bearing an intrastrand G*$_5$-G*$_6$ crosslink using molecular dynamics simulations with imposed G*$_6$-G*$_{11}$ crosslinking and long-range atom–atom distances constraints measured on the desk model. Then, the long-range constraints were gradually replaced by the experimental H−H distance constraints obtained from 50 ms NOESY spectra in D$_2$O and H$_2$O solutions. This procedure was considered safer than trying to obtain a starting structure from distance geometry calculations using the (intrinsically short-range) NOE distances.

Details of materials used, HPLC, PAGE, NMR, and MD simulations are given in the Supporting Information.

[1] J. Reedijk, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3611–3616.
[2] P. J. Sadler, *ChemBioChem* **2009**, *10*, 73–74.
[3] J. R. Jamieson, S. J. Lippard, *Chem. Rev.* **1999**, *99*, 2467–2497.
[4] V. Brabec in *Platinum-Based Drugs in Cancer Therapy* (Eds.: L. R. Kelland, N. P. Farrell), Humana Press, Totowa, NJ, **2000**, pp. 37–61.
[5] J. Malina, O. Novakova, M. Vojtiskova, G. Natile, V. Brabec, *Biophys. J.* **2007**, *93*, 3950–3962.
[6] F. Herman, J. Kozelka, V. Stoven, E. Guittet, J.-P. Girault, T. Huynh-Dinh, J. Igolen, J.-Y. Lallemand, J.-C. Chottard, *Eur. J. Biochem.* **1990**, *194*, 119–133.
[7] S. S. Wijmenga, B. N. M. van Buuren, *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287–387.
[8] S. O. Ano, Z. Kuklenyik, L. G. Marzilli in *Cisplatin: Chemistry and Biochemistry of a Leading Anticancer Drug* (Ed.: B. Lippert), Verlag Helvetica Chimica Acta, Zürich, **1999**, pp. 247–291.
[9] L. G. Marzilli, S. O. Ano, F. P. Intini, G. Natile, *J. Am. Chem. Soc.* **1999**, *121*, 9133–9142.
[10] J. O. Thomas, A. A. Travers, *Trends Biochem. Sci.* **2001**, *26*, 167–174.
[11] D. Yang, S. S. G. E. van Boom, J. Reedijk, J. H. van Boom, A. H.-J. Wang, *Biochemistry* **1995**, *34*, 12912–12920.

## 5.3    Recognition of Transcription Termination Signal by NAB3

In this publication we have studied structure of protein-RNA complex and have shown that the RRM of Nab3 binds specifically the YCU sequence (where Y stands for pyrimidine). This was/is in a good agreement with previous functional data that led to the proposal of the UCUU sequence as the Nab3 termination element.

# Recognition of Transcription Termination Signal by the Nuclear Polyadenylated RNA-binding (NAB) 3 Protein[*]

**Fruzsina Hobor**[‡1,2], **Roberto Pergoli**[‡1,3], **Karel Kubicek**[‡], **Dominika Hrossova**[‡], **Veronika Bacikova**[‡],
**Michal Zimmermann**[§2,4], **Josef Pasulka**[‡], **Ctirad Hofr**[§], **Stepanka Vanacova**[‡], and **Richard Stefl**[‡5]

*From the* [‡]*National Centre for Biomolecular Research and* [§]*Department of Functional Genomics and Proteomics, Central European Institute of Technology, Faculty of Science, Masaryk University, Brno CZ-62500, Czechia*

Non-coding RNA polymerase II transcripts are processed by the poly(A)-independent termination pathway that requires the Nrd1 complex. The Nrd1 complex includes two RNA-binding proteins, the nuclear polyadenylated RNA-binding (Nab) 3 and the nuclear pre-mRNA down-regulation (Nrd) 1 that bind their specific termination elements. Here we report the solution structure of the RNA-recognition motif (RRM) of Nab3 in complex with a UCUU oligonucleotide, representing the Nab3 termination element. The structure shows that the first three nucleotides of UCUU are accommodated on the β-sheet surface of Nab3 RRM, but reveals a sequence-specific recognition only for the central cytidine and uridine. The specific contacts we identified are important for binding affinity *in vitro* as well as for yeast viability. Furthermore, we show that both RNA-binding motifs of Nab3 and Nrd1 alone bind their termination elements with a weak affinity. Interestingly, when Nab3 and Nrd1 form a heterodimer, the affinity to RNA is significantly increased due to the cooperative binding. These findings are in accordance with the model of their function in the poly(A) independent termination, in which binding to the combined and/or repetitive termination elements elicits efficient termination.

RNA Polymerase II (RNA Pol II)[6] transcribes messenger RNA (mRNA), but also a subset of small nuclear and small nucleolar RNAs (snRNAs/snoRNAs), micro-RNA precursors, and a class of intergenic and antisense RNAs (1). RNA Pol II uses two different mechanisms for transcription termination of these "coding" and "non-coding" RNAs. Although the RNA Pol II termination of mRNA requires a large multiprotein complex that recognizes the poly(A) signal in the nascent transcript (2), the termination of the non-coding RNAs requires no poly(A) signal (2–4).

In the poly(A)-independent mechanism, transcription termination requires a specific factor, the Nrd1 complex. This complex consists of three proteins: the nuclear pre-mRNA down-regulation (Nrd) 1 protein, the nuclear polyadenylated RNA-binding (Nab) 3 protein, and the putative RNA helicase Sen1 (5–7). The Nrd1 complex interacts with the exosome, a complex of 10–12 exoribonucleolytic and RNA-binding proteins (8) and the Trf4-Air2-Mtr4 polyadenylation (TRAMP) complex (9–11), which are involved in the 3' end processing of non-coding RNA transcripts (3, 4, 7).

In yeast, transcription termination mediated by the Nrd1 complex requires binding to both the nascent RNA and the carboxyl-terminal domain of RNA Pol II, which consists of 26 repeats of the sequence Tyr[1]-Ser[2]-Pro[3]-Thr[4]-Ser[5]-Pro[6]-Ser[7] (1, 12). Interestingly, the Nrd1 complex binds the carboxyl-terminal domain when it is phosphorylated at Ser[5], a typical feature of the early elongation phase of the transcription cycle. The Ser[5]-phosphorylated carboxyl-terminal domain is recognized by the carboxyl-terminal domain-interacting domain of Nrd1 (13, 14). The RNA-binding subunits of the Nrd1 complex, Nrd1 and Nab3, recognize their specific RNA sequences (called terminator elements) in the nascent transcripts of RNA Pol II. It is believed that this specific binding of Nrd1 complex to the terminator elements is the initial step in the assembly of termination machinery.

A number of studies narrowed the sequence regions with terminator elements (5, 6, 15–17) that were subsequently identified as GUAR (where R stands for purine) and UCUU sequences (18). GUAR and UCUU terminator elements are recognized by Nrd1 and Nab3, respectively, via their fragments encompassing RNA recognition motifs (RRMs) (18). These terminator sequences are located downstream of snRNA and snoRNA genes (18) although their relative orientation and spacing are not highly conserved. In addition, it was demonstrated that Nrd1 and Nab3 form a stable het-

erodimer and bind to snoRNA terminators that contain multiple Nrd1- and Nab3-binding sequences (19).

Both Nrd1 and Nab3 contain RRM that likely mediates the binding to their specific RNA sequences. The RRM is the most abundant RNA-binding domain in higher vertebrates; *e.g.* the RRM is present in about 2% of human genes (20). It is a small protein domain of ~90 amino acids with a typical $\beta\alpha\beta\beta\alpha\beta$ topology that forms a four-stranded $\beta$-sheet packed against two $\alpha$-helices (21–23). The structure of this domain is relatively well defined despite a little sequence conservation among various RRMs. The solved structures of RRM bound to RNA show the complexity of protein-RNA recognition mediated by the RRM, which often involves not only RRM-RNA interactions but also RRM-RRM and other RRM-protein interactions. The main protein surface of the RRM involved in the interaction with the RNA is the four-to-five-stranded $\beta$-sheet, which typically contacts two or three nucleotides. Frequently, RRM-containing proteins bind more than three nucleotides and recognize longer single-stranded RNA or even internal RNA loops by employing of $\beta$-strand loops and N- or C-terminal flanking regions of RRMs (21–23).

To better understand the structural basis behind the poly(A) independent transcription termination pathway, we initiated an NMR study of *Saccharomyces cerevisiae* Nab3. Here, we present the three-dimensional solution structure of the Nab3 RRM in free form and in complex with the 5′-UCUU-3′ RNA substrate. The structure of the complex reveals recognition of the YCU sequence (where Y stands for pyrimidine) by the Nab3 RRM. We confirmed the sequence-specific intermolecular contacts by site-directed mutagenesis and fluorescence anisotropy (FA) measurements, and their physiological role was also confirmed by yeast phenotypic analyses. Finally, we demonstrate that the weak RNA binding of the isolated RRMs of Nab3 and Nrd1 is greatly enhanced when Nab3 and Nrd1 form a heterodimer and bind the RNA cooperatively.

## EXPERIMENTAL PROCEDURES

*Cloning, Expression, and Purification of Proteins*—The coding sequence corresponding to the RRM of the *Nab3* gene from *S. cerevisiae* (961–1245) was amplified by polymerase chain reaction (PCR), and cloned into a pET22b expression vector (Novagen) via NdeI and XhoI restriction sites. The resulting C-terminal His$_6$-tagged construct was verified by DNA sequencing. The protein was overexpressed in *Escherichia coli* BL21-Codon Plus (DE3)-RIPL (Stratagene), transformed with the pET22b-RRM Nab3 construct at 37 °C in M9 minimal medium, supplemented with 50 mg/liter of ampicillin. For isotope labeling, the medium was supplemented with $^{15}NH_4Cl$ and [U-$^{13}C_6$]glucose. Cells were grown at 37 °C to $A_{600}$ ~1 and induced with 1 mM isopropyl $\beta$-D-thiogalactoside. Cells were harvested by centrifugation (6000 × g for 10 min), resuspended in lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM $\beta$-mercaptoethanol, pH 8), and disrupted by sonication. The cell debris was cleared by centrifugation (14,000 × g for 60 min). Soluble lysate was loaded on a nickel-nitrilotriacetic acid column (Qiagen), equilibrated with lysis buffer, washed with a high salt buffer (50 mM sodium phosphate, 500 mM NaCl, 10 mM $\beta$-mercaptoethanol, 5 mM imidazole, pH 8), and eluted with imidazole gradient (50–500 mM) of elution buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM $\beta$-mercaptoethanol, pH 8). The protein was subsequently loaded on a Superdex 75 gel filtration column (GE Healthcare), equilibrated with lysis buffer. The protein fractions from gel filtration were dialyzed against lysis buffer. The purified protein was 99% pure, as judged by Coomassie-stained SDS-PAGE. For NMR measurements the pure protein was concentrated to 2.5 mM in 550 $\mu$l of 50 mM sodium phosphate (pH 8.0), containing 300 mM NaCl, and 10 mM $\beta$-mercaptoethanol. The cloning, expression, and purification of Nrd1 RRM-(340–410) were carried out in the same way as for Nab3 RRM.

The expression and purification of the Nrd1-Nab3 heterodimer have been done in a similar manner as reported previously (19). To improve the yield of expression, we used *E. coli* BL21-Codon Plus (DE3)-RIPL (Stratagene). We used the following final buffer (50 mM Tris (pH 8.0), containing 150 mM NaCl, and 10 mM $\beta$-mercaptoethanol) to have the same conditions for all fluorescence anisotropy measurements. Prior to RNA titration, all proteins were tested for the residual RNase activity using RNaseAlert Lab Test (Ambion). RNA oligonucleotides were purchased from Thermo Fisher Scientific/Dharmacon and Sigma.

*Generation of Nab3 RRM Mutants*—Site-specific mutagenesis was performed using the QuikChange site-directed mutagenesis kit (Stratagene) with complementary sense and antisense (AS) oligonucleotide primers as follows: R331A (S), 5′-gcacaatattcctccgaagtcagcattattcattggtaatttgccg-3′ and (AS), 5′-cggcaaattaccaatgaataatgctgacttcggaggaatattgtgc-3′; N361A (S), 5′-tccatacggtcatatcatgcaaatcgctatcaaaaatgcctttggattcatt-3′ and (AS), 5′-aatgaatccaaaggcattttttgatagcgatttgcatgatatgaccgt-atgga-3′; E397A (S), 5′-gcaaaaagttgatcctggcagtttctagctcgaatg-c-3′ and (AS), 5′-gcattcgagctagaaactgccaggatcaactttttgc-3′; E397K (S), 5′-tggcaaaaagttgatcctgaaagtttctagctcgaatgc-3′ and (AS) 5′-gcattcgagctagaaactttcaggatcaactttttgcca-3′; S399A (S), 5′-agttgatcctggaagttgctagctcgaatgctcgt-3′ and (AS), 5′-acgag-cattcgagctagcaacttccaggatcaact-3′; S399K (S), 5′-actttggcaaaa-agttgatcctggaagttaaaagctcgaatgctcgtcc-3′ and (AS), 5′-ggacga-gcattcgagcttttaacttccaggatcaactttttgccaaagt-3′. All mutations were verified by DNA sequence analysis.

*NMR Spectroscopy*—All NMR spectra of 2.5 mM uniformly $^{15}N,^{13}C$-labeled Nab3 RRM in 50 mM sodium phosphate buffer (pH 8.0), 300 mM NaCl, 10 mM $\beta$-mercaptoethanol (90% $H_2O$/10% $D_2O$) were recorded on Bruker AVANCE 600 and 900 MHz spectrometers equipped with a cryoprobe at a sample temperature of 30 °C. All spectra were processed with Topspin 2.1 (Bruker BioSpin) and analyzed with Sparky 3.0 (T. G. Goddard and D. G. Kneller, University of California, San Francisco). The $^1H$, $^{13}C$, and $^{15}N$ chemical shifts of Nab3 RRM were assigned as described previously (24). All distance restraints were derived from the three-dimensional $^{15}N$- and $^{13}C$-edited NOESYs and two-dimensional $^1H$-$^1H$-labeled NOESY (with mixing time of 150 ms) collected at 900 MHz spectrometer.

Akin to the free Nab3 RRM, the backbone resonance assignments of Nab3 RRM in the bound form were achieved

using three-dimensional triple resonance experiments: HNCA, HNCACB, and CBCA(CO)NH (24–26). Resonances of the aliphatic side chains were assigned by a combination of three-dimensional HCCH TOCSY, three-dimensional HNHA, and three-dimensional $^{13}$C-edited NOESY spectra. Resonances of the aromatic side chains were assigned using two-dimensional homonuclear NOESY, three-dimensional $^{13}$C$_{arom}$-edited NOESY, and two-dimensional (HB)CB(CGC-D)HD spectra. All distance restraints were derived from the three-dimensional $^{15}$N- and $^{13}$C-edited NOESYs (with mixing time of 150 ms) collected at 900 MHz spectrometer. The RNA resonances in complex were assigned using a combination of standard through-space and through-bond experiments (27, 28). We could not assign all sugar resonances unambiguously due to a high resonance overlap in the spectra (only unlabeled RNA was used). The sugar pucker conformation for all nucleotides is *C2′-endo*, as identified in the two-dimensional homonuclear TOCSY spectrum (strong cross-peaks between the H1′ and H2′ resonances). Intermolecular distance constraints were obtained from the three-dimensional $^{13}$C F$_1$-edited, F$_3$-filtered NOESY-HSQC experiment (29), which was recorded in H$_2$O (with WATERGATE water suppression) as a two-dimensional filter NOESY omitting $^{13}$C chemical shift evolution.

*Structure Calculations*—The preliminary structure determinations of the free and bound Nab3 RRM were performed with the automated NOE assignment module implemented in the CYANA program (30). This automated NOE assignment procedure is a re-implementation of the former CANDID algorithm (31) on the basis of a probabilistic treatment of the NOE assignment. CYANA carries out automated assignment and distance calibration of NOE intensities, removal of meaningless restraints, structure calculation with torsion angle dynamics, and automatic upper distance limit violation analysis. The resultant NOE cross-peak assignments were subsequently confirmed by visual inspection of the spectra. The predicted protein backbone $\phi$ and $\psi$ torsion angle from the chemical shifts (32) for the secondary structure elements were also included in the calculations. In the next step, CYANA-generated restraints along with manually assigned protein-RNA intermolecular restraints were used for further refinement of the preliminary structures with AMBER 10.0 software (33). This calculations employed a modified version (AMBER ff99SB) of the force field described by Cornell *et al.* (34) along with a refinement protocol described in Padrta *et al.* (35), and an explicit solvent. From 40 refined structures, the 20 conformers with the lowest AMBER energy were selected to form the final ensemble of structures. Structural quality was assessed using PROCHECK (36) and WHAT IF (37). Molecular graphics were generated using MOLMOL (38) and PyMOL (57).

*Fluorescence Anisotropy Measurements*—The equilibrium binding of Nab3 RRM to different oligonucleotides was analyzed by fluorescence anisotropy. The RNA oligonucleotides were either 5′-labeled with TAMRA or fluorescein attached via a hexyl linker. The measurements were conducted on a FluoroMax-4 spectrofluorometer (Horiba Jobin-Yvon Edison, NJ). The instrument was equipped with a thermostatted cell

holder with a Neslab RTE7 water bath (Thermo Scientific). The whole system was operated using FluorEssence software (version 2.5.3.0, Horiba Jobin-Yvon). The TAMRA fluorophore was excited at 561 nm and its emission was collected at 581 nm. The widths of both excitation and emission monochromatic slits were 8 nm and integration time was set to 3 s. The fluorescein fluorophore was excited at 488 nm and its emission was collected at 520 nm. For measurement with the individual domains (Nrd1-(340–410) and Nab3-(331–415)) the width of both excitation and emission monochromatic slits were 7 nm and the integration time was set to 3 s. 10 nM labeled oligonucleotide (volume 1.4 ml) was titrated with increasing amounts of the protein in 50 mM sodium phosphate buffer (pH 7.5), supplemented with 150 mM NaCl and 10 mM $\beta$-mercaptoethanol. For measurement with the heterodimer (Nab3-(191–565)–Nrd1-(1–548)), 1 nM fluorescein-labeled RNA was used. Both excitation and emission monochromatic slits were 14 nm, the integration time was set to 3 s.

In all measurements, an identical concentration of the oligonucleotide was included in the protein stock solution to prevent dilution of the RNA during titration. All experiments were carried out at 25 °C in a stirred 1.5-ml quartz cuvette. Protein aliquots were added stepwise until the cuvette was filled. After this point, a certain volume of the sample was always removed from the cuvette before addition of a protein aliquot of the same volume. A fixed delay of 30 s was set between each aliquot addition and start of the measurement to allow the reaction to reach equilibrium. This delay was sufficient, as no further change in anisotropy was observed. Each data point is an average of five measurements. Neither TAMRA nor fluorescein labels showed binding with any of the protein constructs.

The data were analyzed in SigmaPlot 11 software (Systat Software). The experimental isotherms were fit to a single-site binding model according to Heyduk and Lee (39) using nonlinear least squares regression. The data were normalized for visualization purposes.

*Yeast Strains and Plasmids*—The plasmid for the expression of wild-type Nab3 in yeast contain 550 bp of NAB3 promoter upstream of the AUG and a 286-bp sequence downstream of the stop codon. SV320 contains the wild-type NAB3 inserted in pRS415 (a LEU2 CEN plasmid) (40). Plasmids SV321-SV326 contain point mutants R331A, N361A, E397A, E397K, S399A, and S399K, respectively (see above for the primers). Strain DLY889 containing the endogenous NAB3 under control of the GAL1 promoter (3) was transformed with plasmids containing either wild-type Nab3 or Nab3 RRM point mutants with a LEU2 selectable marker (SV320, SV321, SV322, SV323, SV324, SV325, and SV326, respectively). The resulting strains were used for growth tests and Western blot analyses.

*Growth Test Analyses*—To test whether the mutated residues were essential for growth, the resulting transformants were grown in SC-LEU-HIS + 2% galactose at 30 °C to an $A_{600}$ 1.0. The cultures were then serially diluted in 96-well plates by a factor of 10, and spotted onto SC-LEU-HIS medium containing 2% glucose to repress the expression of the endogenous NAB3 or control medium (SC-HIS + 2% galac-
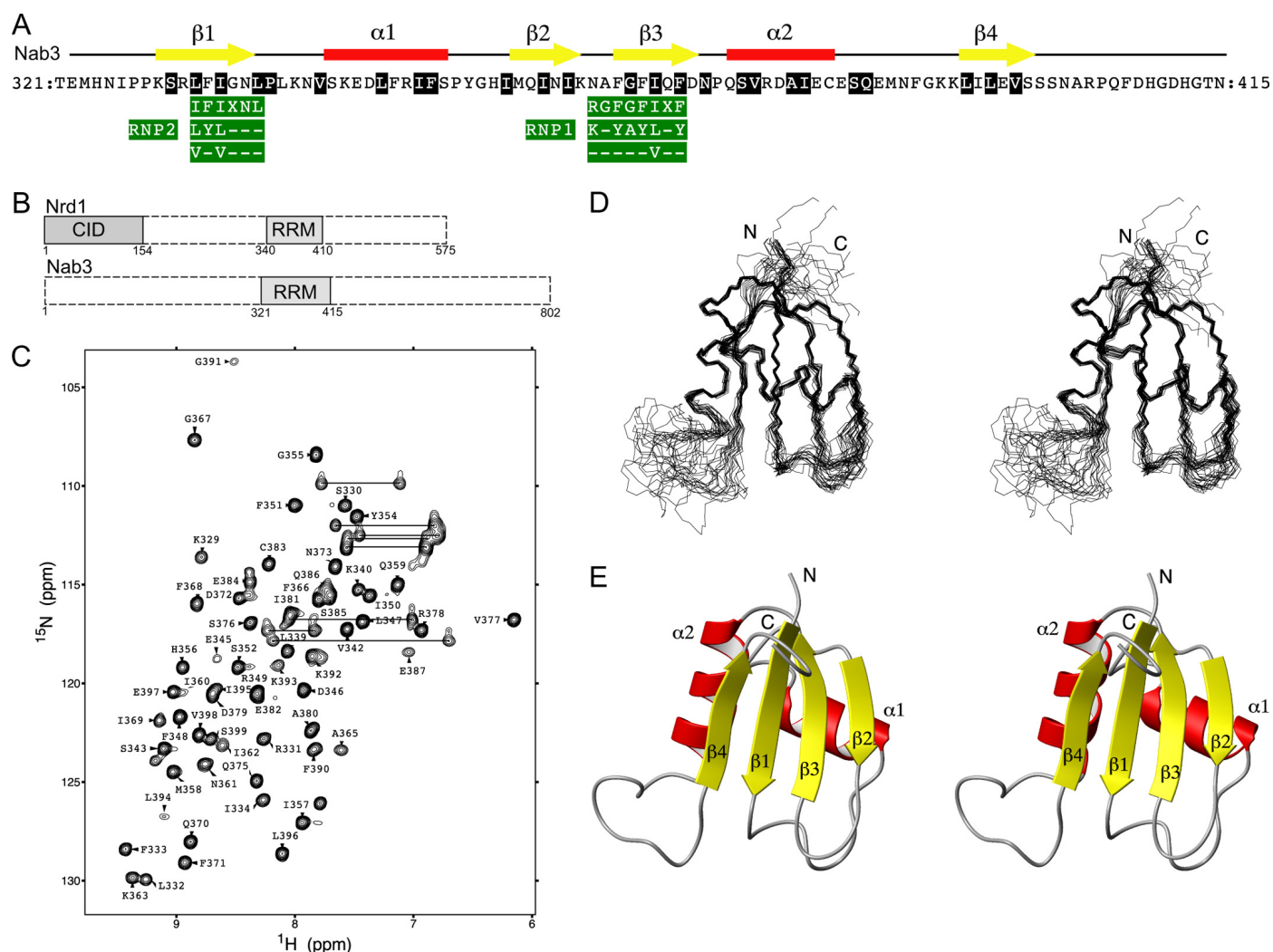
FIGURE 1. **Overview of the RRM of Nab3 sequence, topology, NMR spectra, solution structure, and domain structure in Nab3 and Nrd1.** *A*, amino acid sequence of the *S. cerevisiae* Nab3 RRM along with its secondary structure elements and general consensus of RNP1 and RNP2 motifs. *B*, a schematic drawing of the domain structure of Nab3 and Nrd1. *C*, two-dimensional $^1$H-$^{15}$N HSQC spectrum of 2.5 mM uniformly $^{15}$N,$^{13}$C-labeled Nab3 RRM in 50 mM sodium phosphate buffer (pH 8.0), 300 mM NaCl, and 10 mM $\beta$-mercaptoethanol (90% $H_2O$, 10% $D_2O$). The spectrum was acquired at 303 K on a Bruker Avance 600 MHz spectrometer. The assignments are labeled by the one-letter code of amino acids accompanied by a sequence number. The side chain resonances of asparagine and glutamine are connected by *horizontal lines*. *D*, stereo view of the 20 lowest energy structures of Nab3 RRM. The protein backbone is shown as a wire model. *E*, stereo view of the representative (the lowest energy) structure of Nab3 RRM shown as a *ribbon diagram*. The figure was generated with MOLMOL (38).

tose). These plates were incubated at 24, 30, and 37 °C for 3 days.

*Western Blot Analysis*—Protein extracts were prepared from cultures grown either on galactose containing medium and cultures shifted to glucose containing medium (as described above). Proteins were resolved on a 12% SDS-PAGE gel, transferred to nitrocellulose membrane by a semi-dry electroblotter (Bio-Rad), and probed for the presence of Nab3p with the mAb 2F12 (41), or antibodies directed against the HA epitope (sc805, Santa Cruz Biotechnology) present on the endogenous Nab3. For loading control we used the antibodies against Air2 protein (42).

## RESULTS

*Structure of Nab3 RRM*—The RRM of *S. cerevisiae* Nab3 was examined by NMR spectroscopy (Fig. 1, *A* and *B*). The $^1$H, $^{13}$C, and $^{15}$N chemical shift assignments were obtained as described previously (24). All NMR experiments were mea-

sured at a high salt concentration (300 mM NaCl, 50 mM sodium phosphate (pH 8.0), and 10 mM $\beta$-mercaptoethanol) to prevent protein precipitation. The solution structure determination of Nab3 RRM employed homonuclear and heteronuclear NMR techniques. The $^{15}$N-$^1$H HSQC experiment shows a well dispersed spectrum (Fig. 1*C*), indicating a folded domain. There are a number of missing peaks in this spectrum that mainly correspond to the N- and C-terminal regions of the studied protein construct (outside of the RRM domain). These regions were included in the study as they often form additional structural elements ($\beta$-strand or $\alpha$-helix) in RRMs or contribute to the RNA binding. In addition, several residues in the loops showed no NMR signals. These missing signals are likely a result from the relatively high pH used in the NMR study that was necessary to prevent the precipitation of Nab3 RRM. The three-dimensional structure of Nab3 RRM was determined by combined automated NOESY cross-peak assignment (30) and structure calculations with

**TABLE 1**

**NMR and refinement statistics for Nab3 RRM and Nab3 RRM–UCUU complex**

| NMR distance and dihedral angle restraints | Nab3 RRM | Nab3 RRM–UCUU complex |
|---|---|---|
| **Distance restraints** | | |
| Total NOEs | 857 | 852 |
| Intra-residue | 201 | 228 |
| Inter-residue | | |
| Sequential ($|i–j| = 1$) | 218 | 187 |
| Medium range ($1 \langle|i–j|\rangle 5$) | 164 | 116 |
| Long range ($|i–j|[\text{bar} \geq 5$) | 274 | 310 |
| Hydrogen bond restraints | 25 | 25 |
| Intermolecular | | 11 |
| Dihedral angle restraints | | |
| $\phi$ and $\psi$ | 76 | 88[a] |
| **Structure statistics**[b] | | |
| Residual NOE violations (mean $\pm$ S.D.) | | |
| Number > 0.20 Å | 1.5 ($\pm$ 0.83) | 4 ($\pm$ 2) |
| Maximum (Å) | 0.24 ($\pm$ 0.03) | 0.46 ($\pm$ 0.08) |
| Residual dihedral angle violations | | |
| Number > 10.0° | 0 | 0 |
| Maximum (°) | 0 | 0 |
| **Ramachandran plot statistics**[b,c,d] | | |
| Residues in most favored regions (%) | 90.8 | 88.6 |
| Residues in additionally allowed regions (%) | 9.1 | 9.6 |
| Residues in generously allowed regions (%) | 0.1 | 1.7 |
| Residues in disallowed regions (%) | 0.0 | 0.1 |
| Deviations from idealized geometry | | |
| Bond length (Å) | 0.0011 $\pm$ 0.0001 | 0.0010 $\pm$ 0.0001 |
| Bond angles (Å) | 1.46 $\pm$ 0.02 | 1.48 $\pm$ 0.02 |
| **Average root mean square deviation to mean structure (Å)**[b] | | |
| Protein | | |
| Backbone atoms[b] | 0.57 $\pm$ 0.12 | 0.49 $\pm$ 0.10 |
| Heavy atoms[b] | 1.46 $\pm$ 0.16 | 1.24 $\pm$ 0.12 |
| RNA | | |
| All RNA heavy atoms[e] | | 0.98 $\pm$ 0.21 |
| Complex | | |
| Protein[b] and RNA heavy atoms[e] | | 1.27 $\pm$ 0.12 |
| WHAT IF[f] structure $Z$-scores[d,g] | | |
| Packing quality | −2.3 | −1.9 |
| Ramachandran plot appearance | −3.5 | −3.1 |

[a] Includes $C2'$-*endo* sugar pucker and *anti* conformation of the glycosidic bond was used for all nucleotides (56).

[b] Calculated for an ensemble of the 20 lowest energy structures.

[c] Based on PROCHECK analysis (32).

[d] Calculated for the structured part of the protein construct.

[e] Calculated for $U_1C_2U_3$.

[f] Based on WHAT IF analysis (37).

[g] $Z$-score (54, 55) is defined as the deviation from the average value for this indicator observed in a database of high-resolution crystal structures.

torsion angle dynamics implemented in the program CYANA 2.1 (43), followed by refinement in explicit solvent using AMBER 10 (33). An ensemble of the 20 lowest energy structures along with the best energy structure are shown in Fig. 1, *D* and *E*, respectively. These structures have an average backbone root mean square deviation of 0.57 $\pm$ 0.12 Å for the secondary structure elements. A full summary of structural statistics including the backbone $\phi$-$\psi$ angle distribution is given in Table 1.

The three-dimensional structure of Nab3 RRM adopts a compact fold with an $\beta1\alpha1\beta2\beta3\alpha2\beta4$ topology that is similar to the canonical fold of RRM family (21, 22). The fold is composed of two $\alpha$-helices and a $3_{10}$ helix that are packed along a face of a four-stranded antiparallel $\beta$-sheet. A central hydrophobic core composed of the residues shown in Fig. 1*A* stabilizes the fold of the domain. Nab3 RRM contains a well conserved signature of the RRM family, RNP1 and RNP2 sequences (44–46). These two conserved amino acid sequences found between Leu[332]–Leu[337] and Asn[364]–Phe[371] are located on the $\beta3$- and $\beta1$-strands, respectively. Their sequence compositions correspond to the general RNP2 and RNP1 consensus (ILV)-(FY)-(ILV)-*X*-N-L and

(RK)-G-(FY)-(GA)-(FY)-(ILV)-*X*-(FY), respectively, except for the first two amino acids of the RNP1 (Fig. 1*A*). Nab3 RRM has asparagine and alanine in these positions (Fig. 1*A*). The presence of aromatic residues in RNP1 and RNP2 sequences, which usually mediates the stacking interaction with RNA bases, along with a number of basic and polar residues on the $\beta$-sheet surface, indicate a potential role of Nab3 RRM in RNA binding.

*Characterization of the Nab3-UCUU Interactions by NMR*—To investigate the interaction and binding mode between Nab3 RRM and RNA, we carried out an NMR chemical shift perturbation study with a UCUU element, which has been shown to elicit transcription termination via the Nrd1 pathway. In the RNA titration experiment, we observed that the protein amide resonances moved upon RNA binding from their initial positions, corresponding to the free form, in a stepwise directional manner until they reach their final positions that correspond to the fully bound state, with stoichiometry of 1:1 (Fig. 2). Additional RNA aliquots resulting in excess RNA resulted in no further change of chemical shifts, confirming the 1:1 stoichiometry of the complex. These titration data suggest that protein amide resonances are in a fast exchange regime between
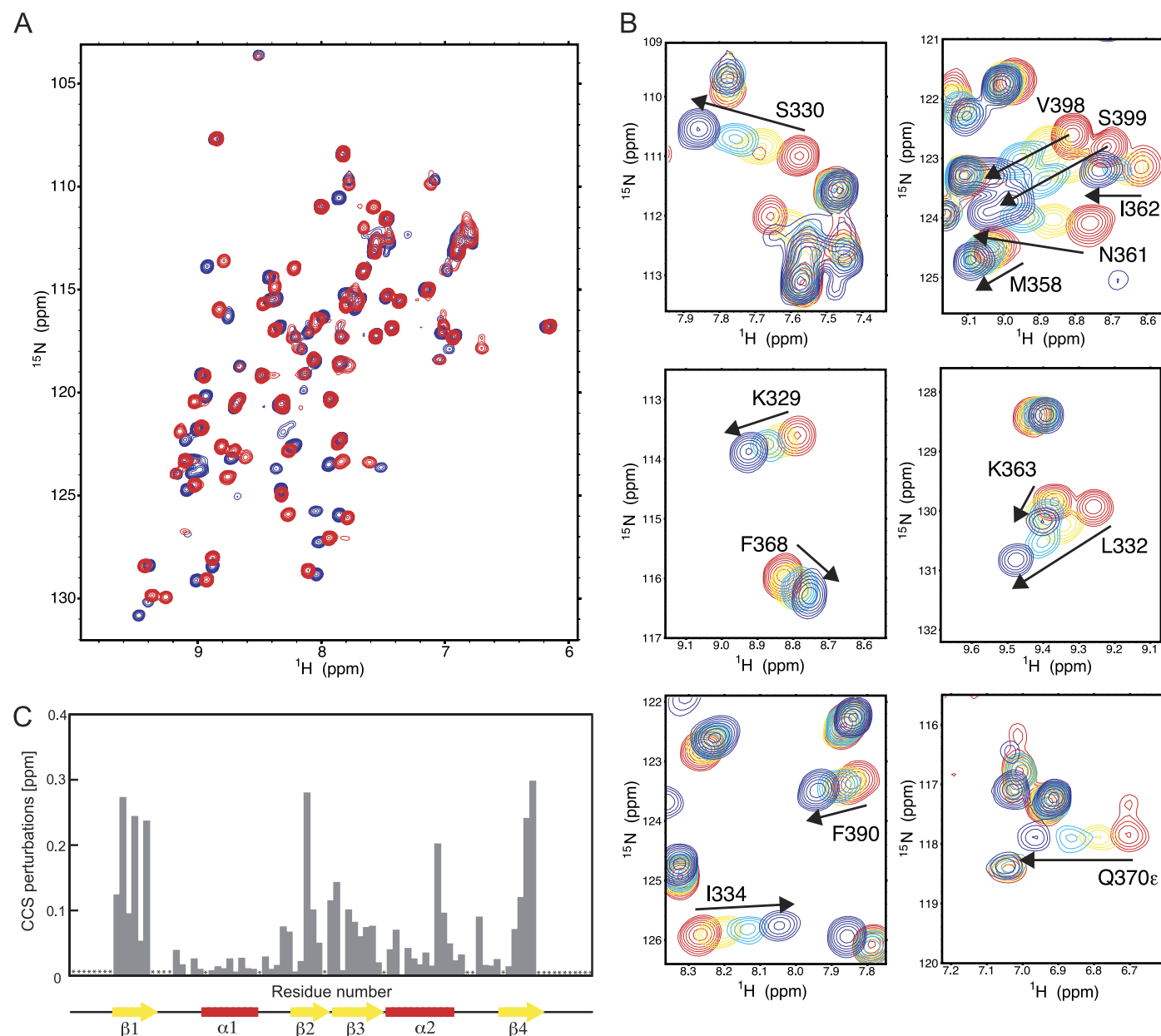
FIGURE 2. **NMR titration experiments of Nab3 RRM with UCUU RNA.** *A*, $^1$H-$^{15}$N HSQC spectra of Nab3 RRM alone (in *red*) and in the presence of 1 eq of 5′-UCUU-3′ (in *blue*) at 303 K. *B*, close-up views of the $^1$H-$^{15}$N HSQC spectra, showing selected chemical shift changes during the titration. *C*, quantification of chemical shift perturbations of Nab3 RRM upon binding to UCUU RNA. The combined chemical shift perturbations ($[\omega_{HN}\Delta\delta_{HN}]^2 + [\omega_N\Delta\delta_N]^2)^{1/2}$, where $\omega_{HN} = 1$ and $\omega_N = 0.154$ are weight factors of the nucleus (52), are plotted *versus* the amino acid residue number. Large changes occur on the β-sheet surface. The assignments of residues indicated by *asterisks* could not be obtained for neither the free nor bound protein, or indicates proline residues.

their free and bound forms relative to NMR time scale. The binding of UCUU to the RRM of Nab3 induces chemical shift perturbation of the residues shown in Fig. 2.

These chemical shift changes indicate that the above mentioned residues are involved in binding to the RNA, or alternatively, could undergo a conformational change upon RNA binding. Mapping the perturbed residues on the sequence of Nab3 RRM delineates that the Nab3 RRM binds the RNA through its β-sheet surface and also through the β2β3 loop (Fig. 2C).

*Structure of Nab3 RRM in Complex with UCUU*—When solving the structure of Nab3 RRM bound to RNA, we extensively tested different lengths of RNA, buffer conditions, and temperatures with the aim to optimize the NMR spectral quality of the complexes. Longer RNA substrates, a UCUU

core motif with flanking sequences, resulted in the significant broadening of NMR signals of the complexes. Interestingly, we obtained the NMR spectra of better quality (for both protein and RNA in complex) with a four-nucleotide UCUU despite the fact that this RNA has lower affinity to Nab3 RRM compared with the longer substrates (see below). Similar improvement of the NMR spectral quality by using of a minimal specific RNA sequence has also been observed for other protein-RNA complexes investigated by NMR (47– 49). Therefore, we pursued the structure determination of the Nab3 RRM-UCUU complex.

The RRM of Nab3 in complex with RNA display the canonical RRM-fold with an β1α1β2β3α2β4 topology and is similar to that of the unbound form (Figs. 1 and 3). Akin to the free
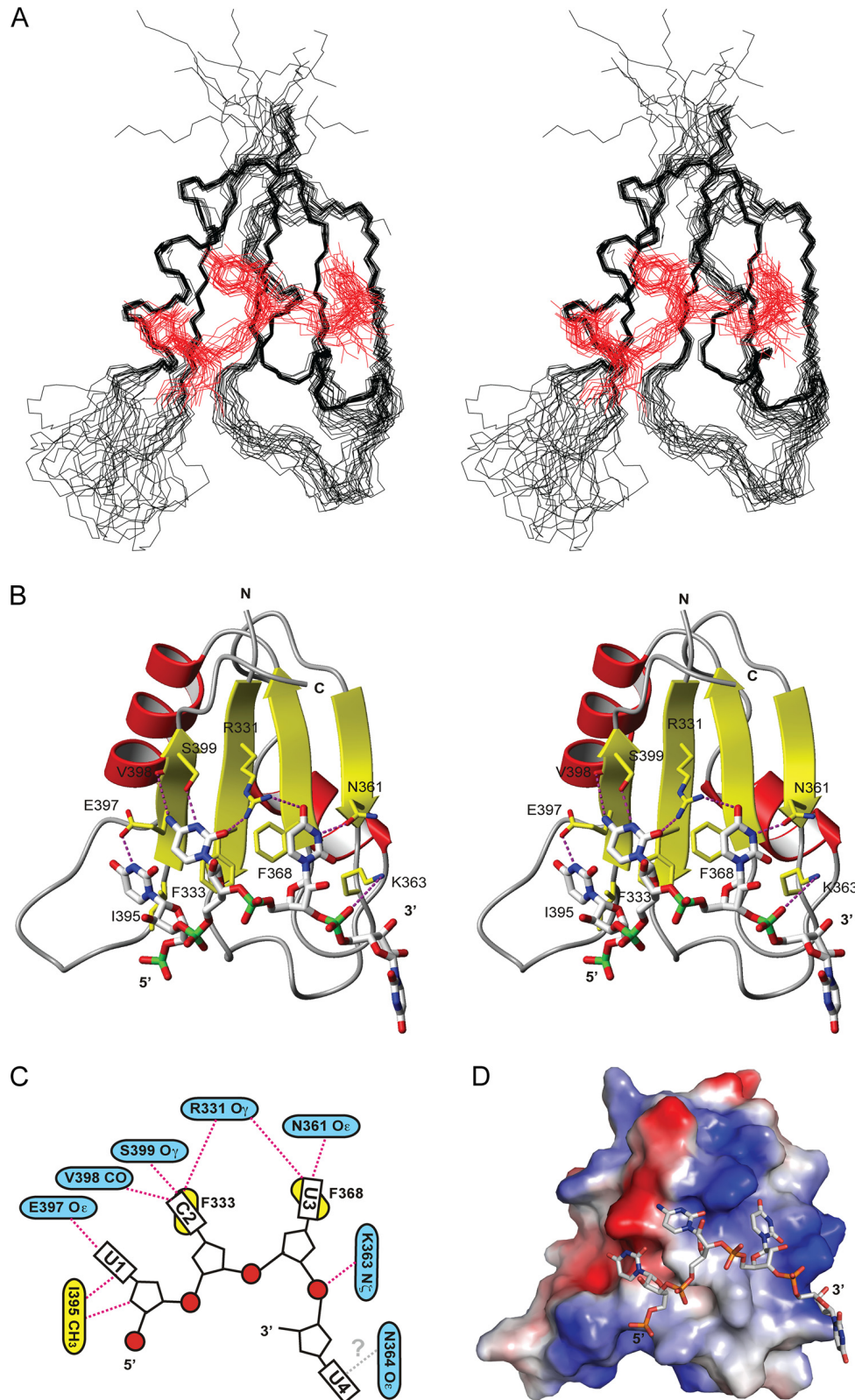
FIGURE 3. **Overview of the solution structure of the Nab3 RRM in complex with UCUU.** *A*, stereo view of the 20 lowest energy structures of the Nab3 RRM-UCUU complex. The protein backbone is shown as a wire model in *black*. The RNA heavy atoms are shown as a wire model in *red*. *B*, stereo view of the representative (the lowest energy) structure of the Nab3 RRM-UCUU complex. The RNA is represented as a *white stick model* and the protein is shown as a *ribbon* model with residues that contact the RNA shown in *yellow*. Putative hydrogen bonds are shown by *dotted magenta lines*. *C*, scheme showing contacts between Nab3 RRM and the UCUU RNA. Protein residues that form putative hydrogen bonds to the RNA are shown in *blue* and the one having hydrophobic interactions are in *yellow*. A hypothetic recognition of U4 is labeled by a *gray question mark*. *D*, solvent-accessible surface representation of Nab3 RRM colored by electrostatic potential (*blue*, positive; *red*, negative) and stick representation for the RNA of the representative structure of the complex. Figures were generated with MOLMOL (38).

form of Nab3 RRM, the N- and C-terminal regions as well as the long $\alpha2\beta4$ loop are structurally undefined due to the lack of experimental data. The UCUU RNA adopts a single-stranded conformation and the first three nucleotides are positioned over the whole $\beta$-sheet surface in a canonical arrangement in which the 5' end is located on the first half of the $\beta$-sheet ($\beta4\beta1$) and the 3' end on the second half ($\beta3\beta2$) (21, 22) (Fig. 2). The overall position of RNA on the $\beta$-sheet coincides with the perturbed residues from the titration experiment (Figs. 2 and 3). All bases have an *anti* conformation of the glycosidic bond and *C2'-endo* conformation of the sugar pucker.

The NMR spectra provided a limited number of intermolecular NOEs (11 unambiguous intermolecular NOEs) that loosely define the position of $U_1C_2U_3$ on the $\beta$-sheet surface of Nab3 RRM (Fig. 3*A*), but are sufficient to reveal the molecular basis of $U_1C_2U_3$ recognition by Nab3 RRM (Fig. 3, *B* and *C*; the protein-RNA hydrogen bonds described below are inferred from the final ensemble of structures and thus they should be considered as putative hydrogen bonds). We could not define the position of $U_4$ due to the lack of intermolecular NOEs. Based on NMR titration data, we speculate that $U_4$ could be recognized by the asparagine side chain or a main chain of the $\beta2-\beta3$ loop that are in proximity to the base of $U_4$, as displayed in the representative structure (Fig. 3*B*). In our NMR structure, $C_2$ and $U_3$ are involved in base stacking with the aromatic rings of Phe[333] and Phe[368], respectively. The Watson-Crick edge of $C_2$ is recognized by the main chain carbonyl group of Val[398] and the hydroxyl group of Ser[399] that form hydrogen bonds with the amino and imino groups of $C_2$, respectively. One-half of the 20 structures in the final ensemble has Arg[331] in a position in which it contacts the $O_2$ oxygen of $C_2$ (for the importance of Arg[331] see below). In addition, it is likely also that Ser[400] could be involved in the recognition of $C_2$ as the resonances of this residue broadened beyond detection upon RNA binding. Our structure also rules out the possibility that a purine could be accommodated in the $C_2$ position due to a steric restriction imposed by the Glu[397] side chain.

The recognition of $U_3$ is mediated by the Arg[331] and Asn[361] side chains. The side chain $NH_2$ group of Arg[331] contacts the O4 carbonyl functional group of the base and the side chain carbonyl group of Asn[361] forms a hydrogen bond with the imino proton of $U_3$. Akin to $C_2$, the position of $U_3$ cannot be exchanged by a purine due to a steric hindrance of Arg[331]. The sugar of the $U_3$ residue is further contacted by the aliphatic region of the Lys[363] side chain.

In contrast to $C_2$ and $U_3$, the recognition of $U_1$ is less evident from the structure. A single hydrogen bond is formed between the imino proton of $U_1$ and the $O\epsilon$ of Glu[397]. There are also hydrophobic contacts between the sugar and the base of $U_1$ and the side chain of Ile[395]. However, these contacts do not explain fully the sequence specificity of a uridine nucleotide. A cytidine nucleotide in this position could also form a similar interaction with the glutamate. Altogether, our NMR structure indicates that Nab3 RRM recognizes the YCU sequence (where Y stands for pyrimidine).

*Nab3 RRM Binds the Nab3 Termination Element with Low Affinity*—The FA measurements were carried out to further characterize the binding of Nab3 RRM to various RNA substrates. In FA measurements, formation of the protein-RNA complex is monitored directly from an increase of the FA value that occurs when the protein binds fluorescently labeled RNA. Binding curves were recorded in the course of titration experiments, where protein aliquots were added to 10 nM fluorescently labeled RNAs (Fig. 4).

First, we assayed the binding affinity of Nab3 RRM to UCUU that has been reported as the minimal Nab3 terminator element. A tetranucleotide GUAA was used as a nonspecific control substrate. The comparison of the anisotropy data for specific and nonspecific four-nucleotide substrates along with determined equilibrium dissociation constants ($K_d$) are shown in Fig. 4*A*. Corresponding logarithmic values of $K_a$ ($K_a = 1/K_d$) are shown in graph in Fig. 4*E*. Nab3 RRM binds the specific recognition sequence UCUU with more than 6-fold higher affinity compared with the nonspecific substrate GUAA.

As terminator elements often occur in multiple repeats, we tested a longer substrate with three UCUU repeats. The substrate sequence was derived from the snR47 that is terminated by the Nrd1 pathway (Fig. 4*F*). Nab3 RRM binds the three UCUU-containing substrate with a $K_d$ of $48 \pm 2$ $\mu M$, 1 order of magnitude stronger than we observed for a single UCUU motif (Fig. 4*B*). Furthermore, we assayed snR13, which is another naturally occurring Nrd1-dependent terminator that contains two UCUU and one CCU motifs (Fig. 4*F*). As expected, Nab3 RRM binds this substrate with a $K_d$ of $46 \pm 1$ $\mu M$, a similar binding affinity to that of snR47 (Fig. 4*B*). As a control, we used SL RNA, which has a similar size and its sequence lacks UCUU or even CU recognition motifs (Fig. 4*F*). For this nonspecific substrate, Nab3 RRM binding is reduced more than 30-fold compared with the snR13 and snR47 terminators (Fig. 4, *B* and *E*).

*Nrd1 RRM Binds the Nrd1 Termination Element with Low Affinity*—As Nab3 RRM binds its termination motif with a low affinity, we therefore decided to investigate the RNA-binding properties of Nrd1 RRM, which is the second RNA-binding domain occurring in the Nrd1-dependent termination complex (Fig. 1*B*). Akin to Nab3 RRM, we assayed the binding affinity of Nrd1 RRM to RNA using FA. We found that Nrd1 RRM-(340–410) binds to the minimal termination RNA sequence GUAA (GUAR is known as the Nrd1-termination element (where R stands for purine) (18) with a low affinity in the mid-micromolar range (Fig. 4*C*; $K_d$ of $66 \pm 1$ $\mu M$). Next, we assayed a longer RNA substrate, snR13, which contains two GUAR motifs (Fig. 4*F*). The titration curve for snR13 yielded a $K_d$ of $11 \pm 1$ $\mu M$ (Fig. 4*C*).

*A Complex of Nrd1-Nab3 Binds RNA with High Affinity*—As the Nrd1 and Nab3 proteins form a heterodimer *in vivo* and *in vitro* (19), we assayed RNA binding of the co-expressed Nrd1-Nab3 heterodimer (19) using FA. As a substrate, we used RNA derived from the snR13 terminator that contains two copies of each recognition element (Fig. 4*F*). This RNA substrate binds the Nrd1-Nab3 heterodimer with a $K_d$ of $2.4 \pm 0.3$ nM (Fig. 4*D*), about 4–5 orders of magnitude stron-
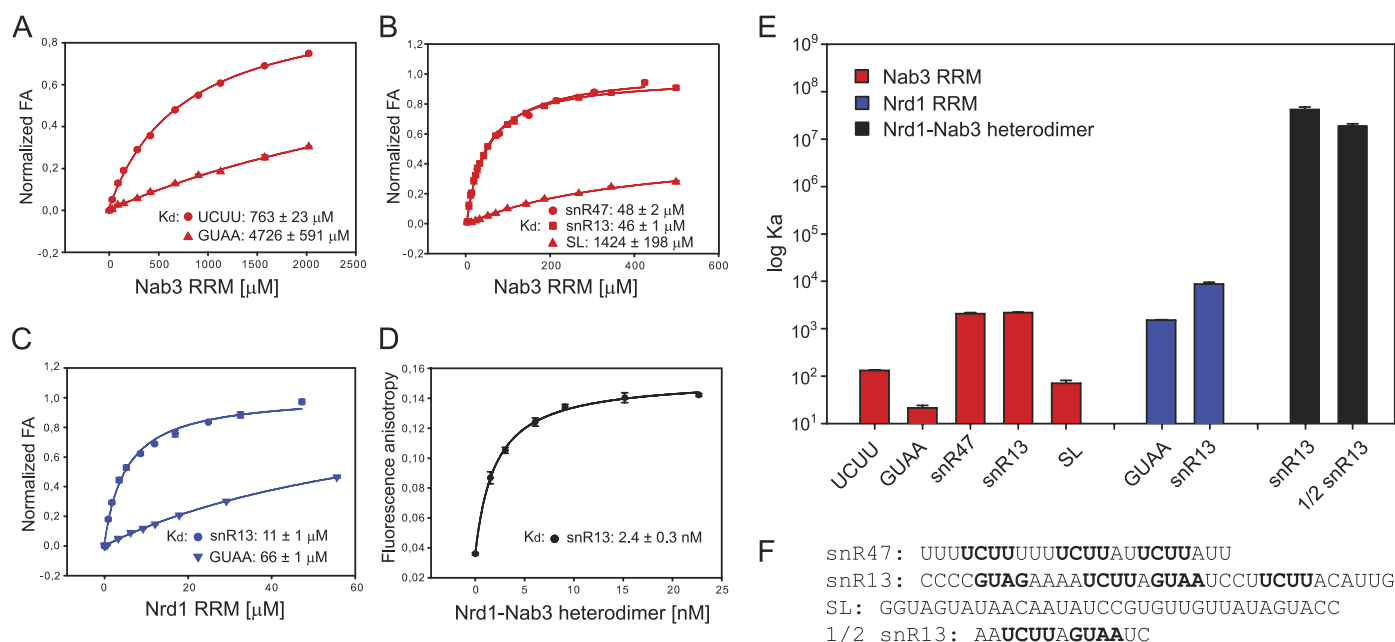
FIGURE 4. **Equilibrium binding of Nab3 RRM, Nrd1 RRM, and Nrd1-Nab3 heterodimer with fluorescently labeled RNA monitored by fluorescence anisotropy.** *A*, Nab3 RRM was titrated with UCUU and GUAA (each 10 nM), and their binding isotherms are shown as *red circles* and *triangles*, respectively. *B*, Nab3 RRM was titrated with snR47, snR13, and SL substrates (each 10 nM), and their binding isotherms are shown as *red circles*, *squares*, and *triangles*, respectively. *C*, Nrd1 RRM was titrated with GUAA and snR13 (each 10 nM) and their binding isotherms are shown as *blue inverted triangles* and *circles*, respectively. *D*, Nrd1-Nab3 heterodimer was titrated with snR13 (100 pM). *E*, summary of the association constants ($K_a$) for the RRM of Nrd1 (in *blue*) and Nab3 (in *red*) in their free forms as well as for the Nrd1-Nab3 heterodimer (in *black*). Logarithmic scale of $K_a$ is shown to cover a wide range of affinities. *F*, RNA sequences used in the affinity measurements. The buffers contained the same ion strength and pH values for all proteins. Equilibrium dissociation constant ($K_d$) was calculated from the best fit to the data using a single-site binding isotherm. *Error* is denoted as S.E. The data were normalized for visualization purposes (*A–C*).

ger compared with the individual RRMs of Nrd1 and Nab3 (Fig. 4*E*). Another snR13-derived substrate, ½ snR13, containing only one copy of each recognition element (Fig. 4*F*), binds the Nrd1-Nab3 heterodimer with only a slightly lower affinity ($K_d$ of 5.3 ± 0.6 nM). However, when one of the recognition elements (either Nrd1 or Nab3) is removed from the ½ snR13 substrate, the affinity is reduced such that the dissociation constant cannot be determined (saturation of the binding curve cannot not be reached due to the low solubility limit of the Nrd1-Nab3 heterodimer; data not shown).

*Impacts of Nab3 RRM Point Mutations on RNA Binding Affinity*—It has been shown previously that mutations of $C_2$ and $U_3$ in the UCUU motif reduce the binding affinity to Nab3 (18) or to the Nrd1-Nab3 heterodimer (19). We performed the converse experiments in which we assessed Nab3 RRM mutants for their ability to bind snR47 RNA in a quantitative solution binding assay by fluorescence anisotropy titration experiments. We mutated the non-canonical amino acid residues on the β-sheet surface (R331A, N361A, S399A, and E397A) that specifically recognize the bases of the $U_1C_2U_3$ sequence (Fig. 3, *B* and *C*). These residues surround the conserved residues of RNP1 and RNP2 consensuses. We found that mutants R331A and S399A showed a 3–4-fold decrease in binding affinity of that demonstrated by the wild-type protein (Fig. 5*A*). In contrast, mutants N361A and E397A showed binding affinity similar to the wild-type Nab3 RRM (Fig. 5*A*).

*Functional Significance of the Nab3 RRM Residues That Contact RNA*—To address the importance of the specific contacts identified in the Nab3 RRM-UCUU complex for Nab3 function *in vivo*, single amino acid mutants (R331A, S399A,

S399K, E397A, and E397K) were prepared in a yeast expression vector and introduced into a yeast strain in which the endogenous NAB3 promoter was replaced with the GAL1 promoter (3). To test whether the mutated residues were essential for growth, the resulting transformants were spotted onto glucose containing plates. The shift to glucose represses the expression of the GAL1-driven endogenous NAB3, which completely impairs cell viability (Fig. 5*B*). This lethality was rescued by wild-type Nab3 (Fig. 5*B*). Mutating the three residues involved in the recognition of $C_2$ and $U_3$ (Fig. 3), R331A, N361A and S399A, or S399K, caused lethality (Fig. 5*B*), providing further support for the functional significance of these contacts. In contrary, the Glu[397] mutant displayed only slow growth at 24 and 37 °C.

To further confirm that mutant lethality was not due to impaired protein expression, we performed Western blot analysis with antibodies directed against Nab3 and antibodies against the HA tag present on the endogenous Nab3 only. We detected similar levels of Nab3 in all proteins expressed episomally, whereas no HA-tagged endogenous Nab3 was detected in yeast grown on glucose medium (Fig. 5*C*).

## DISCUSSION

*RNA Recognition by Nab3 RRM and Its Comparison with Other RRMs*—We have solved the structure of Nab3 RRM in free form and in complex with UCUU. In both forms, the fold of Nab3 RRM is very similar, with minor adjustments upon RNA binding, and resembles canonical RRM (21–23). However, it has shorter α-helices than observed in the structures
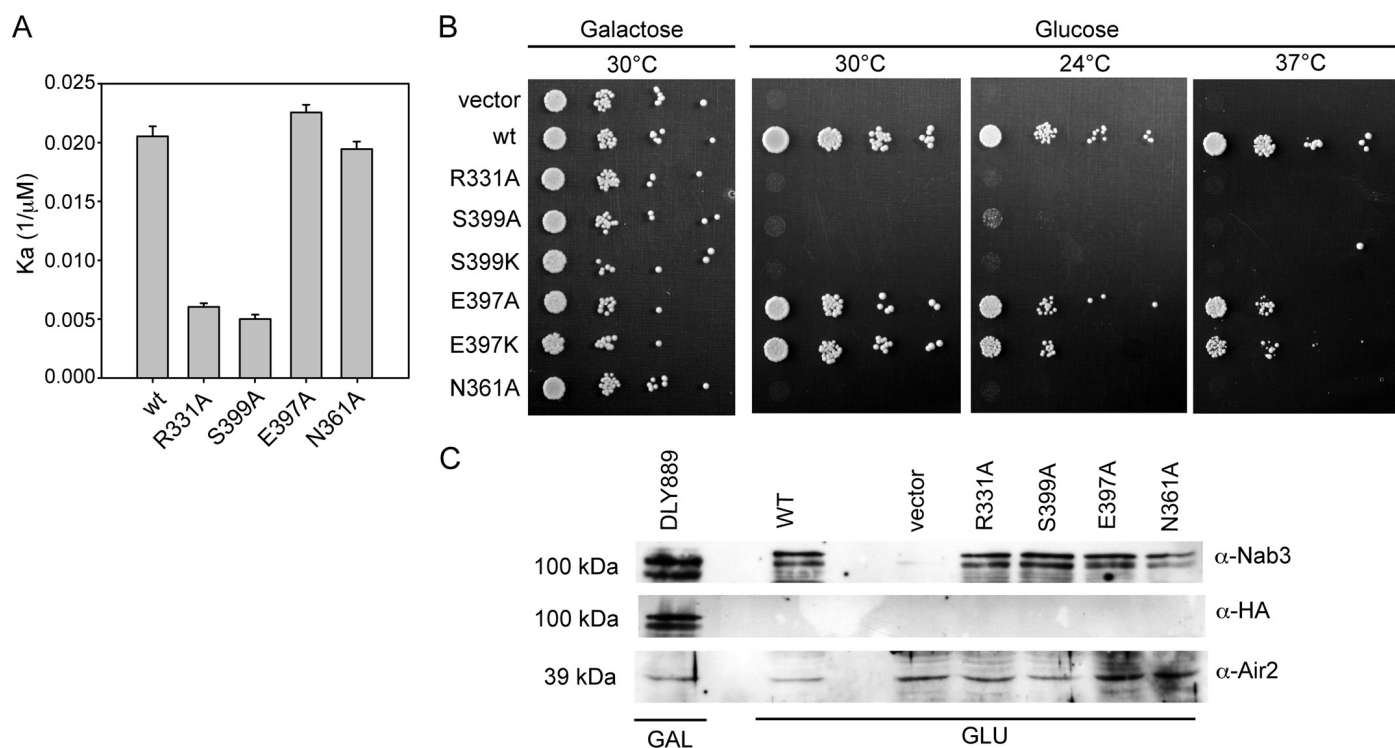
A



B

C

FIGURE 5. **The important residues of Nab3 RRM that are required for RNA binding and cell viability.** *A*, equilibrium binding of Nab3 RRM mutants with fluorescently labeled RNA monitored by fluorescence anisotropy. The Nab3 RRM R331A, N361A, S399A, and E397A mutants along with the wild-type of Nab3 RRM were titrated with fluorescently labeled snR47 substrate. Equilibrium association constants ($K_a$) are shown for individual mutants with S.E. *B*, residues Arg[331], Ser[399], and Asn[361] are required for yeast viability. The indicated Nab3 RRM mutants were expressed episomally from pRS415 plasmids in the yeast strain with the endogenous NAB3 driven by the GAL1 promoter. Mutant strains were spotted on plates containing 2% glucose and a control galactose plate and incubated for 3 days at the indicated temperatures. Growth on glucose-containing plates leads to the repression of GAL1-driven wild-type Nab3, and thus shows the functionality of the different Nab3 mutants. *Vector* is a control where the GAL1::NAB3 strain contains an empty pRS415 plasmid, *wt* is the wild-type NAB3. *C*, expression of Nab3 proteins from pRS415 in glucose-containing medium. Western blot analysis was performed with protein extracts from the original GAL1::NAB3 strain (DLY889) grown in galactose-containing medium and extracts from DLY889 transformed with plasmids carrying wild-type and mutant NAB3 grown for 20 h in glucose-containing medium. *Air2* was used as a loading control.

of canonical RRMs, and an extra $3_{10}$ helix between the $\alpha 1$ helix and $\beta 2$ strand (22).

Nab3 RRM binds UCUU in a canonical manner as observed in other RRM-RNA complexes (21–23). Specifically, Nab3 RRM contains a well conserved signature of the RRM family, RNP1 and RNP2 sequences (44–46). Two conserved phenylalanine residues are used in RNP1 (Phe[333]) and RNP2 (Phe[368]) to mediate the stacking interaction with RNA bases $C_2$ and $U_3$, respectively. These aromatic residues are surrounded by basic and polar amino acid residues that mediate the sequence-specific recognition of $C_2U_3$ and in part of $U_1$. The specifically recognized nucleotides are accommodated on the $\beta$-sheet only and neither loops nor N-/C-terminal regions to the RRM are involved in the recognition process.

Interestingly, the binding preference for CU has also been reported for the polypyrimidine tract-binding protein (PTB) (49). Unlike Nab3 RRM, PTB RRMs display additional topology elements to the canonical RRM involved in RNA recognition and significantly differ from RNP1 and RNP2 consensus (Fig. 6A). In particular, PTB RRMs lack aromatic residues in RNP1 and RNP2 that usually make extensive stacking interactions with the RNA bases and sugars as in the case of Nab3 RRM (Fig. 6). In addition, the structures of Nab3 and PTB RRMs revealed that these domains are sequentially unrelated on the entire RNA interaction surface, except for the last serine residue of $\beta$-strand 4 that is present in all PTB and Nab3

RRMs (Fig. 6A). The serine residue is involved in recognition of a cytosine in the structures of PTB (49) and Nab3 bound to RNA (Fig. 6). Furthermore, in the structures of PTB RRM1-RNA and Nab3 RRM-RNA, recognition of the 5′-end uridine is mediated in a similar way in which its imino proton is contacted by glutamine and glutamate, respectively (Fig. 6). In both structures, a cytidine could be tolerated instead of a uridine in this position. In contrast, recognition of the 3′-end uridine is mediated differently in these two structures. Whereas Nab3 RRM utilizes the side chains of arginine (in $\beta 1$) and asparagine (in $\beta 2$) to recognize the uridine, PTB RRM1 uses the main chain of leucine and lysine located in the C-terminal extension to the RRM to facilitate the uridine binding.

The importance of the serine residue, which is discussed above, is demonstrated by our affinity measurements with the Nab3 RRM S399A mutant that decreases the binding affinity to its UCUU-containing substrate (Fig. 5A). Correspondingly, *in vivo* analyses of Nab3 S399A or S399K mutants show that the serine residue is essential for yeast viability (Fig. 5B). Similarly, the decreased affinity of the Nab3 RRM R331A mutant to the RNA is manifested by the lethal effect in yeast (Fig. 5). Interestingly, Asn[361], which specifically recognizes $U_3$ along with Arg[331], shows lethality *in vivo* for the N361A mutant, whereas the affinity of this mutant is only slightly reduced compared with the wild-type (Fig. 5). Furthermore, the
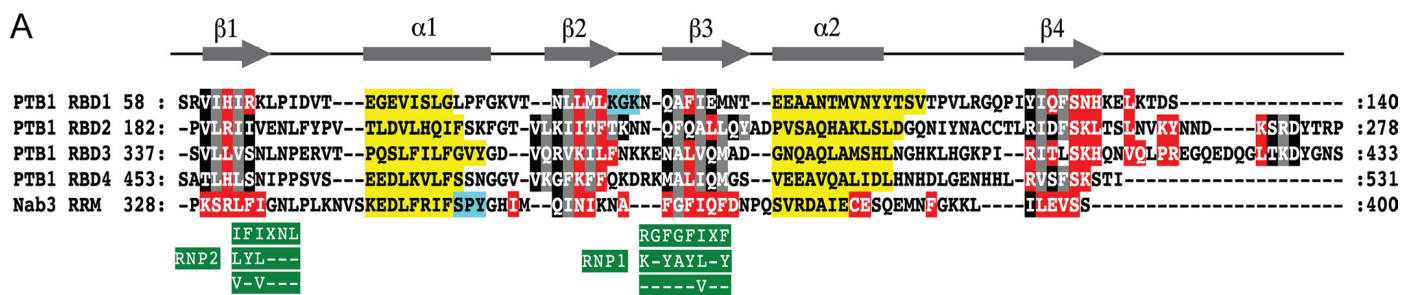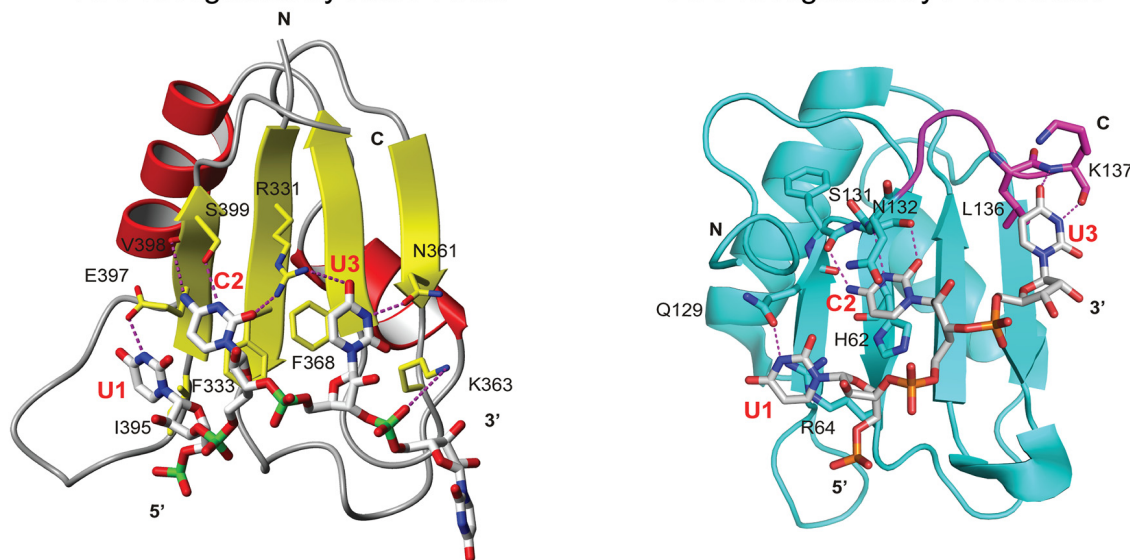
FIGURE 6. **Recognition of YCU by Nab3 and PTB RRMs.** *A*, sequence alignment of PTB RRM1, -2, -3, and -4 and Nab3 RRM whose structures have been solved. The alignment was performed using ClustalW (53) and manually optimized using the three-dimensional structural information (49). For the RRMs of PTB, amino acids interacting with the RNA are shown in *red boxes*, residues in *gray* and *black boxes* are located in the β-sheet and residues in *yellow* and *cyan boxes* are in the α- or 3₁₀ helices, respectively (49). Residues in *gray boxes* form the hydrophobic core of the domains. For the RRM of Nab3, residues in *red boxes* are significantly perturbed upon RNA binding. *B*, comparison of Nab3 RRM (*left*: in *yellow* and *red schematics*) and PTB RRM1 (*right*: in *cyan schematics*) binding to UCU nucleotides (represented as a *stick model*). The protein residues that mediate the specific recognition are highlighted as a stick model. The C-terminal region of PTB RRM1 that mediates the recognition of $U_3$ is shown in *magenta*.

N361A mutant yeast strain shows accumulation of cryptic unstable transcripts when shifted to glucose,[7] supporting the functional significance of this contact in the termination via the Nrd1 pathway. The Glu[397] mutant does not have an impact on RNA binding *in vitro* and causes only slow growth at higher temperatures, indicating a minor role of this contact for the function of Nab3. Altogether, these functional data corroborate with our structural findings that Nab3 RRM specifically recognizes the YCU sequence.

*Association between Nab3 and Nrd1 Increases the Affinity to the Termination Sequences*—Our FA experiments showed that Nab3 RRM binds UCUU with the high micromolar range of equilibrium dissociation constant ($K_d$), yet the Nab3 RRM binds nonspecific four-nucleotide RNA with a $K_d$ in the low millimolar range (Fig. 4). The observed affinity for the specific RNA substrate is weaker than the affinity usually observed for single canonical RRMs (21, 47). The affinity of Nab3 RRM binding to UCUU is lower by 2-fold compared with the affinity of PTB RRM1 to CUCU that also recognizes the YCU motif (Fig. 6) (47). Interestingly, the apparent $K_d$ of Nab3 RRM binding to

longer RNAs with multiple UCUU motifs (snR47 and snR13) are ~50 μM, more than 1 order of magnitude stronger than we observed for a single UCUU motif (Fig. 4). The increased binding affinity likely originates from the presence of multiple binding sites and due to the presence of flanking sequences to the UCUU motif in snR47 and snR13 RNA substrates. The effect of flanking sequences is likely nonspecific, mediated by electrostatic interactions of additional phosphate groups. A similar increase in the binding affinity, when multiple binding motifs are present, has also been observed for the RRMs of PTB (47). The formation of multiple complexes between the snR47-derived RNA substrate and an RRM-containing Nab3 construct (Nab3-(277–565)) has also been observed previously using electrophoretic mobility shift assay (18). It has been also demonstrated that mutations in the 2nd, 3rd, and 4th position of the UCUU motif decrease the binding affinity to the Nab3-(277–565) (18). Furthermore, our data showed that Nrd1, the second RNA-binding subunit of the Nrd1 complex containing a single RRM, binds its termination element, GUAA, with a low affinity ($K_d$ of 66 ± 1 μM; Fig. 4).

Therefore, it is very likely that Nrd1 and Nab3 bind RNA in a cooperative manner to achieve the nanomolar range of af-

---

[7] F. Hobor, D. Hrossova, S. Vanacova, and R. Stefl, unpublished data.

finity to snR13 RNA, as previously estimated using electrophoretic mobility shift assay (EMSA) (19). Indeed, our FA measurements showed that the Nrd1-Nab3 heterodimer binds snR13 RNA with a $K_d$ in the low nanomolar range, 4–5 orders of magnitude stronger compared with the individual RRMs of Nrd1 and Nab3 (Fig. 4*E*). The RRMs of Nrd1 and Nab3 do not bind each other (data not shown) and the regions that mediate the formation of the heterodimer are located near the N-terminal of each RRM (13).[7] It is difficult to assay the binding affinity of full-length Nrd1 and Nab3 individually due to their instability (19); the Nrd1 alone rapidly aggregates.[7] However, other regions outside of the RRMs of Nrd1 and Nab3 are not expected to contribute significantly to RNA binding as they do not contain an identifiable RNA-binding domain. In a similar way, the cooperative RNA binding of two RRM-containing proteins U2AF[65] and U2AF[35] is utilized to enhance the affinity and selectivity in the process of defining the site of spliceosomal assembly (50, 51).

*Implication for Poly(A) Independent Transcription Termination*—We have shown that the RRM of Nab3 binds specifically the YCU sequence (where Y stands for pyrimidine). This is in good agreement with previous functional data that led to the proposal of the UCUU sequence as the Nab3 termination element (6, 18). However, the first position of the UCUU motif is not fully conserved in some snRNA downstream sequences (18); it can be either U or C that perfectly matches our structural findings. Furthermore, the last position of UCUU is also not fully conserved but we cannot explain the recognition of another nucleotide due to the lack of experimental data for the recognition of this nucleotide in our structure. Considering the specificity only for the CU dinucleotide and relatively weak affinity of Nab3 RRM to YCU-like sequences, it is evident that Nab3 alone cannot recruit the Nrd1 complex to the correct termination sites. Indeed, we show the first quantitative evidence that the association of Nrd1 and Nab3 (each protein contains a single RRM) facilitates high affinity binding and sequence selectivity. It remains to be seen whether additional sequence elements are recognized, in addition to YCU and GUAR, upon the association of Nrd1-Nab3.

## REFERENCES

1. Richard, P., and Manley, J. L. (2009) *Genes Dev.* **23,** 1247–1269
2. Bentley, D. (2002) *Curr. Opin. Cell Biol.* **14,** 336–342
3. Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006) *Mol. Cell* **23,** 853–864
4. Arigo, J. T., Eyler, D. E., Carroll, K. L., and Corden, J. L. (2006) *Mol. Cell* **23,** 841–851
5. Steinmetz, E. J., and Brow, D. A. (1996) *Mol. Cell. Biol.* **16,** 6993–7003
6. Steinmetz, E. J., Conrad, N. K., Brow, D. A., and Corden, J. L. (2001) *Nature* **413,** 327–331
7. Vasiljeva, L., and Buratowski, S. (2006) *Mol. Cell* **21,** 239–248
8. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997) *Cell* **91,** 457–466
9. Vanácová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005) *PLoS Biol.* **3,** e189
10. LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005) *Cell* **121,** 713–724
11. Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., Libri, D., and Jacquier, A. (2005) *Cell* **121,** 725–737
12. Proudfoot, N. J., Furger, A., and Dye, M. J. (2002) *Cell* **108,** 501–512
13. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008) *Nat. Struct. Mol. Biol.* **15,** 795–804
14. Gudipati, R. K., Villa, T., Boulay, J., and Libri, D. (2008) *Nat. Struct. Mol. Biol.* **15,** 786–794
15. Steinmetz, E. J., and Brow, D. A. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95,** 6699–6704
16. Conrad, N. K., Wilson, S. M., Steinmetz, E. J., Patturajan, M., Brow, D. A., Swanson, M. S., and Corden, J. L. (2000) *Genetics* **154,** 557–571
17. Morlando, M., Greco, P., Dichtl, B., Fatica, A., Keller, W., and Bozzoni, I. (2002) *Mol. Cell Biol.* **22,** 1379–1389
18. Carroll, K. L., Pradhan, D. A., Granek, J. A., Clarke, N. D., and Corden, J. L. (2004) *Mol. Cell Biol.* **24,** 6241–6252
19. Carroll, K. L., Ghirlando, R., Ames, J. M., and Corden, J. L. (2007) *RNA* **13,** 361–373
20. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D.,

Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001) *Science* **291,** 1304–1351

21. Maris, C., Dominguez, C., and Allain, F. H. (2005) *FEBS J.* **272,** 2118–2131
22. Stefl, R., Skrisovska, L., and Allain, F. H. (2005) *EMBO Rep.* **6,** 33–38
23. Cléry, A., Blatter, M., and Allain, F. H. (2008) *Curr. Opin. Struct. Biol.* **18,** 290–298
24. Pergoli, R., Kubicek, K., Hobor, F., Pasulka, J., and Stefl, R. (2010) *Biomol. NMR Assign.* **4,** 119–121
25. Bax, A., and Grzesiek, S. (1993) *Acc. Chem. Res.* **26,** 131–138
26. Sattler, M., Schleucher, J., and Griesinger, C. (1999) *Prog. Nucl. Magn. Reson. Spectrosc.* **34,** 93–158
27. Varani, G., Aboulela, F., and Allain, F. H. (1996) *Prog. Nucl. Magn. Reson. Spectrosc.* **29,** 51–127
28. Peterson, R. D., Theimer, C. A., Wu, H., and Feigon, J. (2004) *J. Biomol. NMR* **28,** 59–67
29. Zwahlen, C., Legault, P., Vincent, S. J., Greenblatt, J., Konrat, R., and Kay, L. E. (1997) *J. Am. Chem. Soc.* **119,** 6711–6721
30. Güntert, P. (2004) *Methods Mol. Biol.* **278,** 353–378
31. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) *J. Mol. Biol.* **319,** 209–227
32. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) *J. Biomol. NMR* **44,** 213–223
33. Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., and Kollman, P. A. (2008) *Amber 10,* University of California, San Francisco, CA
34. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc.* **117,** 5179–5197
35. Padrta, P., Stefl, R., Králík, L., Zídek, L., and Sklenár, V. (2002) *J. Biomol. NMR* **24,** 1–14
36. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* **8,** 477–486
37. Vriend, G. (1990) *J. Mol. Graph.* **8,** 52–56, 29

38. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graph.* **14,** 51–55, 29–32
39. Heyduk, T., and Lee, J. C. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87,** 1744–1748
40. Sikorski, R. S., and Hieter, P. (1989) *Genetics* **122,** 19–27
41. Wilson, S. M., Datar, K. V., Paddy, M. R., Swedlow, J. R., and Swanson, M. S. (1994) *J. Cell Biol.* **127,** 1173–1184
42. San Paolo, S., Vanacova, S., Schenk, L., Scherrer, T., Blank, D., Keller, W., and Gerber, A. P. (2009) *PLoS Genet.* **5,** e1000555
43. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol.* **273,** 283–298
44. Swanson, M. S., Nakagawa, T. Y., LeVan, K., and Dreyfuss, G. (1987) *Mol. Cell. Biol.* **7,** 1731–1739
45. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1986) *Mol. Cell. Biol.* **6,** 2932–2943
46. Dreyfuss, G., Swanson, M. S., and Piñol-Roma, S. (1988) *Trends Biochem. Sci.* **13,** 86–91
47. Auweter, S. D., Oberstrass, F. C., and Allain, F. H. (2007) *J. Mol. Biol.* **367,** 174–186
48. Hargous, Y., Hautbergue, G. M., Tintaru, A. M., Skrisovska, L., Golovanov, A. P., Stevenin, J., Lian, L. Y., Wilson, S. A., and Allain, F. H. (2006) *EMBO J.* **25,** 5126–5137
49. Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L., and Allain, F. H. (2005) *Science* **309,** 2054–2057
50. Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001) *Cell* **106,** 595–605
51. Singh, R., Valcárcel, J., and Green, M. R. (1995) *Science* **268,** 1173–1176
52. Mulder, F. A., Schipper, D., Bott, R., and Boelens, R. (1999) *J. Mol. Biol.* **292,** 111–123
53. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) *Nucleic Acids Res.* **31,** 3497–3500
54. Spronk, C. A. E. M., Nabuurs, S. B., Krieger, E., Vriend, G., and Vuister, G. W. (2004) *Prog. Nucleic Magn. Res. Spectro.* **45,** 315–337
55. Hooft, R. W., Sander, C., and Vriend, G. (1997) *Comput. Appl. Biosci.* **13,** 425–430
56. Trantírek, L., Stefl, R., Masse, J. E., Feigon, J., and Sklenár, V. (2002) *J. Biomol. NMR* **23,** 1–12
57. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System,* DeLano Scientific, Palo Alto, CA

## 5.4   Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1

This article presents structure revealing a direct recognition of phoshorylated serine 5 (pSer5) of short CTD construct by the protein Nrd1 the protein. The Nrd1 requires for the CTD recognition *cis* conformation of the upstream pSer5–Pro6 peptidyl-prolyl bond of the CTD. Additionally, Nrd1 requires also phosphorylation of the RNAPII CTD. This suggests that the coupling of covalent and noncovalent changes in the CTD structure regulated by kinasesi / phosphatases and isomerases is crucial for the dynamic process of recruitment and displacement of appropriate processing factors during the transcriptional cycle. Furthermore, the paper shows that specific recognition of the pSer5 CTD by the Nrd1 CID is important for the processing and degradation of ncRNAs *in vivo*, suggesting that these events occur cotranscriptionally.

## RESEARCH COMMUNICATION

# Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1

Karel Kubicek,[1,3] Hana Cerna,[1,3] Peter Holub,[1] Josef Pasulka,[1] Dominika Hrossova,[1] Frank Loehr,[2] Ctirad Hofr,[1] Stepanka Vanacova,[1,4] and Richard Stefl[1,4]

[1]CEITEC-Central European Institute of Technology, Masaryk University, Brno, 62500, Czech Republic; [2]Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe-University Frankfurt, 60438 Frankfurt am Main, Germany

**Recruitment of appropriate RNA processing factors to the site of transcription is controlled by post-translational modifications of the C-terminal domain (CTD) of RNA polymerase II (RNAP II). Here, we report the solution structure of the Ser5 phosphorylated (pSer5) CTD bound to Nrd1. The structure reveals a direct recognition of pSer5 by Nrd1 that requires the *cis* conformation of the upstream pSer5–Pro6 peptidyl-prolyl bond of the CTD. Mutations at the complex interface diminish binding affinity and impair processing or degradation of noncoding RNAs. These findings underpin the interplay between covalent and noncovalent changes in the CTD structure that constitute the CTD code.**

Supplemental material is available for this article.

The C-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNAP II) consists of multiple tandem repeats of the heptapeptide consensus Tyr1–Ser2–Pro3–Thr4–Ser5–Pro6–Ser7 that is conserved from yeast to humans. The CTD is essential and forms a flexible tail of RNAP II. It serves as a binding platform for various cotranscriptional processing factors (Hirose and Manley 2000; Maniatis and Reed 2002; Meinhart et al. 2005). Phosphorylation and dephosphorylation of Ser2, Ser5, and Ser7 create a unique pattern in coordination with the transcription cycle (Komarnitsky et al. 2000; Meinhart et al. 2005; Phatnani and Greenleaf 2006; Kim et al. 2010; Mayer et al. 2010; Tietjen et al. 2010; Bataille et al. 2012). This phosphorylation pattern, often called the CTD code (Buratowski 2003, 2009; Chapman et al. 2008; Egloff and Murphy 2008), controls the recruitment, activation, and

displacement of various factors involved in transcription and RNAP II transcript processing (Meinhart et al. 2005). It has been proposed that the CTD code is also affected by noncovalent changes in the CTD structure, such as peptidyl-prolyl bond isomerization (Buratowski 2003); however, the structural basis for such tuning of the CTD code and its role in recruitment of RNA processing factors are not yet fully understood. The importance of the CTD isomerization for Ser5 dephosphorylation has recently been demonstrated for the human and insect Ssu72 phosphatases (Xiang et al. 2010; Werner-Allen et al. 2011).

The pSer5 marks occur predominantly in the early elongation phase and are essential in 3′ end processing of short noncoding genes (Komarnitsky et al. 2000; Gudipati et al. 2008; Vasiljeva et al. 2008a). These marks are specifically recognized by Nrd1 and are required for the Nrd1-dependent termination pathway, used at small nuclear/nucleolar RNAs (sn/snoRNAs), cryptic unstable transcripts (CUTs), and other short RNAP II transcripts (Steinmetz et al. 2001; Arigo et al. 2006; Thiebaut et al. 2006; Vasiljeva et al. 2008a). Recent genome-wide studies in yeast demonstrated the co-occurrence of pSer5 and pSer7 marks at some genes in the early elongation phase (Kim et al. 2010; Mayer et al. 2010; Tietjen et al. 2010). In addition to specific phosphorylation, the Nrd1 termination pathway requires the Ess1 (Pin1 in humans) peptidyl-prolyl isomerase (Singh et al. 2009) that specifically isomerizes the pSer5–Pro6 peptidyl-prolyl bond in the CTD (Gemmill et al. 2005). In yeast, Ess1 stimulates dephosphorylation of pSer5–Pro6 in vivo (Singh et al. 2009), and therefore it has been hypothesized that it regulates the Nrd1 association with the CTD (Singh et al. 2009). To gain insights into the recruitment process of Nrd1 to the 5′ regions of genes, we determined the solution structure of the *Saccharomyces cerevisiae* Nrd1 CTD-interacting domain (CID) in complex with a CTD peptide phosphorylated at Ser5. The Nrd1 CID structure reveals a conserved CTD-binding site that engages the β-turn motif of the CTD formed by $Ser2_b$–$Pro3_b$–$Thr4_b$–$pSer5_b$ and a site recognizing selectively the upstream $pSer5_a$ and the *cis* conformation of the $pSer5_a$–$Pro6_a$ peptidyl-prolyl bond of the CTD. Furthermore, we show that the specific recognition of pSer5 CTD by Nrd1 CID is important for the processing and degradation of noncoding RNAs (ncRNAs) in vivo.

## Results and Discussion

### Affinity of Nrd1 to the CTD with 'early' phosphorylation marks

A previous study demonstrated that Nrd1 favors binding to the CTD with "early" pSer5 marks over the CTD with "late" pSer2 marks (Vasiljeva et al. 2008a). To test the effect of the unphosphorylated CTD and the CTD with the "early" pSer7 and pSer5 CTD marks on the affinity to Nrd1, we performed a quantitative solution-binding assay using fluorescence anisotropy (FA) experiments. We found that Nrd1 binds the pSer5 CTD with a significantly stronger affinity compared with the pSer7 CTD or unphosphorylated CTD (Fig. 1A). Nrd1 shows also only a slightly weaker binding to the doubly phosphorylated pSer5/7 CTD than to the pSer5 CTD (Fig. 1A),
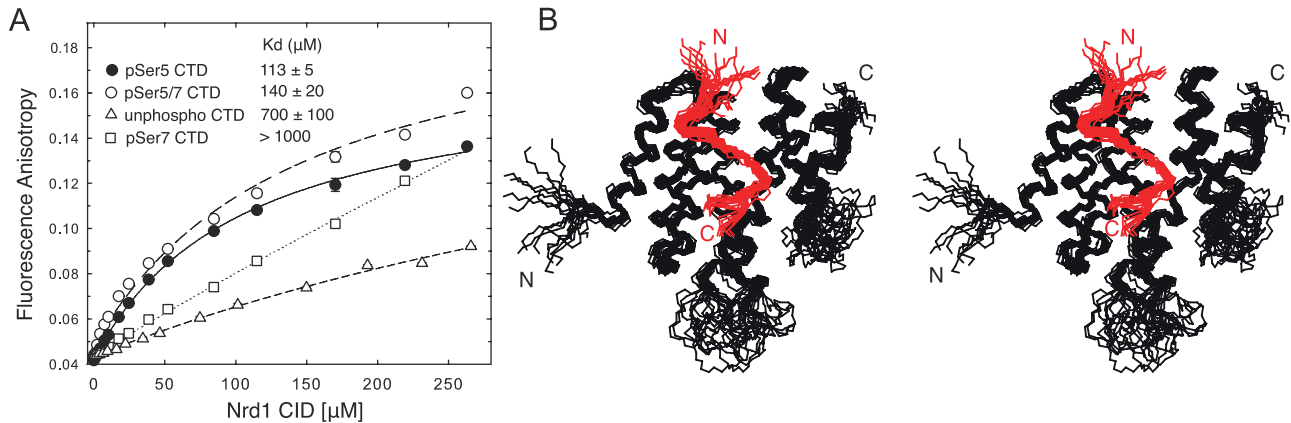
**Figure 1.** Structure of the Nrd1 CID–pSer5 CTD complex. (*A*) Equilibrium binding of the Nrd1 CID with differently phosphorylated CTD peptides monitored by FA. Binding isotherms and dissociation constants ($K_d$) are shown. (*B*) Overlay of the 20 lowest-energy structures of the Nrd1 CID–pSer5 CTD complex shown in stereo view. The backbone of the Nrd1 CID and pSer5 CTD is shown in black and red, respectively.

suggesting that the co-occurrence of the pSer5 and pSer7 marks may not impair the recruitment of the Nrd1 complex in the early elongation phase.

### Structure of Nrd1 CID bound to the phosphorylated CTD

To understand how the pSer5 CTD is recognized by Nrd1, we determined the solution structure of a reconstituted complex consisting of the CID (residues 1–153) of Nrd1 and a 14-amino-acid peptide, the pSer5 CTD (two repeats of the heptapeptide CTD consensus phosphorylated at Ser5; $Tyr1_a$–$Ser2_a$–$Pro3_a$–$Thr4_a$–$pSer5_a$–$Pro6_a$–$Ser7_a$–$Tyr1_b$–$Ser2_b$–$Pro3_b$–$Thr4_b$–$pSer5_b$–$Pro6_b$–$Ser7_b$) (Fig. 1B; Supplemental Table S1; Supplemental Fig. S1). The $^1H$, $^{13}C$, and $^{15}N$ chemical shift assignments for the bound Nrd1 CID were obtained as described previously (Kubicek et al. 2011). The structure of the Nrd1 CID is formed by eight α helices in a right-handed superhelical arrangement (Fig. 1B) and is virtually identical to the structure of the Nrd1 CID in the free form (Vasiljeva et al. 2008a). The pSer5 CTD peptide contacts helices α2, α4, and α7 of the Nrd1 CID (Fig. 2A).

### Recognition of the phosphorylated CTD by Nrd1

The CTD peptide adopts a β-turn conformation at $Ser2_b$–$Pro3_b$–$Thr4_b$–$pSer5_b$ and docks into a hydrophobic pocket of the Nrd1 CID that is formed by Ile29, Tyr67, Leu127, Ile130, and Met126 using $Tyr1_b$ and $Pro3_b$ residues (Fig. 2A,B). The hydroxyl group of $Tyr1_b$ forms a hydrogen bond with a conserved aspartate (Asp70) of Nrd1 (Fig. 2A,B). The binding mode of the peptide at the β-turn conformation resembles other previously determined structures of the CTD bound to CIDs of Pcf11, SCAF8, and Rtt103 (Supplemental Fig. S2; Meinhart and Cramer 2004; Becker et al. 2008; Lunde et al. 2010). However, in contrast to these CID–CTD complexes, Nrd1 binds more residues upstream of the pSer5 CTD via a conserved region at the N-terminal tip of helix α2 (Fig. 2A; Supplemental Fig. S3A). This unique region of Nrd1 is used to specifically recognize $pSer5_a$ via hydrogen bonding of Ser25 and Arg28 to the phosphate group of $pSer5_a$ (Fig. 2A,B). Another region that is more upstream in the α1–α2 loop has been previously suggested as the phosphoserine-

binding site of Nrd1 based on a sulfate ion that was found in the crystal structure of the free Nrd1 CID (Vasiljeva et al. 2008a). The sulfate ion located ~8 Å away from
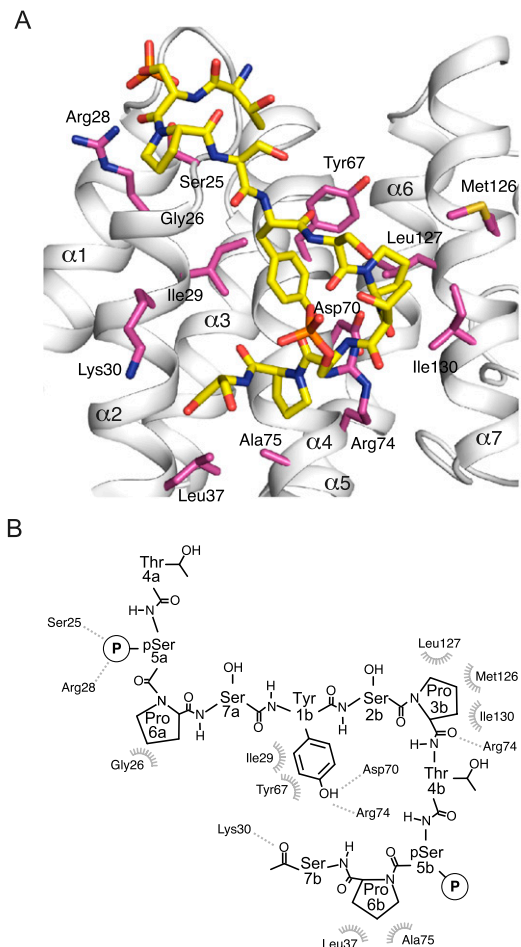


**Figure 2.** Recognition of the pSer5 CTD by Nrd1. (*A*) Scheme showing contacts between the Nrd1 CID and the CTD peptide. Protein residues that form hydrogen bonds and hydrophobic contacts to the CTD peptide are shown in white sticks. (*B*) Scheme showing contacts between the Nrd1 CID and the CTD peptide.

the phosphoserine location identified in our structure is coordinated using nonconserved amino acids. The structure of the Nrd1 CID–pSer5 CTD complex shows that the specific recognition of pSer5$_a$ is facilitated by the *cis* conformation of the pSer5$_a$–Pro6$_a$ peptidyl-prolyl bond. This *cis* conformation maximizes the intermolecular contacts and prevents the peptide from clashing with the α1–α2 loop of Nrd1. A conserved G26 in Nrd1 (all other CTD-containing proteins have a bulky and charged residues in this position) (Supplemental Fig. S3B) allows for loading of pSer5$_a$ into a highly electropositive pocket (Supplemental Fig. S3A). This is the first CID–CTD structure in which a phosphoserine–proline bond is observed in the *cis* conformation and in which a direct recognition of pSer5 is found. Importantly, both features are interconnected and thus required for the efficient binding. A similar conformation of the pSer5 CTD peptide was found in the crystal structure of the Ssu72–pSer5 CTD complex (Supplemental Fig. S4; Xiang et al. 2010; Werner-Allen et al. 2011).

## Interaction between Nrd1 and the CTD is important for cell viability and the processing or degradation of ncRNAs

Specific association of the pSer5 CTD with the Nrd1 CID was further tested in a quantitative in vitro binding assay using FA. We titrated the wild-type and mutant Nrd1 CID against the fluorescently labeled pSer5 CTD. Alanine or aspartate (charge-swapping) substitutions at positions Ser25, Gly26, Arg28, Ile29, and Lys30 significantly decreased the binding affinity with the pSer5 CTD (Fig. 3A; Supplemental Fig. S6, control mutations of nonessential residues). In comparison, the effect of mutants at nonconserved positions Leu20, Lys21, and Ser22 in the region that was previously suggested to bind the phosphoserine (see above) is much smaller (Vasiljeva et al. 2008a). D70R and R74D variants of Nrd1 could not be assayed due to their instability at the high concentrations required for FA measurements. Next, we tested the effect of removal of phosphorylation in the downstream CTD repeat (at Ser5$_b$). We found that the CTD peptide with a single phosphorylation [pSer5(1P)] has the same affinity for Nrd1 as the CTD peptide phosphorylated at both Ser5s (Supplemental Fig. S5). Furthermore, mutations at Pro6$_a$ in the CTD peptide (P6$_a$A CTD and P6$_a$R CTD) have a larger negative effect on affinity to Nrd1 than mutations at Pro6$_b$ (P6$_b$A CTD and P6$_b$R CTD), confirming the requirement of the *cis* conformation at pSer5$_a$–Pro6$_a$ for the binding to Nrd1 (Supplemental Fig. S5). The effect of the P6$_a$R mutation is larger than for P6$_a$A, as the bulkier side chain (the side chain at position 6$_a$ is solvent-exposed) creates more unfavorable interactions with the adjacent side chain of pSer5, decreasing the stability of the *cis* conformation. This corroborates previous studies that showed that proline-to-alanine mutations do not necessarily alter the *cis* conformation if
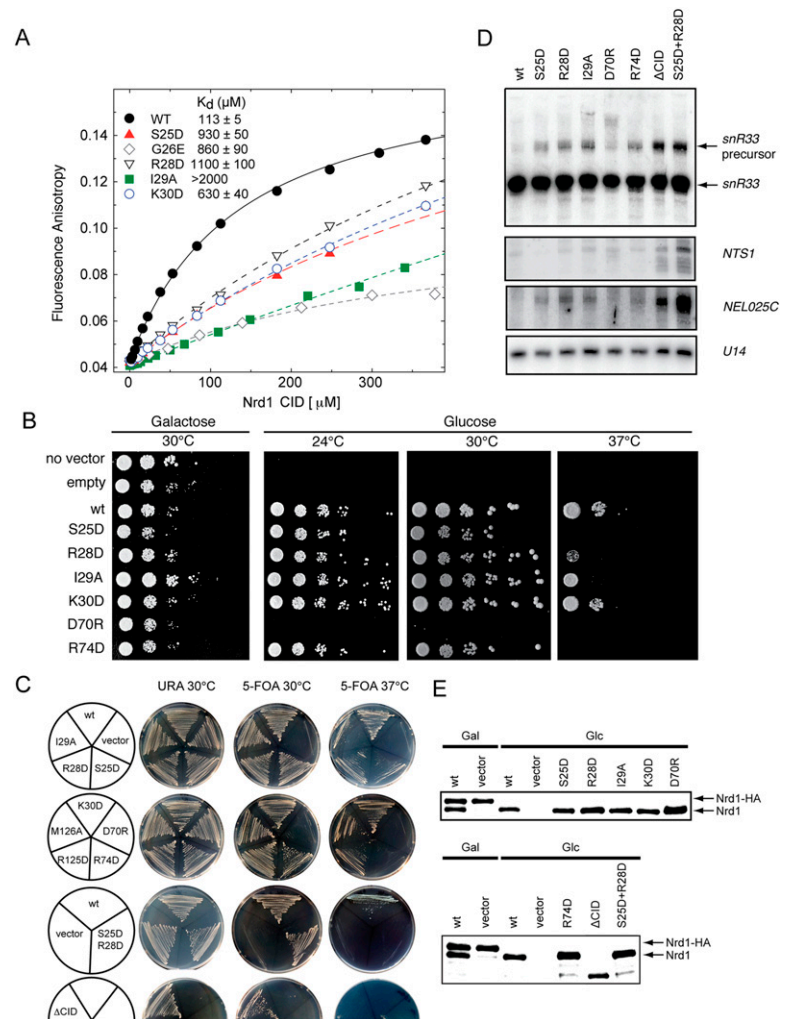


**Figure 3.** Critical residues of the Nrd1 CID that are required for CTD binding, cell viability, and RNA processing and degradation. (*A*) Equilibrium binding of the Nrd1 CID mutants with the pSer5 CTD peptide monitored by FA. Binding isotherms and dissociation constants ($K_d$) are shown for individual mutants. (*B*) Phenotypic analysis of the Nrd1 CID mutants. The mutants were expressed from *pRS415* plasmids in the *GAL1::NRD1* strain background. Growth on glucose-containing plates leads to the repression of *GAL1*-driven wild-type Nrd1 and thus shows the functionality of the different Nrd1 CID mutants. (*C*) Phenotypic analysis of the Nrd1 CID mutants. The mutants were expressed episomally from *pRS415* plasmids in the yeast strain where the endogenous *NRD1* was deleted and growth was supplemented with *NRD1* on the *URA3* plasmid (*pRS316*). To test the functionality of the different Nrd1 CID mutants, cells were grown on 5-FOA-containing plates for 3 d at the indicated temperatures. Wild-type *NRD1* was used as positive control, empty *pRS415* plasmid was used as negative control, and Nrd1$_{\Delta1–150}$ was the ΔCID. (*D*) *snR33* snoRNA processing efficiency and stability of *NTS1* and *NEL025C* CUTs analyzed by Northern blot analysis. Mature *U14* snoRNA represents a loading control. Total RNA was purified from cells expressing wild-type *NRD1* or the indicated mutants grown in glucose-containing medium. (*E*) Western blot analysis of expression levels of the wild-type and mutant Nrd1 originating from *pRS415* (faster-migrating band) in the yeast strain where the endogenous *NRD1* is under the galactose promoter. Protein extracts were prepared from the original *GAL1::NRD1* strain transformed with plasmids carrying wild-type and mutant *NRD1* grown in either galactose-containing medium (Gal) or glucose-containing medium (Glc). (*Top* band) The genomic *NRD1* copy contains a fusion HA tag, resulting in slower gel migration. Proteins were detected using specific antibodies against Nrd1p. Vectors with no insertion (vector) or containing the wild-type *NRD1* were used as negative and positive controls, respectively.

it is enforced by the structural context (Mayr et al. 1994; Xiong et al. 2000). Altogether, the FA data strongly support the phosphorylation-specific recognition ob-

Kubicek et al.

served in the structure of the Nrd1–pSer5 CTD complex and that the interaction relies on the presence of the invariant basic residues in the CID domain (Supplemental Fig. S3A).

To determine the importance of these individual residues for Nrd1 function in vivo, we monitored cell viability and ncRNA processing/stability in single-amino-acid mutants (Fig. 3B–D). The Nrd1 strain lacking the CID (Nrd1Δ1–150) was used as a reference for the CID-related function. Deletion of the CID is not lethal (Vasiljeva et al. 2008a); however, we observed that deletion of the CID led to inviability at 37°C (Fig. 3C). Similarly, to a lesser extent, a temperature-sensitive (ts) growth defect was observed in mutants of Ser25, Arg28, Ile29, and Arg74 (Fig. 3B,C). The double mutant in the residues contacting the phosphorylated serine (Ser25+Arg28) exhibited the same growth phenotype as ΔCID Nrd1 (Fig. 3C). Nrd1 CID deletion causes an accumulation of snR33 precursors in vivo (Vasiljeva et al. 2008a). We observed that point mutants with the ts growth phenotype showed snoRNA processing and CUT degradation defects demonstrated by an accumulation of pre-snR33 snoRNA and NEL025c and NTS1 CUTs, respectively (Fig. 3D,E; Supplemental Fig. S7). None of the Nrd1 CID mutants tested displayed transcription termination defects that would be represented by readthrough product accumulation. Importantly, the Ser25+Arg28 double mutant showed processing and degradation defects comparable with those of the mutant lacking the entire Nrd1 CID. Taken together, these data demonstrate that specific recognition of the pSer5 CTD by the Nrd1 CID is important for the processing and degradation of ncRNAs in vivo.

The Nrd1 complex associates with the exosome and the TRAMP complex (Vasiljeva and Buratowski 2006). Based on our data, it is tempting to speculate that the Nrd1 CID mediates interaction with this processing and degradation apparatus. As the Nrd1 truncation lacking CID (Δ39–169) has no effect on the exosome copurification (demonstrated for Rrp6) (Vasiljeva and Buratowski 2006), it is likely that other RNA processing and degradation auxiliary factors, such as the TRAMP complex (LaCava et al. 2005; Vanacova et al. 2005), are recruited through the Nrd1 CID.

Previous chromatin immunoprecipitation (ChIP) experiments showed that Ess1 promotes the release of Nrd1 from terminator regions (Singh et al. 2009). This observation has been attributed to the indirect effect of Ess1 in which it stimulates dephosphorylation by Ssu72 (Singh et al. 2009). Ssu72 targets specifically the cis conformation of the peptidyl-prolyl bond of the pSer-Pro-containing peptides (Xiang et al. 2010; Werner-Allen et al. 2011). Here we suggest that Ess1 can also directly regulate the association of Nrd1 with the RNA Pol II CTD, as it specifically recognizes the pSer5–Pro6 CTD in the cis conformation, indicating that Ess1 may play a dual role in regulating the Nrd1 pathway. A detailed understanding of this mechanism will require further studies, but it is an exciting possibility, particularly in light of recent observations that suggested that Ssu72 may be a less "conformation-specific" phosphatase (acting also on the pSer7–Tyr1 CTD peptide that is unlikely to exist in the cis conformation) than previously expected (Bataille et al. 2012; Zhang et al. 2012).

## Conclusions

The structure of Nrd1 CID–pSer5 CTD presented here reveals that the CTD recognition by Nrd1 requires both phosphorylation and isomerization of the RNAPII CTD.

This suggests that the coupling of covalent and noncovalent changes in the CTD structure regulated by kinases/phosphatases and isomerases is crucial for the dynamical process of recruitment and displacement of appropriate processing factors during the transcriptional cycle. In addition, we show that specific recognition of the pSer5 CTD by the Nrd1 CID is important for the processing and degradation of ncRNAs in vivo, suggesting that these events occur cotranscriptionally.

## Materials and methods

### Protein expression and mutagenesis

The DNA encoding the Saccharomyces cerevisiae Nrd1 CID domain (residues 1–153) was amplified and cloned into a pET22b expression vector (Novagen) via NdeI and XhoI restriction sites. Details on cloning, expression, and purification of the Nrd1 CID construct have been described previously (Kubicek et al. 2011). Protein mutants were designed on the basis of the NMR structure of the Nrd1–pSer5 CTD complex and were prepared using the QuikChange site-directed mutagenesis kit (Stratagene).

### The CTD of RNAP II

It has been established previously that the CTD mimic consisting of two repeats of the CTD canonical heptad yields the same binding affinity to Nrd1 CID as the CTD mimic of four repeats (Vasiljeva et al. 2008a). Thus, we used a 14-amino-acid peptide, the pSer5 CTD (two repeats of the heptapeptide CTD consensus phosphorylated at two Ser5s; $Tyr1_a–Ser2_a–Pro3_a–Thr4_a–pSer5_a–Pro6_a–Ser7_a–Tyr1_b–Ser2_b–Pro3_b–Thr4_b–pSer5_b–Pro6_b–Ser7_b$), in our study to mimic the CTD phosphorylated at Ser5. Similarly, the unphosphorylated CTD, the pSer7 CTD, and the doubly phosphorylated pSer5/7 CTD were used. The peptides were purchased from Clonestar Peptide Services.

### NMR

All NMR spectra for the backbone and side chain assignments of 2.0 mM uniformly $^{15}N,^{13}C$-labeled Nrd1 CID in 50 mM sodium phosphate buffer (pH 8.0), 100 mM NaCl, and 10 mM β-mercaptoethanol (90% $H_2O$/10% $D_2O$) were recorded on Bruker AVANCE 600- and 950-MHz spectrometers equipped with a cryoprobe at a sample temperature of 20°C. The spectra were processed using an NMRPipe package (Delaglio et al. 1995), and the protein resonances were assigned manually using Sparky software (T.G. Goddard and D.G. Kellner, University of California at San Francisco). The $^1H$, $^{13}C$, and $^{15}N$ chemical shifts of the bound form of the Nrd1 CID were assigned as described elsewhere (Kubicek et al. 2011). All distance constraints were derived from the three-dimensional (3D) $^{15}N$- and $^{13}C$-separated NOESYs and two-dimensional (2D) $^1H$-$^1H$ NOESY (with a mixing time of 80 msec) collected on a 950-MHz spectrometer. Intermolecular distance constraints were obtained from the 3D $F_1$-$^{13}C/^{15}N$-filtered NOESY-[$^{13}C,^1H$]-HSQC experiment (Zwahlen et al. 1997; Peterson et al. 2004), with a mixing time of 150 msec on a 950-MHz spectrometer. Intramolecular distance constraints of the bound CTD peptide (unlabeled) were derived from a 2D $F_1,F_2$-$^{13}C/^{15}N$-filtered [$^1H,^1H$]-NOESY ($τ_m$ = 150 msec) (Zwahlen et al. 1997; Peterson et al. 2004). The NOEs were semiquantitatively classified based on their intensities in the 2D and 3D NOESY spectra.

### Structure calculations

The preliminary structure determinations of the Nrd1–pSer5 CTD complex were performed with the automated NOE assignment module implemented in the CYANA program (Guntert 2004). In the next step, CYANA-generated restraints along with manually assigned protein–CTD intermolecular restraints were used for further refinement of the preliminary structures with AMBER 10.0 software (Case et al. 2005). These calculations used a modified version (AMBER ff99SB) of the force field described by Cornell et al. (1995) using a protocol described previously (Stefl et al. 2010; Hobor et al. 2011). From 40 refined structures, the 20 conformers with the lowest AMBER energy were selected to form the final ensemble of structures. Molecular graphics were generated using MOLMOL (Koradi et al. 1996) and PyMOL (http://www.pymol.org). The

atomic coordinates and restraints for the Nrd1 CID–pSer5 CTD complex have been deposited in the Protein Data Bank under ID code 2lo6.

### FA

The equilibrium binding of the Nrd1 CID to the differently phosphorylated CTD was analyzed by FA. The CTD peptides were N-terminally labeled with the 5,6-carboxyfluorescein (FAM). The measurements were conducted on a FluoroMax-4 spectrofluorometer (Horiba Jobin-Yvon). The instrument was equipped with a thermostatted cell holder with a Neslab RTE7 water bath (Thermo Scientific). Samples were excited with vertically polarized light at 477 nm, and both vertical and horizontal emissions were recorded at 525 nm. All measurements were conducted at 10°C in 50 mM phosphate buffer (pH 8.0) containing 100 mM NaCl and 10 mM β-mercaptoethanol. Each data point is an average of five measurements. The experimental binding isotherms were analyzed by nonlinear least-squares regression in SigmaPlot 11 software (Systat Software) using a single-site binding model according to Heyduk and Lee (1990).

### Construction of yeast plasmids

The *pRS415* plasmid (*CEN*, *LEU2*) with insertion of the wild-type *NRD1* gene surrounded by the *NRD1* promoter and terminator (Vasiljeva et al. 2008a) was used as a template for QuikChange site-directed mutagenesis (Stratagene). See Supplemental Table S2 for primer sequences and Supplemental Table S3 for constructs generated in this study. The *NRD1* ΔCID region was amplified as follows: Fragment 1: −340 nt up to +6 nt of *NRD1* with SVO F71 and SVO F72; Fragment 2: +453 nt up to TAA +300 nt from the 3′ untranslated region (UTR) with SVO F73 and SVO F74 primers. The two resulting PCR products were ligated together and inserted into the above-mentioned *pRS415*-based construct.

### Yeast cultures and manipulation

Yeast were cultured under standard conditions in media with selective markers corresponding to particular strains and vectors. Yeast transformations were performed by the lithium acetate method.

### Yeast growth test analysis

W303 (*GAL1::NRD1*)-derived strains (Supplemental Table S4) carrying appropriate mutant *NRD1* plasmids were grown in SD-LEU-HIS + 2% galactose at 30°C to an O.D. of 1.0. The cultures were serially diluted by a factor of 10 and spotted onto SD-LEU-HIS medium containing 2% glucose to repress the expression of the endogenous *NRD1* or control medium (SD-HIS + 2% galactose). Plates were incubated at 25°C, 30°C, and 37°C. EJS101-9d-derived strains (Supplemental Table S4) were grown for 3 d on SC-LEU plates, then spread on SC-LEU plates with or without 5-FOA and incubated at 25°C, 30°C, and 37°C.

### Whole-cell protein extract preparation for Western blot analysis

Protein extracts were prepared from cultures grown on either galactose- or glucose-containing medium to an O.D. of 1.0. Five milliliters of culture was harvested and lyzed by 1.85 M NaOH for 15 min on ice, and proteins were subsequently precipitated with ice-cold trichloroacetic acid. Pellets were resuspended in 5% SDS and 8 M urea buffer prior to SDS-PAGE analysis. Proteins were resolved on a 12% SDS-PAGE gel, transferred to a nitrocellulose membrane by a semidry electroblotter (Bio-Rad), and probed for the presence of Nrd1p with the anti-Nrd1 sera kindly provided by David Brow (Steinmetz and Brow 1998).

### RNA isolation and analysis

For RNA analysis, cells were inoculated to an O.D. of 0.1 in glucose- or galactose-containing SD medium and grown for 16 h at 30°C. RNA was isolated by hot phenol extraction and stored at −80°C. Five micrograms of total RNA was denatured in 25% formamide, separated on an 8% denaturing (8 M urea) polyacrylamide gel, and transferred to nylon membrane using semidry electro-transfer. RNA was cross-linked to the membrane by UV light (120 mJ/cm$^2$) and hybridized with a probe in Ultra-Hyb buffer (Ambion) according to the manufacturer's instructions.

### Preparation of DNA probes for Northern blot analysis

The DNA probe for *snR33* was amplified from *S. cerevisiae* S288C genomic DNA with primers Forward, 5′-CGGAACGGTACATAAGAA TAGAAGAG-3′, and Reverse, 5′-TAAAGAAAACGATAAGAACTAA CCTC-3′. The *NTS1* probe 1 was prepared according to Vasiljeva et al. (2008b), by using primers Forward, 5′-TGAGTGCTTGTATAAGTTTA GAGAATTGA-3′, and Reverse, 5′-TTAATACTTTCCTCTTCGTCTTTT TCTAC-3′. The *NEL025c* probe was amplified with primers Forward, 5′-CCTGTTGACATTGCAGACAA-3′, and Reverse, 5′-GCAAAGATCTG TATGAAAGG-3′. The resulting PCR products were used as templates for random primed labeling using [α-$^{32}$P]dATP and the commercial kit (Roche). To detect *U14* snoRNA, the oligonucleotide 5′-TCACTCAGACATCC TAGG-3′ was 5′-phosphate-labeled by T4 polynucleotide kinase (New England Biolabs) and [γ-$^{32}$P]ATP.

## References

Arigo JT, Eyler DE, Carroll KL, Corden JL. 2006. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell* **23:** 841–851.

Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F. 2012. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* **45:** 158–170.

Becker R, Loll B, Meinhart A. 2008. Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **283:** 22659–22669.

Buratowski S. 2003. The CTD code. *Nat Struct Biol* **10:** 679–680.

Buratowski S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36:** 541–546.

Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. 2005. The Amber biomolecular simulation programs. *J Comput Chem* **26:** 1668–1688.

Chapman RD, Heidemann M, Hintermair C, Eick D. 2008. Molecular evolution of the RNA polymerase II CTD. *Trends Genet* **24:** 289–296.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J Am Chem Soc* **117:** 5179–5197.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biol NMR* **6:** 277–293.

Egloff S, Murphy S. 2008. Cracking the RNA polymerase II CTD code. *Trends Genet* **24:** 280–288.

Gemmill TR, Wu X, Hanes SD. 2005. Vanishingly low levels of Ess1 prolyl-isomerase activity are sufficient for growth in *Saccharomyces cerevisiae*. *J Biol Chem* **280:** 15510–15517.

Gudipati RK, Villa T, Boulay J, Libri D. 2008. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat Struct Mol Biol* **15:** 786–794.

Guntert P. 2004. Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278:** 353–378.

Heyduk T, Lee JC. 1990. Application of fluorescence energy transfer and polarization to monitor *Escherichia coli* cAMP receptor protein and *lac* promoter interaction. *Proc Natl Acad Sci* **87:** 1744–1748.

Hirose Y, Manley JL. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev* **14:** 1415–1429.

Hobor F, Pergoli R, Kubicek K, Hrossova D, Bacikova V, Zimmermann M, Pasulka J, Hofr C, Vanacova S, Stefl R. 2011. Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J Biol Chem* **286:** 3645–3657.

Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, Megee PC, Bentley DL. 2010. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* **17:** 1279–1286.

Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14:** 2452–2460.

Koradi R, Billeter M, Wuthrich K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graph* **14:** 51–55.

Kubicek K, Pasulka J, Cerna H, Lohr F, Stefl R. 2011. 1H, 13C, and 15N resonance assignments for the CTD-interacting domain of Nrd1 bound to Ser5-phosphorylated CTD of RNA polymerase II. *Biomol NMR Assign* **5:** 203–205.

LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121:** 713–724.

Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **17:** 1195–1201.

Maniatis T, Reed R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416:** 499–506.

Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P. 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* **17:** 1272–1278.

Mayr LM, Willbold D, Rösch P, Schmid FX. 1994. Generation of a non-prolyl *cis* peptide bond in ribonuclease T1. *J Mol Biol* **240:** 288–293.

Meinhart A, Cramer P. 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature* **430:** 223–226.

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. 2005. A structural perspective of CTD function. *Genes Dev* **19:** 1401–1415.

Peterson RD, Theimer CA, Wu H, Feigon J. 2004. New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA–protein complexes. *J Biol NMR* **28:** 59–67.

Phatnani HP, Greenleaf AL. 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20:** 2922–2936.

Singh N, Ma Z, Gemmill T, Wu X, Defiglio H, Rossettini A, Rabeler C, Beane O, Morse RH, Palumbo MJ, et al. 2009. The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. *Mol Cell* **36:** 255–266.

Stefl R, Oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L, Maris C, Peng L, Hofr C, Emeson RB, et al. 2010. The solution structure of the ADAR2 dsRBM–RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **143:** 225–237.

Steinmetz EJ, Brow DA. 1998. Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. *Proc Natl Acad Sci* **95:** 6699–6704.

Steinmetz EJ, Conrad NK, Brow DA, Corden JL. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3′-end formation of RNA polymerase II transcripts. *Nature* **413:** 327–331.

Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D. 2006. Transcription termination and nuclear degradation of cryptic unstable transcripts: A role for the nrd1–nab3 pathway in genome surveillance. *Mol Cell* **23:** 853–864.

Tietjen JR, Zhang DW, Rodriguez-Molina JB, White BE, Akhtar MS, Heidemann M, Li X, Chapman RD, Shokat K, Keles S, et al. 2010. Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* **17:** 1154–1161.

Vanacova S, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W. 2005. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol* **3:** e189. doi: 10.1371/journal.pbio.0030189.

Vasiljeva L, Buratowski S. 2006. Nrd1 interacts with the nuclear exosome for 3′ processing of RNA polymerase II transcripts. *Mol Cell* **21:** 239–248.

Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. 2008a. The Nrd1–Nab3–Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **15:** 795–804.

Vasiljeva L, Kim M, Terzi N, Soares LM, Buratowski S. 2008b. Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. *Mol Cell* **29:** 313–323.

Werner-Allen JW, Lee CJ, Liu P, Nicely NI, Wang S, Greenleaf AL, Zhou P. 2011. *Cis*-proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. *J Biol Chem* **286:** 5717–5726.

Xiang K, Nagaike T, Xiang S, Kilic T, Behav MM, Manley JL, Tong L. 2010. Crystal structure of the human symplekin–Ssu72–CTD phosphopeptide complex. *Nature* **467:** 729–733.

Xiong Y, Juminaga D, Swapna GV, Wedemeyer WJ, Scheraga HA, Montelione GT. 2000. Solution NMR evidence for a *cis* Tyr–Ala peptide group in the structure of [Pro93Ala] bovine pancreatic ribonuclease A. *Protein Sci* **9:** 421–426.

Zhang DW, Mosley AL, Ramisetty SR, Rodríguez-Molina JB, Washburn MP, Ansari AZ. 2012. Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *J Biol Chem* **287:** 8541–8551.

Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE. 1997. Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage λ N-peptide/boxB RNA complex. *J Am Chem Soc* **119:** 6711–6721.

## 5.5   Recognition of asymmetrically dimethylated arginine by TDRD3

The results presented in this study extend our understanding of how the methylarginine binding Tudor domains coordinate their ligands. The publication reports the solution structure of the Tudor domain of TDRD3 bound to the asymmetrically dimethylated CTD. A comparison of the ligand-bound TDRD3 structure with other Tudor domains, supported by the mutational analysis and quantum chemical calculations provide a framework to understand dimethylarginine recognition by the Tudor domains at a molecular level.

# Recognition of asymmetrically dimethylated arginine by TDRD3

Tomas Sikorsky, Fruzsina Hobor, Eva Krizanova, Josef Pasulka, Karel Kubicek and Richard Stefl*

CEITEC—Central European Institute of Technology, Masaryk University, CZ-62500 Brno, Czech Republic

## ABSTRACT

Asymmetric dimethylarginine (aDMA) marks are placed on histones and the C-terminal domain (CTD) of RNA Polymerase II (RNAP II) and serve as a signal for recruitment of appropriate transcription and processing factors in coordination with transcription cycle. In contrast to other Tudor domain-containing proteins, Tudor domain-containing protein 3 (TDRD3) associates selectively with the aDMA marks but not with other methylarginine motifs. Here, we report the solution structure of the Tudor domain of TDRD3 bound to the asymmetrically dimethylated CTD. The structure and mutational analysis provide a molecular basis for how TDRD3 recognizes the aDMA mark. The unique aromatic cavity of the TDRD3 Tudor domain with a tyrosine in position 566 creates a selectivity filter for the aDMA residue. Our work contributes to the understanding of substrate selectivity rules of the Tudor aromatic cavity, which is an important structural motif for reading of methylation marks.

## INTRODUCTION

Arginine methylation is a frequent post-translational modification of proteins that regulates a variety of cellular processes, including transcriptional regulation, RNA processing, trafficking, signal transduction and DNA repair (1–3). There are three major forms of methylated arginine identified in mammals, monomethylarginine (MMA), asymmetric dimethylarginine (aDMA) and symmetric dimethylarginine (sDMA). These methylation marks are introduced by the protein arginine methyltransferases' (PRMTs) family, in which, type I PRMTs (PRMT1, 2, 3, 4, 6 and 8) generate MMA and aDMA modifications, whereas type II PRMTs (PRMT5 and 7) produce MMA and sDMA modifications. The MMA modifications introduced by both type I and type II PRMTs are likely generated as 'intermediates' on the way to dimethylarginines. PRMTs methylate a large number of protein targets, involved in various aspects of regulation of gene expression (2).

The co-activator-associated arginine methyltransferase 1 (CARM1/PRMT4) deposits an asymmetric dimethylation at the R2 and R17 sites of histone H3 (H3R17 and H3R2) and at the R3 site of histone H4 (H4R3) (4). It also introduces the aDMA mark at R1810 of the mammalian carboxy-terminal domain (CTD) of RNA Polymerase II (RNAP II) (5). The CTD is an important region of RNAP II that undergoes structural remodeling throughout the transcriptional cycle, which allows the association and dissociation of a multitude of effector molecules (6–11). These temporal and spatial interactions couple transcription with most, if not all, pre-mRNA processing steps (12–24). The CTD methylation at R1810 is present on the hyperphosphorylated CTD in vivo, and it facilitates the expression of small nuclear and nucleolar RNAs (snRNAs and snoRNAs) (5).

Pull-down experiments showed that Tudor domain-containing protein 3 (TDRD3) displays a specific interaction with the asymmetrically dimethylated R1810-containing CTD peptide (aDMA-CTD) and histones H3 and H4, but not with unmodified and monomethylated or symmetrically dimethylated peptides (4,5). Consistently, using fluorescence polarization, it was reported that TDRD3 preferentially binds the aDMA marks compared with other methylarginine species (25). Other members of the Tudor family, such as the Survival of Motor Neuron (SMN), survival of motor neuron-related splicing factor 30 (SPF30), staphylococcal nuclease domain-containing protein 1 (SND1) and Tur11 do not discriminate between the aDMA- and sDMA-containing peptides (26–28). They are capable of binding to both dimethylarginine isoforms, with slightly higher affinity to sDMA, which also represents their physiological ligand (29,30). The structures of the Tudor domains of SMN, SPF30, SND1 and Tur11 bound to their ligands have been reported (26–28).

In order to reveal the structural basis of selective recognition of the aDMA marks by TDRD3 that has recently been implicated in transcription activation and RNA processing (4,5), we have determined the solution structure of

*To whom correspondence should be addressed. Tel: +42 054 94 92 436; Fax: +42 054 94 92 556; Email: richard.stefl@ceitec.muni.cz

the Tudor domain of TDRD3 in complex with asymmetrically dimethylated CTD.

## MATERIALS AND METHODS

### Cloning, expression and purification of Tudor domain of TDRD3 (residues 554–608)

The Tudor domain of TDRD3 (residues 554–608) was cloned into pET22b expression vector to generate C-terminal 6x His-tagged protein. The protein was over-expressed in *Escherichia coli* BL21-Codon Plus (DE3)-RIPL (Stratagene) overnight at 16°C after induction by 1 mM IPTG and purified by affinity chromatography on Ni-NTA resin (Qiagen) and further purified on Superdex75 gel filtration column. For nuclear magnetic resonance (NMR) experiments, the protein was concentrated into a buffer containing 50 mM $Na_2HPO_4$, pH 8.0, 150 mM NaCl and 10 mM β-mercaptoethanol. Mutant constructs were prepared by QuikChange$^{TM}$ Site-directed Mutagenesis Kit with complementary sense (S) and antisense (AS) oligonucleotide primers as follows:

Y566F  (S), 5′-cctggagatgaatgtttttgcacttttttgggaagacaaca-3′
(AS), 5′-tgttgtcttcccaaaaaagtgcaaaacattcatctccagg-3′;

Y566W  (S), 5′-gaaacctggagatgaatgttttgcactttggtgggaagacaacaag-3′
(AS), 5′-cttgttgtcttcccaccaaagtgcaaaacattcatctccaggtttc-3′;

W567S  (S), 5′-cctggagatgaatgttttgcactttattcggaagacaacaagtt-3′
(AS), 5′-aacttgttgtcttccgaataaagtgcaaaacattcatctccagg-3′;

D569A  (S), 5′-gcactttattgggaagcgaacaagtttaccgggc-3′
(AS) 5′-gcccggtaaaacttgttcgcttcccaataaagtgc-3′.

The mutant proteins were over-expressed in the same *E. coli* line and purified using the same procedure as described above. Peptides for NMR measurements were purchased from Clonestar peptide service (Brno, CZ).

### Nuclear magnetic resonance

All spectra for the backbone and side-chain assignments of ~1.8 mM uniformly $^{15}$N,$^{13}$C-labeled TDRD3 Tudor in 50 mM sodium phosphate buffer (pH 8.0), 150 mM NaCl, 10 mM β-mercaptoethanol (90% $H_2O$/10% $D_2O$) were recorded on Bruker AVANCE 600 MHz spectrometer equipped with a cryoprobe at a sample temperature of 293.15 K. The spectra were processed using NMRPipe package (31) and the protein resonances were assigned manually, using Sparky software (Goddard T.G. and Kellner D.G., University of California, San Francisco, USA). The $^1$H, $^{13}$C and $^{15}$N chemical shifts of TDRD3–aDMA-CTD complex were assigned using standard triple resonance experiments (32). All distance constraints were derived from the 3D $^{15}$N- and $^{13}$C-separated nuclear Overhauser effect spectroscopy (NOESYs) (with mixing time of 110 ms) collected on a 600 MHz spectrometer. Intermolecular distance constraints were obtained from the 3D $F_1$-$^{13}$C/$^{15}$N-filtered NOESY-[$^{13}$C,$^1$H]-HSQC

experiment (33,34), with mixing time of 150 ms on a 600 MHz spectrometer. In the 3D NOESY spectra, intermolecular constraints were semi-quantitatively classified, based on their peak volumes divided by the number of $^1$H spins involved in nuclear Overhauser effect (NOE) cross-relaxation.

### Structure calculation

Structure determination of the TDRD3 Tudor–aDMA-CTD complex was performed with the NOE assignment algorithm implemented in the CYANA program (35). This automated NOE assignment procedure is a re-implementation of the former CANDID algorithm (36) on the basis of a probabilistic treatment of the NOE assignment. CYANA carries out automated assignments, distance calibration of NOE intensities, removal of erroneous restraints, structure calculations with torsion angle dynamics. The resultant NOE cross-peak assignments were subsequently confirmed by visual inspection of the spectra. In the next step, CYANA-generated restraints along with manually assigned TDRD3–aDMA-CTD intermolecular restraints were used for further refinement of 20 preliminary structures with AMBER 11 software (37,38). These calculations employed a modified version (AMBER ff99SB) of the force field described by Cornell *et al.* (39) and an explicit water solvent. The non-standard aDMA residue was parameterized with Gaussian 09 using restrained electrostatic potential charge approach at HF/6-31G(d) level of theory (40). The compound name for aDMA is DA2. Structural quality was assessed using PROCHECK (41) and WHAT IF (42).

### Fluorescence anisotropy

The equilibrium binding of Tudor domain of TDRD3 to the CTD peptides was analyzed by fluorescence anisotropy. The CTD peptides were N-terminally labeled with 5,6-carboxyfluorescein (FAM). The measurements were conducted on a FluoroMax-4 spectrofluorometer (Horiba Jobin-Yvon Edison, NJ, USA). Samples were excited with vertically polarized light at 495 nm, and both vertical and horizontal emissions were recorded at 535 nm. All measurements were conducted at 293.15 K in 50 mM phosphate buffer (pH 8.0) containing 150 mM NaCl. A fixed delay of 30 s was set between each aliquot addition and start of the measurement to allow the reaction to reach equilibrium. This delay was sufficient, as no further change in anisotropy was observed. Each data point is an average of three measurements. The data were analyzed using Gnuplot. The experimental isotherms were fit to a single-site binding model using non-linear least squares regression according to Heyduk and Lee (43).

### NMR binding experiments

$^1$H-$^{15}$N Heteronuclear single quantum coherence (HSQC) spectra were recorded at same conditions as NOESY experiments. To determine the affinity of different TDRD3 mutants with aDMA- and sDMA-containing peptides, a series of $^1$H-$^{15}$N HSQC spectra were recorded. Increasing

amounts of the unlabeled dimethylarginine-containing peptides were added to $^{15}$N-labeled proteins. Combined chemical shift (CCS) is defined as the normalized length of a vector $E_j$, whose components are chemical shift differences $\delta_{ji}$ between observed chemical shift and reference experiment (free form). Index $j$ represents the atom type within the primary sequence of the protein

$$|E_j| = \sqrt{\sum_{i=H,N} w_i \delta_{ji}^2}.$$

Weight factors for each atom type $w_H = 1$ and $w_N = 0.15$ were used. CCS values of three best resolved residues for each titration were used to construct a binding isotherm of 1:1 stoichiometry. Parameters for binding isotherms were found by non-linear least-square regression with Gnuplot. The errors for fitted parameters were calculated within a 95% confidence interval (95% CI).

### Quantum chemical calculations

The geometry of aromatic cavity was taken from the deposited NMR structure. The truncated model was made of aDMA and four residues forming the aromatic cavity. The $C_\alpha$ atoms were substituted with methyl groups that were kept frozen during the optimization to account for backbone covalent interactions. The geometry of aDMA and four aromatic cage amino acids was then optimized *in vacuo* using TPSS-D3/def2-TZVP level of theory with Turbomole 6.3 (44,45). For interaction energy calculations, we used a Hartree–Fock wave function of the monomers as a reference from which, we obtained interaction energy terms in first two orders at symmetry-adapted perturbation theory (SAPT0) approximation using PSI4 suite of codes. Density fitting of 4-index integrals and Laplace transformations of energy denominators were used as described in the following article (46). Reference wavefunction in SAPT analysis was calculated using aug-cc-pVDZ basis set.

## RESULTS

### Solution structure of TDRD3 in complex with aDMA–CTD peptide

Initial NMR titration experiments suggested that two regions of the Tudor domain of TDRD3 (residues 566–573 and 589–598) were significantly perturbed upon binding to the aDMA-CTD peptide (Figure 1). Analysis of NMR titration data under fast exchange regime allowed to evaluate the binding affinity of the TDRD3 Tudor with the sDMA- and aDMA-containing CTD peptides. TDRD3 binds aDMA-CTD with a $K_D$ of $770 \pm 30\,\mu\text{M}$, more than one order of magnitude stronger, compared with the symmetrically methylated isoform (Figure 1D). Similar binding affinity of TDRD3 to fluorescently labeled aDMA-CTD peptide was determined in a quantitative binding assay using fluorescence anisotropy ($K_D = 900 \pm 200\,\mu\text{M}$; Supplementary Figures S1 and S2).

To understand the structural basis for molecular recognition of aDMA-containing CTD by TDRD3, we determined the solution structure of a reconstituted complex consisting of the Tudor domain of TDRD3 (residues 555–610) and a 13-mer peptide aDMA-CTD [YSPSSP(aDMA)YTPQSP; Figure 1A] (Figure 2A and B; Supplementary Figure S2B and C; Supplementary Table S1). The Tudor domain of TDRD3 forms a four-stranded β-barrel fold and is similar to the structure of TDRD3 in free form (Supplementary Figure S2B) (25).

The β1–β2 and β3–β4 loops of the Tudor domain contain aromatic residues (Y566, Y573, F591 and Y594). These residues form a partially exposed aromatic cavity of rectangular cuboid shape that is side-walled by N596 (Figure 2C). Upon binding of TDRD3 to the CTD peptide, the aromatic cavity accommodates aDMA that is placed parallel between two tyrosines (Y566 and Y594). This binding results from cation–π and stacking interactions between positively charged guanidinium group of aDMA and aromatic rings of the two parallel tyrosine residues. Furthermore, CH–π interactions between the methyl groups of aDMA and two remaining aromatic residues (Y573 and F591) stabilize the interaction. One guanidinomethyl group faces Y573, whereas the second guanidinomethyl group faces F591 at the back wall of the cavity (Figure 2C). Interestingly, conserved N596 does not form a H-bond to aDMA-CTD, akin to the structures of s/aDMA bound to the canonical Tudor domains of SPF30 and SMN (26). This conserved asparagine in the structures of extended Tudor domains forms an H-bond to sDMA, altering its role in this Tudor sub-family (27,28).

Our NMR data show no intermolecular contacts between the aromatic cavity and neighboring residues of aDMA (Supplementary Figure S2C). This suggests that those residues are flexible and do not interact with the TDRD3 Tudor domain, and that the CTD peptide with the aDMA mark is recognized in a sequence-independent manner. Furthermore, phosphorylations of aDMA-CTD (at Ser2 and Ser5) showed no effect on the binding affinity to TDRD3.

### TDRD3 Y566 mutants loose specificity to aDMA-CTD peptides

It was shown that substitution of any of the four aromatic residues of the cavity with a non-aromatic amino acid, abrogates dimethylarginine binding (26). In the present study, we have investigated the effect of aromatic substitutions of the least conserved residue within the aromatic cavity (Figures 2C and 3A). In TDRD3, Y566 is a unique residue, whereas SMN, SPF30, SND1 and Tur11 contain tryptophan or phenylalanine in this position. Y566F substitution diminishes binding to aDMA-CTD and it does not increase binding affinity to sDMA-CTD (Figure 3B and Supplementary Figure S5). On the other hand, Y566W substitution promotes complex formation with sDMA-CTD, yet it has similar binding affinity to aDMA-CTD as the wild-type protein. This indicates that both phenylalanine and tryptophan substitutions at position Y566 abrogate TDRD3 selectivity for aDMA-against sDMA-CTD peptides. Residue Y566 is thus the
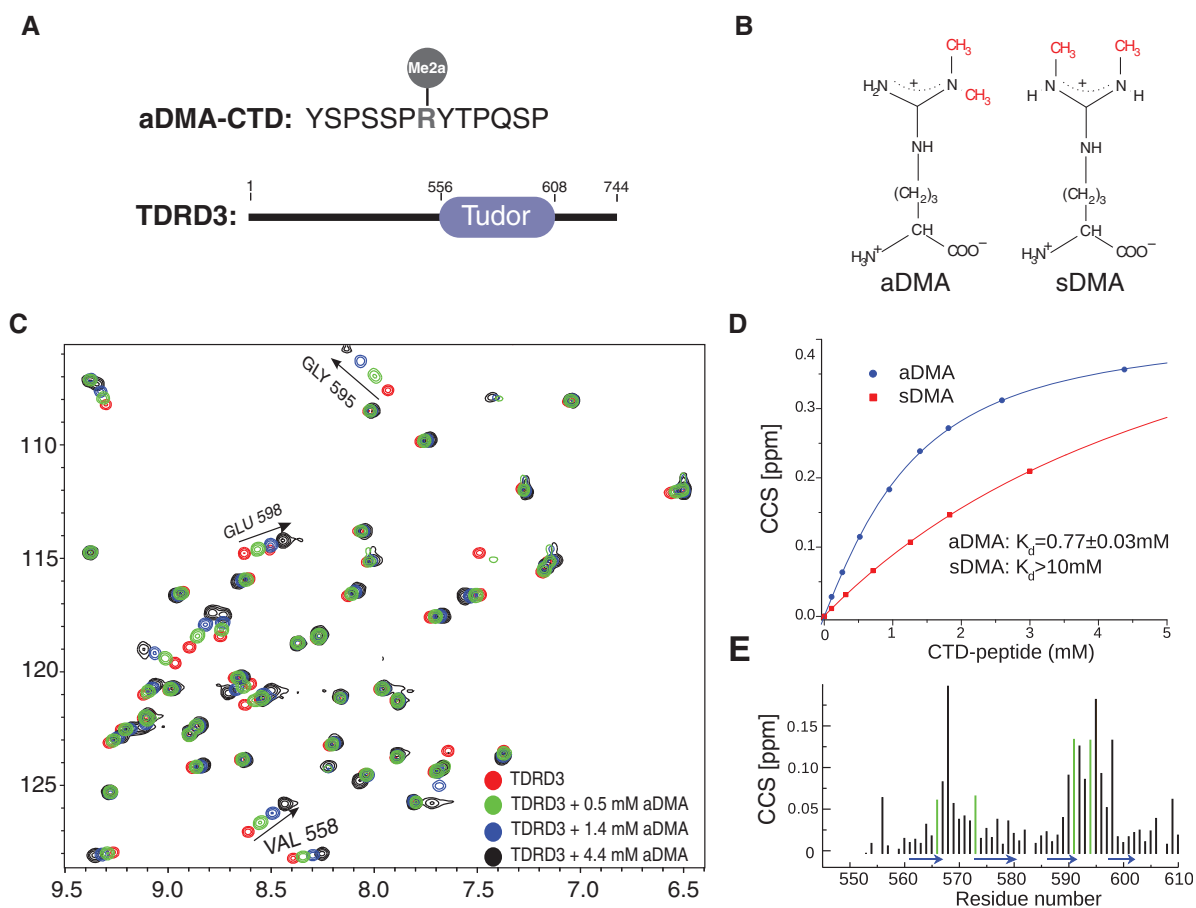
**Figure 1.** Interaction of the TDRD3 Tudor with aDMA-CTD. (**A**) Sequence of a 13 amino acid peptide with aDMA, used as a mimic of methylated CTD; and domain organization of TDRD3. (**B**) Scheme of aDMA and sDMA. (**C**) A 2D $^1$H-$^{15}$N HSQC spectra showing four representative steps of NMR titration of 1.8 mM TDRD3 with aDMA-CTD peptide. Trajectories for the three best resolved signals that were used for construction of binding isotherm are highlighted. (**D**) Estimation of the TDRD3 Tudor–aDMA-CTD (in blue) and TDRD3 Tudor–sDMA-CTD (in red) dissociation constants from NMR titration experiments. CCS values derived from $^1$H-$^{15}$N HSQCs are plotted against the CTD peptide concentration. Errors are denoted as 95% CI. (**E**) Quantification of chemical shift perturbations of the TDRD3 Tudor upon addition of aDMA-CTD. The combined chemical shift perturbations are plotted versus the amino-acid residue number with β-sheets regions shown as blue arrows. Large changes occur in the regions involved in binding of the aDMA-CTD peptide. Green lines represent the residues forming hydrophobic cavity.

key element, which determines the specificity of TDRD3 toward aDMA-containing peptides.

Comparison of the TDRD3–aDMA-CTD complex with previously determined structures of SPF30/SMN–a/sDMA complexes show variations in accommodation of the substrates in the aromatic cavities (Supplementary Figure S3A). The sDMA substrates are inserted at a different angle than aDMA, so that they maximize stacking interactions with the tryptophan aromatic residue in the position 566 (TDRD3 numbering). This underlines the importance of the aromatic residue type (Y/W/F; Figure 3A) in this position for dimethylarginine recognition.

### Pyramidalization of aDMA amino group promotes hydrogen bond formation with TDRD3

Force field approximations along with limited resolution of experimental data often imperfectly describe molecular interactions. In our structure of the TDRD3–aDMA-CTD complex, we identified a possibility of H-bond between the aDMA amino group and the hydroxyl

group of Y566. Such H-bond would require a distortion of the aDMA amino group planarity. Studies of high-resolution x-ray structures have shown amino group pyramidalization allowing hydrogen bond formation in nucleic acids (47). However, empirical force fields used for the structural calculations enforce the planarity of the amino group; therefore, we had to resort to other methods. To probe the existence of this H-bond, we performed a dispersion-corrected density functional theory (DFT-D) study of aDMA bound to the aromatic cavity of TDRD3 (for details, see 'Materials and Methods' section). Geometry of the DFT-D optimized model is very close to the NMR structure with heavy atom RMSD of 0.9 Å (Supplementary Figure S3B). Most importantly, the planarity of aDMA amino group was slightly distorted (due to partial sp$^3$ hybridization), allowing the H-bond formation with the hydroxyl group of Y566 (Figure 3D). As the cavity is formed by hydrophobic residues and its interior is not accessible to solvents, it is likely that this H-bond has significant energy. Y566F substitution in TDRD3 or the lack of
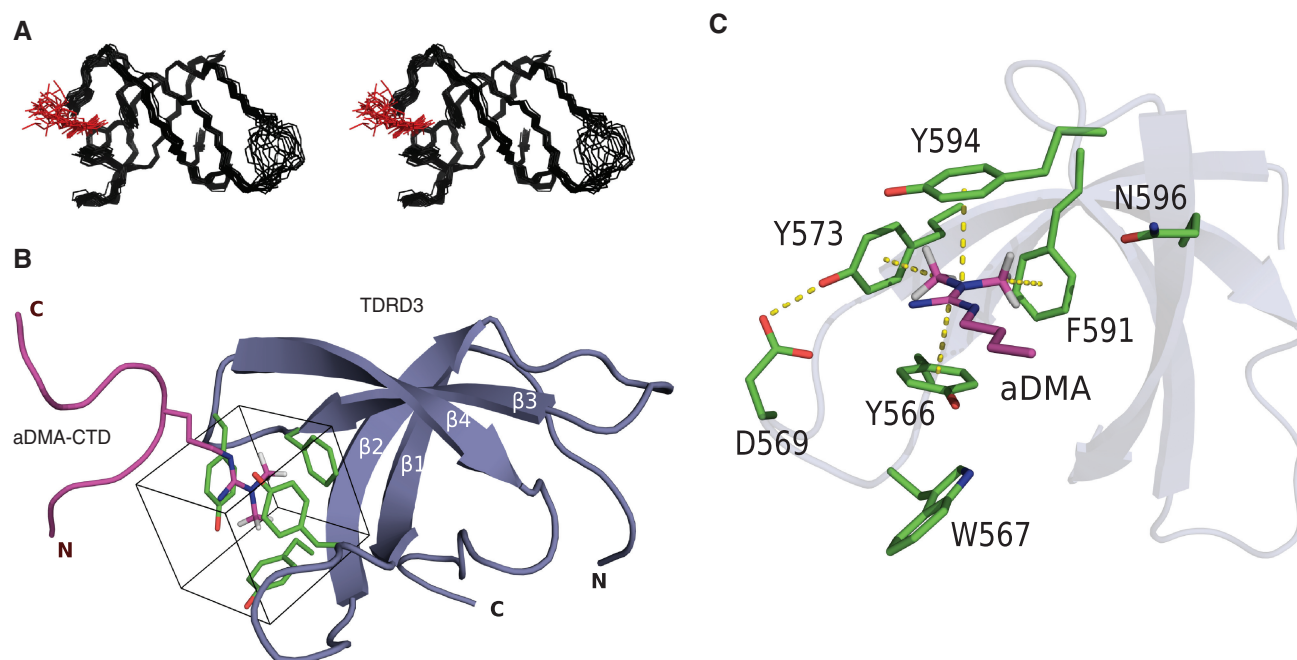
**Figure 2.** NMR structure of the TDRD3 Tudor–aDMA-CTD complex. (**A**) Overlay of the 20 lowest energy structures of the TDRD3 Tudor–aDMA-CTD complex shown in stereo view. Backbone and side-chains of the aromatic cavity (Y566, Y573, F591 and Y594) are shown in wire representation. For clarity, only aDMA residue of the bound peptide is shown in wire representation. The TDRD3 Tudor domain and aDMA are shown in black and red, respectively. (**B**) Structure of the human TDRD3 Tudor domain bound to the aDMA-CTD peptide. The aDMA-CTD peptide is represented as a magenta ribbon with the aDMA residue in sticks (only methyl protons are shown) and the protein is shown as a blue ribbon model. Residues forming the aromatic cavity (Y566, Y573, F591 and Y594) are shown in green sticks and the cube-shaped cavity is highlighted by a square. (**C**) aDMA recognition by the TDRD3 Tudor domains. The hydrogen bond and aDMA interactions with the aromatic cavity are shown with yellow dotted lines. Only aDMA and the side-chain of amino acids that form, or interact with the aromatic cavity are shown (starting from $C_\alpha$ atoms).

amino group in sDMA prevents formation of this H-bond (Figures 1D and 3B). This hydrogen bond explains why, during titration experiments, Y566F substitution abrogated TDRD3 binding with aDMA and why sDMA did not bind to wild-type TDRD3.

### Dispersion interactions play a major role in complex formation

To understand the forces contributing to the complex formation, we calculated the interaction energy of previously optimized DFT-D model of TDRD3–aDMA complex using SAPT calculations (48). In this perturbation approach, the interaction energy is expressed as a sum of physically well defined parts. Electrostatic and exchange-repulsion terms represent the first order contributions, whereas the second order terms are represented by induction and dispersion energy. Calculated interaction energy terms are summarized in Figure 3C.

The 'electrostatic' term is stabilizing. This is consistent with our picture that $N-H\cdots O$ hydrogen bond and cation–π interactions contribute to complex formation. However, the 'exchange-repulsion' over-compensates the electrostatic attraction, as many electron pairs come close when aDMA and aromatic residues form the complex. Therefore, the first order contributions are repulsive. The second order contributions showed that 'induction' term plays a minor role for the overall binding energy. In general, cation–π interactions are dominated by

electrostatic forces and cation-induced polarization of π systems (49). This suggests that induction mainly represents an interaction between a positive guanidinium group of aDMA and π orbitals of parallel tyrosines. The 'dispersion' interaction is the most significant stabilizing term. It should be noted that all non-covalent interactions contain some degree of a dispersion-type component (49). However, the role of dispersion in cation–π and hydrogen bonds is modest (50). As the dispersion energy is almost as large as the total interaction energy, we conclude that the dispersion originates predominantly from stacking interactions between aDMA and two parallel tyrosines (Y566 and Y594).

### Surrounding amino acids stabilize the aromatic cavity

The structure of the TDRD3–aDMA-CTD complex reveals that D569 of the β1–β2 loop forms the H-bond with the hydroxyl group of Y573 and thus, stabilizes the geometry of the aromatic cavity (Figure 2C). If the H-bond formation is impaired by D569A substitution, the [$^1$H,$^{15}$N]-HSQC spectrum is significantly altered when compared with that of the wild-type (Supplementary Figure S4A). The residues of the aromatic cavity and their nearest neighbors display large chemical shift perturbations, suggesting that the geometry of the cavity is disturbed (Supplementary Figure S4B). Furthermore, D569A substitution abrogates binding to the aDMA-CTD peptide, indicating the importance of D569 for
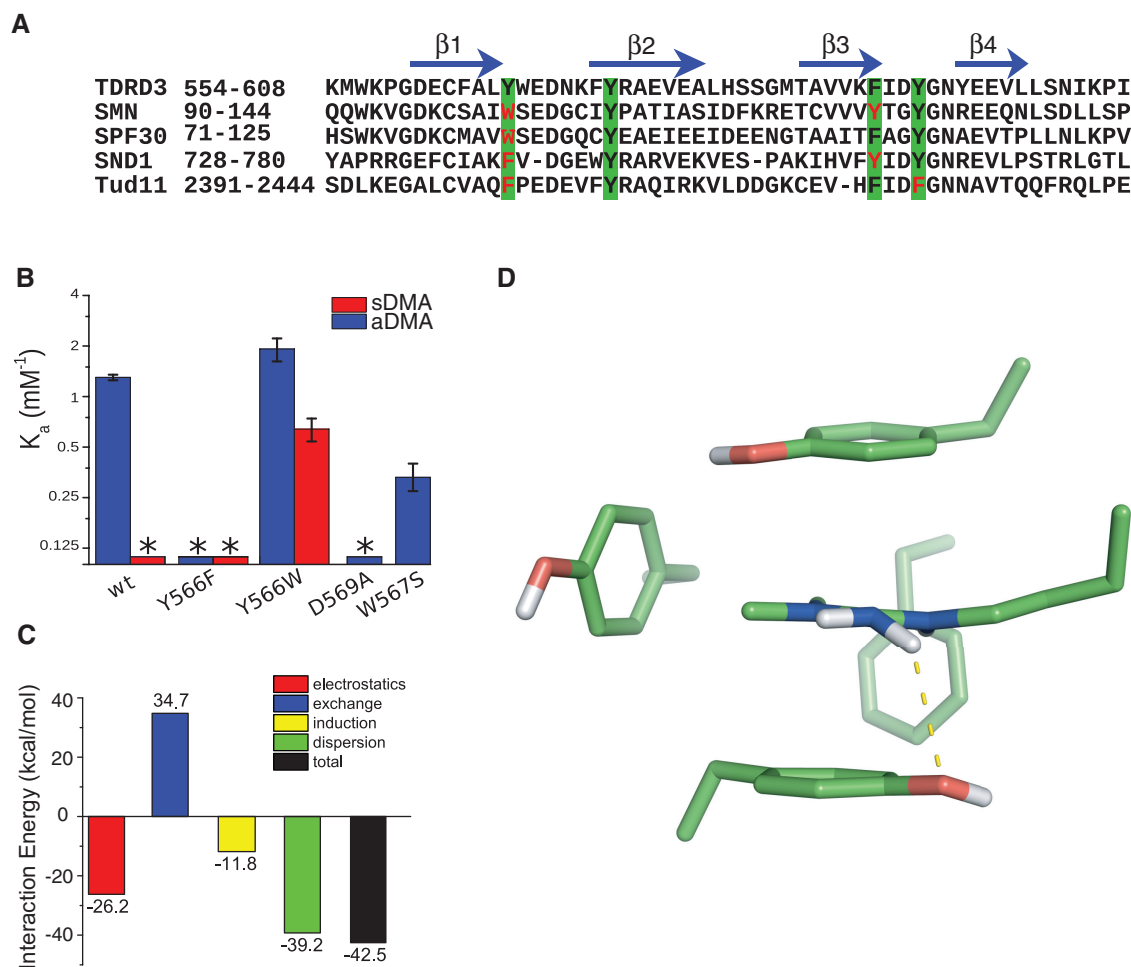
**A**

```
            β1              β2              β3      β4
TDRD3  554-608  KMWKPGDECFAL Y WEDNKF Y RAEVEALHSSGMTAVVK F IDY G NYEEVLLSNIKPI
SMN     90-144  QQWKVGDKCSAI W SEDGCI Y PATIASIDFKRETCVVV Y TGY G NREEQNLSDLLSP
SPF30   71-125  HSWKVGDKCMAV W SEDGQC Y EAEIEEIDEENGTAAIT F AGY G NAEVTPLLNLKPV
SND1   728-780  YAPRRGEFCIAK F V-DGEW Y RARVEKVES-PAKIHVF Y IDY G NREVLPSTRLGTL
Tud11 2391-2444 SDLKEGALCVAQ F PEDEVF Y RAQIRKVLDDGKCEV-H F IDF G NNAVTQQFRQLPE
```

**Figure 3.** Recognition of aDMA by the TDRD3 Tudor. (**A**) Sequence alignment of the human TDRD3 Tudor domain with other dimethylarginine binding Tudor domains of SMN (*Homo sapiens*), SPF30 (*H. sapiens*), SND1 (*H. sapiens*), and the 11th Tudor domain of Tud (*Drosophila melanogaster;* Tud11). Residues forming an aromatic cavity are highlighted in green squares, variable residues are shown in red. The β-sheet regions (β1, β2, β3 and β4) of human TDRD3 Tudor are shown with blue arrows. (**B**) Bar plot of the NMR-derived association constants ($K_a$) of various TDRD3 mutants with the aDMA-CTD peptide (blue) and sDMA-CTD peptide (red) in a logarithmic scale. Asterisk indicates that the binding constant could not be determined. (**C**) Bar plot showing decomposed interactions' energies between the aromatic cavity and aDMA calculated by SAPT. The exchange-repulsion (blue bar) accounts for an interaction caused by tunneling of the electrons between interacting systems and electron−electron repulsion due to the Pauli exclusion principle. The induction interaction (yellow bar) is a second-order energy contribution, which originates from mutual distortion of electron density distribution of interacting molecules. The dispersion interaction (green) arises from the correlated electron fluctuations in the interacting molecules (48). (**D**) Pyramidalization of the aDMA amino group as predicted by DFT-D theory. The hydrogen bond (2.7 Å) that is responsible for aDMA recognition is shown with yellow dotted line. Only heavy atoms and non-polar hydrogens are shown.

the stabilization of the aromatic cavity (Supplementary Figure S4B).

Conserved W567 residue stacks with Y566, stabilizing the geometry of aromatic cavity (Figure 2C and Supplementary Figure S2A). W567S substitution of TDRD3 Tudor yields lower affinity to aDMA-CTD (Figure 3B). These results demonstrate the importance of residues surrounding the cavity for the recognition of aDMA by TDRD3.

## DISCUSSION

Available structures of the SMN, SPF30, SND1 and Tud11 Tudor domains in complex with their corresponding dimethylarginine ligands (26–28), together with the structure of the TDRD3 Tudor domain bound to

aDMA-CTD, define structural determinants for the recognition of sDMA- and aDMA-containing peptides. The extended Tudor domains (SND1 and Tur11) contain an additional α-helix and two β-strands at aminoterminal and several α-helices and β-strands at the carboxy-terminal to the canonical Tudor core. They preferably interact with sDMA-containing peptides, because of hydrogen bond formation between sDMA and asparagine residue (Supplementary Figure S6). In contrast, the structures of the canonical Tudor domains bound to dimethylarginine ligands lack this hydrogen bond (Supplementary Figure S7). They utilize a different strategy for the dimethylarginine recognition. The SMN and SPF30 canonical Tudor domains contain tryptophan at position 566 (Figure 3A). The presence of the fused-ring heterocyclic amino acid alters the stacking interactions

between aromatic cavity and ligand. As a result, both SMN and SPF30 preferably bind to sDMA but also possess a comparable affinity to aDMA.

The presence of tyrosine at position 566 in the canonical Tudor domain of TDRD3, creates unique binding properties of the cavity as it selectively binds the aDMA mark but not the sDMA mark. The TDRD3–aDMA complex formation is driven by a combination of cation–$\pi$, CH–$\pi$ and stacking interactions between aromatic cavity of Tudor domain and aDMA ligand. The complex is further stabilized by the $N-H\cdots O$ hydrogen bond between the aDMA amino group and the hydroxyl group of Y566. Y566F variant of TDRD3, which causes the loss of this hydrogen bond, abrogates the interaction between TDRD3 and aDMA. Interestingly, Y566W variant allows TDRD3 to form a complex with sDMA-CTD and also has a similar affinity to aDMA-CTD (Figure 3B). The presence of tryptophan in this position provides an additional dispersion interaction that is non-specific to the dimethylarginine motifs. Y566W mutant of the TDRD3 Tudor domain is therefore, able to bind both sDMA and aDMA peptides even though the $N-H\cdots O$ hydrogen bond is broken by this substitution. We conclude that the tyrosine residue at position 566 of the TDRD3 Tudor domain works as a selectivity filter for the aDMA-containing peptides.

The results presented in this study extend our understanding of how the methylarginine binding Tudor domains coordinate their ligands. A comparison of the ligand-bound TDRD3 structure with other Tudor domains, supported by the mutational analysis and quantum chemical calculations provide a framework to understand dimethylarginine recognition at a molecular level.

## ACCESSION NUMBERS

The atomic coordinates for the NMR ensemble of the TDRD3–aDMA-CTD complex have been deposited in the Protein Data Bank under accession code 2lto. The chemical shift assignments have been deposited in the Biological Magnetic Resonance Data Bank under accession code 18490.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–7.

## ACKNOWLEDGEMENTS

## FUNDING

## REFERENCES

1. Bedford,M.T. and Richard,S. (2005) Arginine methylation an emerging regulator of protein function. *Mol. Cell*, **18**, 263–272.
2. Bedford,M.T. and Clarke,S.G. (2009) Protein arginine methylation in mammals: who, what, and why. *Mol. Cell*, **33**, 1–13.
3. Siomi,M.C., Mannen,T. and Siomi,H. (2010) How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev.*, **24**, 636–646.
4. Yang,Y., Lu,Y., Espejo,A., Wu,J., Xu,W., Liang,S. and Bedford,M.T. (2010) TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol. Cell*, **40**, 1016–1023.
5. Sims,R.J. 3rd, Rojas,L.A., Beck,D., Bonasio,R., Schuller,R., Drury,W.J. 3rd, Eick,D. and Reinberg,D. (2011) The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science*, **332**, 99–103.
6. Phatnani,H.P. and Greenleaf,A.L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.*, **20**, 2922–2936.
7. Corden,J.L. (2007) Transcription. Seven ups the code. *Science*, **318**, 1735–1736.
8. Chapman,R.D., Heidemann,M., Hintermair,C. and Eick,D. (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet.*, **24**, 289–296.
9. Egloff,S. and Murphy,S. (2008) Cracking the RNA polymerase II CTD code. *Trends Genet.*, **24**, 280–288.
10. Buratowski,S. (2003) The CTD code. *Nat. Struct. Biol.*, **10**, 679–680.
11. Buratowski,S. (2009) Progression through the RNA polymerase II CTD cycle. *Mol. Cell*, **36**, 541–546.
12. Hirose,Y. and Manley,J.L. (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev.*, **14**, 1415–1429.
13. Maniatis,T. and Reed,R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499–506.
14. Meinhart,A., Kamenski,T., Hoeppner,S., Baumli,S. and Cramer,P. (2005) A structural perspective of CTD function. *Genes Dev.*, **19**, 1401–1415.
15. Viladevall,L., St Amour,C.V., Rosebrock,A., Schneider,S., Zhang,C., Allen,J.J., Shokat,K.M., Schwer,B., Leatherwood,J.K. and Fisher,R.P. (2009) TFIIH and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast. *Mol. Cell*, **33**, 738–751.
16. Ghosh,A., Shuman,S. and Lima,C.D. (2011) Structural insights to how mammalian capping enzyme reads the CTD code. *Mol. Cell*, **43**, 299–310.
17. de la Mata,M. and Kornblihtt,A.R. (2006) RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat. Struct. Mol. Biol.*, **13**, 973–980.
18. Munoz,M.J., de la Mata,M. and Kornblihtt,A.R. (2010) The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends Biochem. Sci.*, **35**, 497–504.
19. David,C.J., Boyne,A.R., Millhouse,S.R. and Manley,J.L. (2011) The RNA polymerase II C-terminal domain promotes splicing

activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev.*, **25**, 972–983.

20. de Almeida,S.F., Grosso,A.R., Koch,F., Fenouil,R., Carvalho,S., Andrade,J., Levezinho,H., Gut,M., Eick,D., Gut,I. *et al.* (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.*, **18**, 977–983.

21. Ahn,S.H., Kim,M. and Buratowski,S. (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol. Cell*, **13**, 67–76.

22. Egloff,S., O'Reilly,D., Chapman,R.D., Taylor,A., Tanzhaus,K., Pitts,L., Eick,D. and Murphy,S. (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, **318**, 1777–1779.

23. Johnson,S.A., Kim,H., Erickson,B. and Bentley,D.L. (2011) The export factor Yra1 modulates mRNA 3′ end processing. *Nat. Struct. Mol. Biol.*, **18**, 1164–1171.

24. MacKellar,A.L. and Greenleaf,A.L. (2011) Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. *J. Biol. Chem.*, **286**, 36385–36395.

25. Liu,K., Guo,Y., Liu,H., Bian,C., Lam,R., Liu,Y., Mackenzie,F., Rojas,L.A., Reinberg,D., Bedford,M.T. *et al.* (2012) Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30. *PLoS One*, **7**, e30375.

26. Tripsianes,K., Madl,T., Machyna,M., Fessas,D., Englbrecht,C., Fischer,U., Neugebauer,K.M. and Sattler,M. (2011) Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nat. Struct. Mol. Biol.*, **18**, 1414–1420.

27. Liu,K., Chen,C., Guo,Y., Lam,R., Bian,C., Xu,C., Zhao,D.Y., Jin,J., MacKenzie,F., Pawson,T. *et al.* (2010) Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc. Natl Acad. Sci. USA*, **107**, 18398–18403.

28. Liu,H., Wang,J.Y., Huang,Y., Li,Z., Gong,W., Lehmann,R. and Xu,R.M. (2010) Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev.*, **24**, 1876–1881.

29. Cote,J. and Richard,S. (2005) Tudor domains bind symmetrical dimethylated arginines. *J. Biol. Chem.*, **280**, 28476–28483.

30. Chen,C., Jin,J., James,D.A., Adams-Cioaba,M.A., Park,J.G., Guo,Y., Tenaglia,E., Xu,C., Gish,G., Min,J. *et al.* (2009) Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. *Proc. Natl Acad. Sci. USA*, **106**, 20336–20341.

31. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

32. Bax,A. and Grzesiek,S. (1993) Methodological advances in protein NMR. *Accounts Chem. Res.*, **26**, 131–138.

33. Peterson,R.D., Theimer,C.A., Wu,H. and Feigon,J. (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J. Biomol. NMR*, **28**, 59–67.

34. Zwahlen,C., Legault,P., Vincent,S.J.F., Greenblatt,J., Konrat,R. and Kay,L.E. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *J. Am. Chem. Soc.*, **119**, 6711–6721.

35. Guntert,P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol.*, **278**, 353–378.

36. Herrmann,T., Guntert,P. and Wuthrich,K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, **319**, 209–227.

37. Case,D.A., Darden,T.A., Cheatham,T.E. III, Simmerling,C.L., Wang,J., Duke,R.E., Luo,R., Crowley,M., Walker,R.C., Zhang,W. *et al.* (2008) *AMBER 10*. University of California, San Francisco.

38. Case,D.A., Cheatham,T.E. 3rd, Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.

39. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

40. Bayly,C.I., Cieplak,P., Cornell,W.D. and Kollman,P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the Resp Model. *J. Phys. Chem.*, **97**, 10269–10280.

41. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.

42. Vriend,G. (1990) What If - a molecular modeling and drug design program. *J. Mol. Graphics*, **8**, 52–56.

43. Heyduk,T. and Lee,J.C. (1990) Application of fluorescence energy transfer and polarization to monitor Escherichia coli cAMP receptor protein and lac promoter interaction. *Proc. Natl Acad. Sci. USA*, **87**, 1744–1748.

44. Grimme,S., Antony,J., Ehrlich,S. and Krieg,H. (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, **132**, 154104.

45. Tao,J., Perdew,J.P., Staroverov,V.N. and Scuseria,G.E. (2003) Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, **91**, 146401.

46. Turney,J.M., Simmonett,A.C., Parrish,R.M., Hohenstein,E.G., Evangelista,F.A., Fermann,J.T., Mintz,B.J., Burns,L.A., Wilke,J.J., Abrams,M.L. *et al.* (2012) Psi4: an open-source ab initio electronic structure program. *WIREs: Comput. Mol. Sci.*, **2**, 556–565.

47. Luisi,B., Orozco,M., Sponer,J., Luque,F.J. and Shakked,Z. (1998) On the potential role of the amino nitrogen atom as a hydrogen bond acceptor in macromolecules. *J. Mol. Biol.*, **279**, 1123–1136.

48. Jeziorski,B., Moszynski,R. and Szalewicz,K. (1994) Perturbation-theory approach to intermolecular potential-energy surfaces of Van-Der-Waals complexes. *Chem. Rev.*, **94**, 1887–1930.

49. Cubero,E., Luque,F.J. and Orozco,M. (1998) Is polarization important in cation-pi interactions? *Proc. Natl Acad. Sci. USA*, **95**, 5976–5980.

50. Kim,D., Hu,S., Tarakeshwar,P., Kim,K.S. and Lisy,J.M. (2003) Cation-pi interactions: A theoretical investigation of the interaction of metallic and organic cations with alkenes, arenes, and heteroarenes. *J. Phys. Chem. A*, **107**, 1228–1238.

## 5.6   Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation

In this work, structural and functional evidence is provided demonstrating that the same domain of Nrd1p interacts with RNA polymerase II and Trf4p in a mutually exclusive manner, thus defining two alternative forms of the NNS complex, one involved in termination and the other in degradation.

# Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation

Agnieszka Tudek,[1,6] Odil Porrua,[1,5,6,*] Tomasz Kabzinski,[2,3,6] Michael Lidschreiber,[4] Karel Kubicek,[2,3] Andrea Fortova,[2,3] François Lacroute,[1] Stepanka Vanacova,[2,3] Patrick Cramer,[4] Richard Stefl,[2,3,7,*] and Domenico Libri[1,5,7,*]
[1]Centre de Génétique Moléculaire, CNRS UPR3404, 91190 Gif sur Yvette, France
[2]CEITEC-Central European Institute of Technology, Masaryk University, Brno 62500, Czech Republic
[3]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno 62500, Czech Republic
[4]Gene Center Munich and Department of Biochemistry, Center for Integrated Protein Science CIPSM, Ludwig-Maximilians-Universität München, 81377 Munich, Germany
[5]Present address: Institut Jacques Monod, CNRS, UMR 7592, Univ Paris Diderot, Sorbonne Paris Cité, 75205 Paris, France
[6]Co-first author
[7]Co-senior author
*Correspondence: porrua@ijm.univ-paris-diderot.fr (O.P.), richard.stefl@ceitec.muni.cz (R.S.), libri.domenico@ijm.univ-paris-diderot.fr (D.L.)
http://dx.doi.org/10.1016/j.molcel.2014.05.031

## SUMMARY

The Nrd1-Nab3-Sen1 (NNS) complex is essential for controlling pervasive transcription and generating sn/snoRNAs in *S. cerevisiae*. The NNS complex terminates transcription of noncoding RNA genes and promotes exosome-dependent processing/degradation of the released transcripts. The Trf4-Air2-Mtr4 (TRAMP) complex polyadenylates NNS target RNAs and favors their degradation. NNS-dependent termination and degradation are coupled, but the mechanism underlying this coupling remains enigmatic. Here we provide structural and functional evidence demonstrating that the same domain of Nrd1p interacts with RNA polymerase II and Trf4p in a mutually exclusive manner, thus defining two alternative forms of the NNS complex, one involved in termination and the other in degradation. We show that the Nrd1-Trf4 interaction is required for optimal exosome activity in vivo and for the stimulation of polyadenylation of NNS targets by TRAMP in vitro. We propose that transcription termination and RNA degradation are coordinated by switching between two alternative partners of the NNS complex.

## INTRODUCTION

The ubiquitous presence of transcribing polymerases in the genome is a potential risk to the cell, as it can hamper the appropriate expression of canonical genes by interfering with their transcription (Jensen et al., 2013). Pervasive transcription is controlled at the level of transcription termination and RNA degradation, which can be coupled in *S. cerevisiae*. The main actors of this quality control pathway are the Nrd1-Nab3-Sen1 (NNS) transcription termination complex, the nuclear exosome, and the TRAMP complex (for recent reviews, see Porrua and Libri, 2013a; Jensen et al., 2013). The NNS complex is required

for transcription termination of a large fraction of noncoding RNAs (ncRNAs) transcribed by RNA polymerase II (RNAPII), essentially CUTs (cryptic unstable transcripts), snRNAs, and snoRNAs (Steinmetz et al., 2001; Arigo et al., 2006; Thiebaut et al., 2006). CUTs are short lived in wild-type yeast and are largely nonfunctional, although in a few cases their transcription has been shown to regulate gene expression (Kuehner and Brow, 2008; Thiebaut et al., 2008). Transcripts terminated by the NNS pathway are polyadenylated by the TRAMP complex and targeted by the nuclear exosome for degradation (in the case of CUTs) or 3′ end trimming (in the case of snRNAs and snoRNAs).

The exosome is composed of a ring-shaped core to which two catalytic subunits, Dis3p and Rrp6p, associate. The two enzymes are 3′-5′ exonucleases, and Dis3p is also endowed with endonuclease activity (Chlebowski et al., 2013). Because the central channel of the ring that drives the substrate toward the catalytic subunit Dis3p is only wide enough to accommodate single-stranded RNA, it has been proposed that the presence of an unstructured region of at least 30 residues is required for degradation (Chlebowski et al., 2013). Rrp6p only associates with the nuclear form of the exosome and has overlapping and complementary roles to Dis3p in RNA degradation (Gudipati et al., 2012).

The TRAMP complex is an important cofactor of the exosome that is required for the efficient processing and degradation of a variety of RNAs produced by the three yeast RNA polymerases (Wyers et al., 2005; San Paolo et al., 2009; Wlotzka et al., 2011; Kadaba et al., 2004). TRAMP is composed of a poly(A) polymerase (Trf4p or Trf5p), a zinc knuckle RNA-binding protein (Air1p or Air2p), and the DExH-box RNA helicase Mtr4p (LaCava et al., 2005; Vanácová et al., 2005; Wyers et al., 2005). Polyadenylation of exosome substrates by TRAMP favors their degradation (Callahan and Butler, 2010; Kadaba et al., 2004; LaCava et al., 2005; Rougemaille et al., 2007; Vanácová et al., 2005; Wyers et al., 2005), and it has been proposed that poly(A) tails added by TRAMP provide the unstructured extensions that allow threading of structured substrates through the central channel of the exosome ring. It has also been shown that TRAMP stimulates exosome and Rrp6p activity independently of polyadenylation
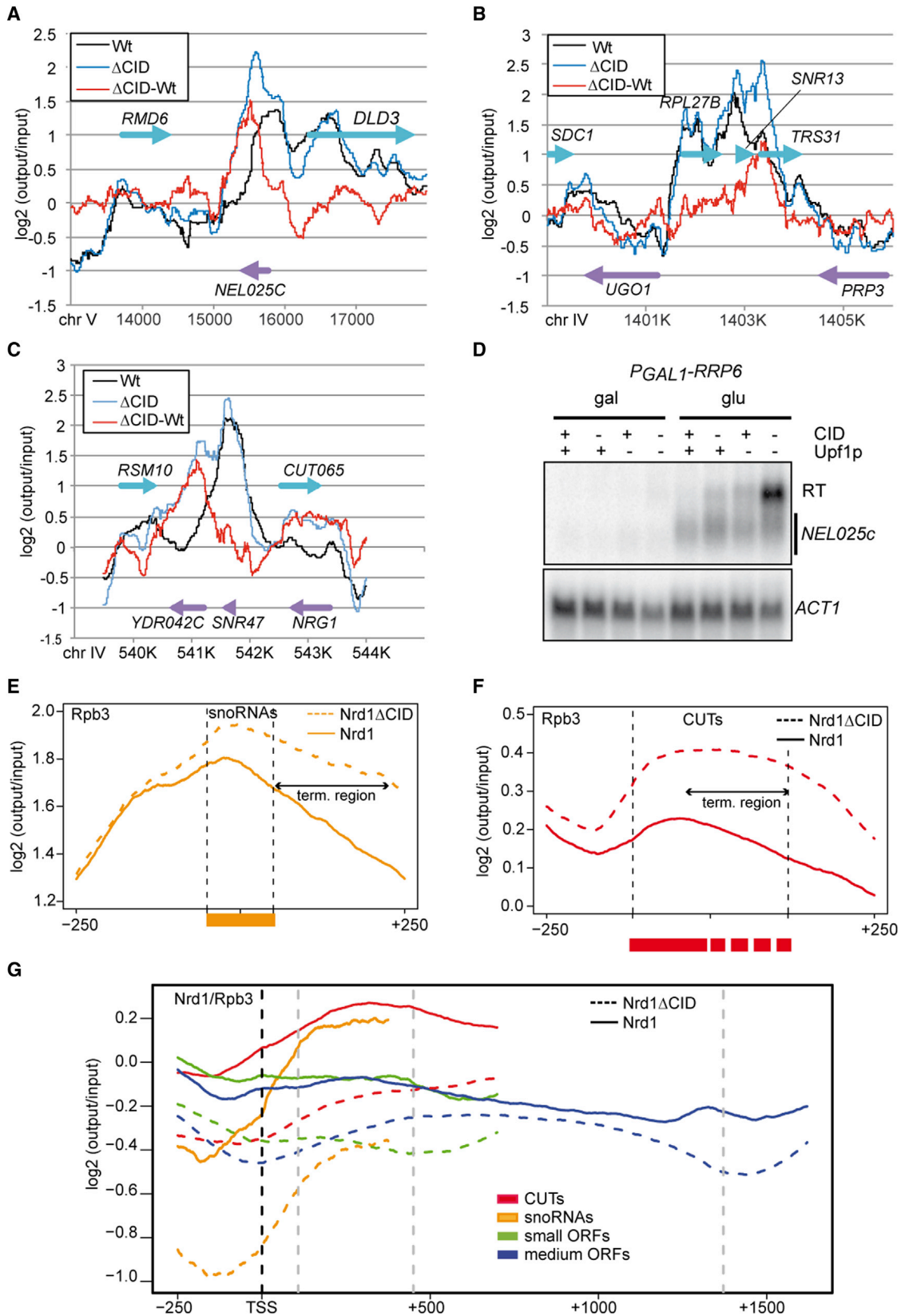
(legend on next page)

(Wyers et al., 2005; Callahan and Butler, 2010; Rougemaille et al., 2007), although the mechanistic details of this stimulation are still unclear.

The NNS complex is composed of the RNA-binding proteins Nrd1p and Nab3p and the superfamily I helicase Sen1p. Binding of the Nrd1-Nab3 complex to specific motifs on the nascent RNA constitutes the essential readout of transcription termination signals (Creamer et al., 2011; Porrua et al., 2012; Wlotzka et al., 2011). The actual termination step is most likely operated by Sen1p that interacts with the Nrd1-Nab3 complex (Hazelbaker et al., 2013; Porrua and Libri, 2013b). The NNS complex has been shown to associate with TRAMP and the exosome, which is thought to favor degradation (Vasiljeva and Buratowski, 2006), although the molecular and mechanistic details of the interaction between the NNS complex, the exosome, and the TRAMP are not well understood.

Nrd1p interacts with the C-terminal domain (CTD) of the largest subunit of RNAPII via a CTD interacting domain (CID). The CID recognizes heptapeptide (YSPTSPS) repeats in the CTD that are phosphorylated on the serine at position five (Ser5P) (Kubicek et al., 2012; Mayer et al., 2012; Vasiljeva et al., 2008). Because this modification mark predominates early in transcription, when NNS-dependent termination preferentially occurs (Buratowski, 2009; Gudipati et al., 2008; Jenks et al., 2008; Steinmetz et al., 2006a), the CID-CTD interaction is believed to determine the regional specificity of termination (Gudipati et al., 2008; Vasiljeva et al., 2008). However, previous studies did not detect significant termination defects in a *nrd1ΔCID* background (Vasiljeva et al., 2008). Surprisingly, RNAs produced by NNS termination were found to be stabilized in *nrd1ΔCID* cells, suggesting that the CID domain might favor degradation/processing by the exosome (Kubicek et al., 2012; Vasiljeva et al., 2008).

Here we analyze the role of the Nrd1p CID in transcription termination and in promoting RNA degradation/processing by the nuclear exosome. We detected widespread termination defects at NNS targets and defective recruitment of Nrd1p to elongation complexes upon deletion of the CID. Surprisingly, we discovered that the CID also mediates the interaction between the NNS complex and TRAMP by recognizing in Trf4p a CTD mimic that we dubbed NIM (for Nrd1 interaction motif). We solved the solution structure of the interaction surface, and we show that the interactions of Nrd1p with TRAMP and RNAPII are mutually exclusive. Importantly, we demonstrate that the Nrd1p-Trf4p interaction stimulates the polyadenylation activity of TRAMP in vitro, suggesting that the CID contributes to efficient degradation of exosome substrates by facilitating TRAMP recruitment and function. Our results demonstrate the existence of two alternative forms of the NNS complex: one associated with RNAPII and functioning in termination and the other associated with TRAMP and promoting RNA degradation.

## RESULTS

### The Nrd1p CID Domain Plays a Role in NNS-Dependent Transcription Termination

To assess the role of the Nrd1p CID in the function of the NNS complex, we reexamined whether this domain is required for efficient transcription termination by the NNS pathway. We compared the RNAPII distribution by chromatin immunoprecipitation (ChIP) and tiling arrays in wild-type and *nrd1ΔCID* cells. Upon deletion of the CID, we observed persistent presence of RNAPII downstream of many genes, such as the CUT *NEL025C*, *SNR13*, *SNR47*, and *CUT065* (Figures 1A–1C), indicating transcriptional readthrough. These findings were confirmed by northern blot analyses (Figures 1D and S2A, available online) and showed that in some cases (e.g., *NEL025c*) poor detection of the readthrough transcripts is due to the combined nuclear and cytoplasmic degradation of these species. Indeed, readthrough transcripts become prominent in cells in which both the nuclear exosome and the cytoplasmic nonsense-mediated decay (NMD) degradation pathways are defective (i.e, in a Rrp6p-depleted, *Δupf1* mutant; Figure 1D).

Metagene analyses suggested that readthrough occurs at the majority of NNS targets, such as CUTs and snoRNAs (Figures 1E and 1F; see also Supplemental Experimental Procedures). Readthrough also occurs at small open reading frames (ORFs; <500 bp; Figure S1A) that have been shown to be partially NNS dependent (Steinmetz et al., 2006b) and possibly at a subset of larger ORFs (Figure S1B), although in both cases it is not possible to clearly distinguish bona fide readthrough events from failure to terminate antisense transcription that is frequently observed at the 3′ end of ORFs (Neil et al., 2009; Xu et al., 2009). Consistent with this notion, genes with no detected antisense transcripts like *RMD6*, *DLD3*, *TRS31*, *UGO1*, and *PRP3* display no significant readthrough (Figures 1A and 1B).

We compared the genome-wide chromatin distribution of Nrd1p in the presence and the absence of the CID by ChIP-chip analysis, normalizing to transcription levels as defined by RNAPII occupancy in both strains (Figure 1G; for the non-normalized Nrd1p occupancy, see Figure S1C). Metagene analyses for the four distinct classes of features showed that

**Figure 1. Effect of CID Deletion on RNAPII and Nrd1p Occupancy Determined by ChIP-Chip at NNS Targets**
(A–C) Rpb3p-TAP occupancy is plotted at the *NEL025c* (A), *SNR13* (B), and *SNR47* (C) loci in a wild-type (black) or an *nrd1ΔCID* (blue) strain as indicated. The difference between the two signals (ΔCID-WT) is also plotted in red. The position of the relevant features on the W or C strand is indicated by blue and violet arrows, respectively. The signal (log2 ratio) is normalized to its genome-wide median level.
(D) Northern blot analysis of *NEL025c* transcripts in the presence and absence of the Nrd1p CID and in NMD *Δupf1* mutant cells. In this experiment, Rrp6p was metabolically depleted using the glucose-repressible *GAL1* promoter as indicated.
(E and F) Metagene analysis of RNAPII distribution for *NRD1* and *nrd1ΔCID* strains at snoRNAs (E) and CUTs (F). All features have been scaled and aligned either to the coordinates of the mature transcript (snoRNAs) or to the annotation based on tiling array analyses (CUTs; Xu et al., 2009). Alignment borders are indicated by gray dotted lines. The approximate range of termination for CUTs is indicated by a double arrow.
(G) Metagene distribution of Nrd1p occupancy for the different classes of features as indicated, in the presence (plain lines) or absence (dotted lines) of the CID. All signals are normalized to Rpb3p occupancy to limit biases due to differences in transcription levels. Raw Nrd1p signals for all classes are shown in Figure S1. Features have been scaled and aligned as in (E) and (F) and in Figure S1.
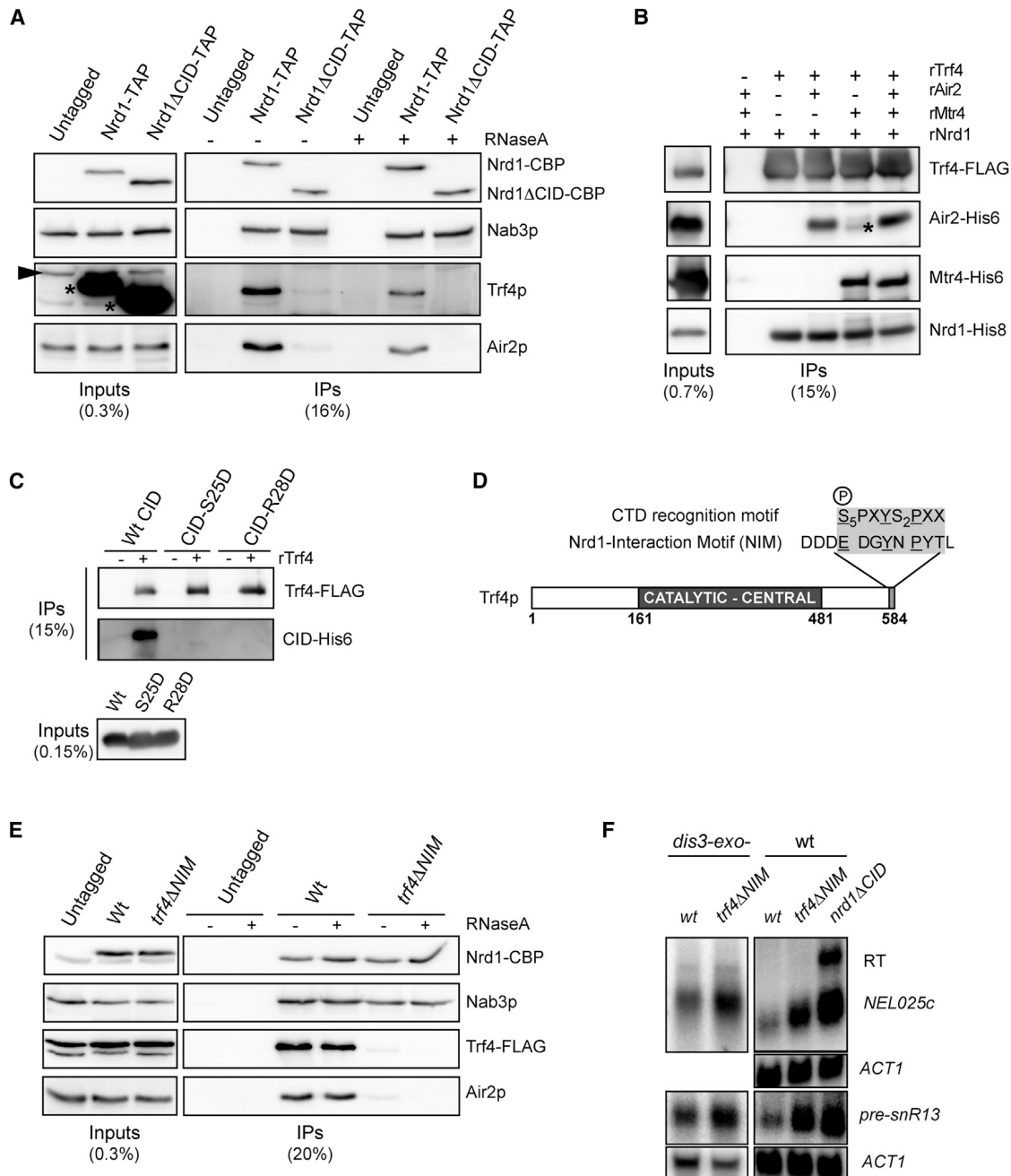
**Figure 2. Direct Interaction between Trf4p and Nrd1p Is Mediated by the CID**

(A) Western blot analysis of Nrd1p-TAP and Nrd1ΔCIDp-TAP immunopurified complexes (IP). Samples were eluted by cleavage with the TEV protease. The indicated proteins were detected with specific antibodies, except for Nrd1p, which was detected with an anti-CBP antibody. The Trf4 signal in the input (indicated by an arrowhead) is partially overlapping with the Nrd1-TAP signal (denoted by an asterisk). The fraction of extract and immunoprecipitated material that is loaded on the gel is indicated.

(B) Immunoblot analysis of pull-down experiments performed with recombinant Trf4-FLAG as bait and *E. coli* extracts expressing recombinant His-tagged Air2, Mtr4, or Nrd1 as indicated. Immunoprecipitations were performed in the presence of RNase. An asterisk indicates a degradation fragment of Mtr4-His$_6$.

(C) Immunoblot analysis as in (B) using recombinant Trf4-FLAG and recombinant CID-His$_6$ or CID mutant derivatives (rCID-S25D-His$_6$ and CID-R28D-His$_6$), defective for the interaction with the CTD. Proteins were detected with an antibody anti-His tag or anti-FLAG.

(D) Scheme of Trf4p indicating the position and sequence of the NIM, compared to a CTD pattern containing the amino acids that mediate major contacts with the CID, including Ser5P (equivalent CTD and NIM regions are shaded; identical amino acids are underlined). Note the presence of a Ser5 phosphomimic (E) in the NIM.

(E) Western blot analysis of Nrd1-TAP immunopurified complexes from a *TRF4* or *trf4ΔNIM* strain as in (A).

*(legend continued on next page)*

wild-type Nrd1p is recruited to higher levels at CUTs and snoRNAs genes, but also at the 5′ end of ORFs as previously reported (Kim et al., 2010; Mayer et al., 2010), although the 5′ end recruitment peak was attenuated by RNAPII normalization (Figures 1G and S1D). Importantly, ablation of the CID domain affected recruitment in all instances, although higher Nrd1p occupancy persisted at CUTs relative to other features in *nrd1ΔCID* cells, presumably because of RNA-mediated recruitment. As expected, no effects of CID deletion were observed at tRNA genes (Figure S1E). Altogether, these results demonstrate that the CID is required for efficient termination at NNS-dependent targets.

## Nrd1p CID Recognizes the Trf4p Component of the TRAMP Complex

Consistent with previous reports (Kubicek et al., 2012; Vasiljeva et al., 2008), we observed that the levels of the *NEL025C* CUT, pre-snR13, and pre-snR47 were increased upon deletion of Nrd1 CID (Figure S2A). Stabilization was stronger when the nuclear degradation machinery was compromised in *rrp6* catalytic mutants (Figure S2A), presumably because of partial redundancy in the degradation pathways. Interestingly, we also detected stabilization of some degradation intermediates derived from the U4 and U5 snRNAs (Figure S2B), suggesting a more general CID requirement for optimal activity of Rrp6-exosome.

Because the NNS complex has previously been shown to copurify with the exosome and TRAMP complexes (Vasiljeva and Buratowski, 2006), we considered that the CID could be involved in mediating such interactions. To address this question, we first performed coimmunoprecipitation assays with wild-type or ΔCID TAP-tagged Nrd1p. We consistently observed strong signals for the TRAMP components Trf4p and Air2p in Nrd1p immunoprecipitates. Strikingly, however, TRAMP signals could not be detected in the absence of the CID in ribonuclease (RNase)-treated extracts (Figure 2A), which was also reported by Heo et al. (2013) while this work was in progress. As expected, deletion of the CID did not affect the interaction between Nrd1p and Nab3p (Figure 2A).

In contrast to the strong TRAMP signals, we only detected weak signals for Rrp6p and Dis3p in Nrd1-TAP immunoprecipitates upon RNase treatment (Figure S3A). This suggests that the TRAMP complex is a major partner of the NNS complex relative to the exosome. In order to assess whether the interaction between Nrd1p and the TRAMP is direct, we performed pull-down experiments using *E. coli* extracts containing recombinant TRAMP components and Nrd1 (Figure 2B). We observed a robust, RNA-independent interaction between rNrd1 and rTrf4, even in the absence of rAir2 and rMtr4. Importantly, rNrd1ΔCID failed to interact with rTRAMP (Figure S3B), and recombinant isolated CID (rCID-His$_6$) alone was efficiently pulled down by rTrf4 (Figure 2C). Taken together, these results demonstrate that Nrd1p recognizes the TRAMP complex via a direct interac-

tion with Trf4p, the CID domain being necessary and sufficient for this interaction.

## The Nrd1p Interaction Motif Is a Short CTD-like Domain in Trf4p

Since the CID interacts with the RNAPII CTD and Trf4p, we considered that the same surface might be involved in the recognition of both targets. Consistent with this notion, two CID variants mutated at positions that are critical for binding to the Ser5P CTD (S25D and R28D; Kubicek et al., 2012) also failed to interact with rTrf4 (Figure 2C). Therefore, we surmised that the Nrd1p CID might recognize a CTD mimic in Trf4p. Because the CTD is intrinsically disordered, we restricted our search to the unstructured N- and C-terminal regions of Trf4p. We found a 9 aa motif at the very C-terminal end of Trf4p, which contains several residues that are critical in the CTD for the interaction with the CID, including a glutamate that could mimic the Ser5P in CTD repeats (Figures 2D and S3C).

To assess the role of this motif, we immunoprecipitated Nrd1-TAP from strains expressing a mutant variant of Trf4p lacking the last nine C-terminal amino acids and the two adjacent aspartic acid residues, which we surmised to be important for binding based on the CID structure (Kubicek et al., 2012). Despite identical steady-state levels of the wild-type and mutant Trf4p, the interaction with Nrd1p was abolished in the mutant, indicating that the C-terminal region is necessary (Figure 2E). Therefore, we dubbed this motif NIM, for Nrd1p interaction motif.

We set out to test whether the lack of interaction between Nrd1p and Trf4p contributes to the degradation/processing defects observed in the *nrd1-ΔCID* mutant (Figure S2). To this end, we analyzed by northern blot the effect of the NIM deletion on the levels of the *NEL025c* CUT and the *SNR13* precursor. As shown in Figure 2F, in *trf4ΔNIM* cells these NNS targets were stabilized, although to levels slightly lower than those in an *nrd1ΔCID* mutant. As for deletion of the CID, deletion of the NIM exacerbates the degradation/processing phenotype of exosome defective cells (*dis3-exo-*; Figure 2F). We did not observe any significant effect of the NIM deletion on termination based on the detection of readthrough transcripts or RNAPII ChIP (Figures 2F and S3E and data not shown), suggesting that the Nrd1p-Trf4p interaction is not required for transcription termination. Taken together, these results demonstrate that Nrd1p recognizes a CTD-like motif, NIM, in the Trf4p C-terminal region via its CID domain and that this interaction contributes to the role of the NNS complex in promoting RNA degradation by the Rrp6-exosome.

## The CID-CTD and CID-NIM Interactions Are Mutually Exclusive

Previous studies have demonstrated that the CID of Nrd1p binds to a fragment consisting of two canonical CTD repeats with the

(F) Stabilization of NNS targets in *trf4ΔNIM* cells. Analysis of *NEL025C* and pre-snR13 RNAs by northern blot in the presence or absence of the NIM, in an exosome-defective (Dis3p catalytic mutant, *dis3-exo⁻*; left panels) or an otherwise wild-type background (right panels). Note that the rightmost panels were more exposed than the other panels to visualize the poorly detectable *NEL025C* and pre-snR13 transcripts in a strain wild-type for the nuclear exosome. Stabilization values in *dis3-exo⁻/trf4ΔNIM* relative to *dis3-exo⁻* are 1.9 ± 0.37 and 2.5 ± 0.56 for pre-snR13 and *NEL025c*, respectively (average and SD from three independent samples). Stabilization of pre-snR13 and *NEL025c* was consistently observed in *trf4ΔNIM* cells but could not be reliably quantified due to the low levels of these RNAs in a wild-type background.
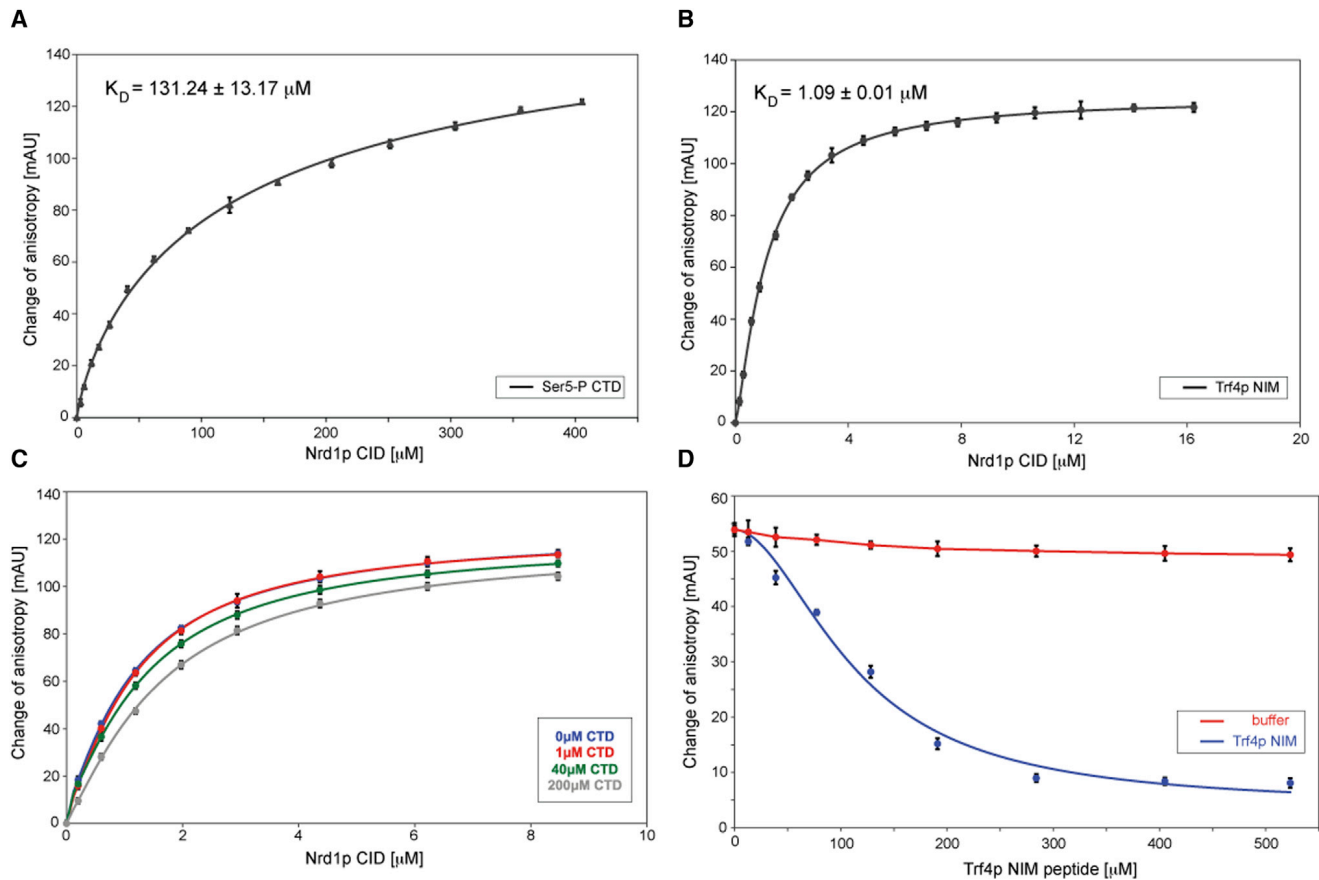
**Figure 3. Fluorescence Anisotropy Analyses of Nrd1p CID Binding to Trf4 NIM and the CTD**

(A and B) Equilibrium binding of Nrd1p CID with Ser5P CTD (A) and Trf4p NIM (B) fluorescently labeled peptides monitored by fluorescence anisotropy (FA). Binding isotherms and dissociation constant ($K_D$) are shown.

(C) FA competition assays between Ser5P CTD and Trf4p NIM for binding to Nrd1p CID. Samples containing 10 nM FAM-labeled Trf4p NIM peptide and 0 μM (blue), 1 μM (red), 40 μM (green), or 200 μM (gray) of unlabeled Ser5P CTD were titrated with Nrd1p CID. Displacement of the binding isotherm with increasing concentration of Ser5P CTD indicates competition for binding to Nrd1p CID.

(D) FA competition assays between pSer5 CTD and Trf4p NIM for binding to Nrd1p CID with a different experimental setup compared to (C). Preformed complex of 10 nM FAM-labeled Ser5P CTD and Nrd1p CID (120 μM final protein concentration) was titrated with different amounts of Trf4p NIM peptide (blue) or buffer (red) as a control. The decrease of fluorescence anisotropy reflects the disassembly of the Ser5P CTD-Nrd1p CID complex.

Ser5P mark located in the upstream repeat (Kubicek et al., 2012). This phospho-CTD fragment binds Nrd1p CID with a $K_D$ of ~130 μM (Figure 3A). To assess the binding affinity of Nrd1p CID to the NIM, we performed a quantitative solution-binding assay using fluorescence anisotropy (FA) measurements. We found that Nrd1p CID binds NIM with an affinity roughly 100-fold stronger compared to the phospho-CTD peptide (a $K_D$ of ~1 μM; Figure 3B). To assess whether the interactions of the CID with CTD and Trf4p are mutually exclusive, we first performed titration of fluorescently labeled NIM peptides with Nrd1p CID in the absence or presence of unlabeled CTD (Figure 3C). The displacement of the binding isotherm in the presence of increasing CTD concentrations demonstrates that the CTD can outcompete Nrd1p CID from binding to the NIM, although the competition was only partially effective, as expected from the stronger affinity of the NIM for the CID. Importantly, increasing concentrations of unlabeled NIM could effectively disassemble a preformed Nrd1p CID-CTD complex

in which the CTD was fluorescently labeled (Figure 3D). The latter competition assay was used to calculate a $K_D$ of ~120 μM for Nrd1p CID-Ser5P CTD complex (using a $K_D$ of 1.08 μM for Nrd1p CID-Trf4p NIM interaction), which is similar to the noncompetitive binding assay (Figure 3B). In both titration experiments we observed no additional increase of anisotropy, indicating that the Nrd1p CID-CTD-NIM ternary complex is not formed. Altogether, the FA data showed that the interactions of Nrd1p CID with the CTD and the NIM are mutually exclusive.

**Solution Structure of Nrd1p CID Bound to Trf4p NIM**

To understand how Trf4p is recognized by Nrd1p, we determined the solution structure of a reconstituted complex consisting of the CID (residues 1–153) of Nrd1p and a 12 aa NIM peptide (Asp573-Asp574-Asp575-Glu576-Asp577-Gly578-Tyr579-Asn580-Pro581-Tyr582-Thr583-Leu584) (Figures 4 and S4; Table 1). The structure of Nrd1p CID consists of eight α helices in a right-handed superhelical arrangement
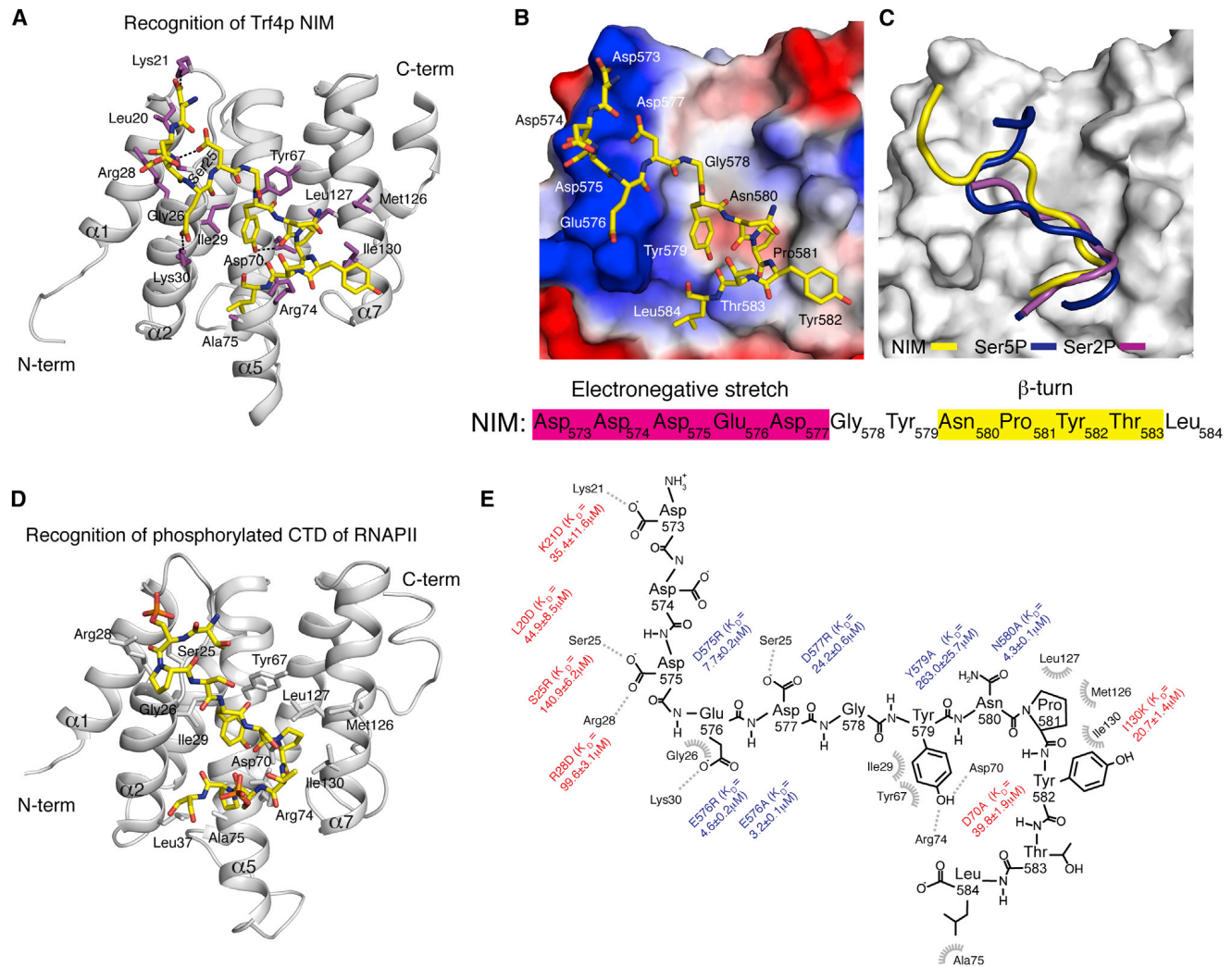
**Figure 4. Recognition of the NIM Peptide by Nrd1p CID**

(A) Solution structure of Nrd1p CID bound to the NIM peptide. The NIM peptide is represented in yellow sticks (only nonhydrogen atoms are shown), and the protein is shown as a gray ribbon model. Protein residues that form hydrophobic contacts and putative hydrogen bonds to the NIM peptide are shown in magenta sticks.

(B) Electrostatic surface representation of Nrd1p CID (electropositive in blue, electronegative in red, neutral in white) with the NIM peptide (represented in yellow sticks; only nonhydrogen atoms are shown). The upstream electronegative stretch of NIM interacts with the electropositive pocket of Nrd1p CID, while the β turn conformation formed by Asn580-Pro581-Tyr582-Thr583 docks in a hydrophobic pocket of Nrd1p CID.

(C) Superposition of Nrd1p CID-Ser5P CTD (blue), Nrd1p CID-Trf4p NIM (yellow), and Pcf11p CID-Ser2P CTD (magenta) complexes, displaying only peptide ribbons on the surface of Nrd1p CID. The comparison highlights the β turn conformation recognition of CTDs and NIM by the CIDs.

(D) Solution structure of Nrd1p CID bound to the Ser5P CTD peptide. The phospho-CTD peptide is represented in yellow sticks (only nonhydrogen atoms are shown), and the protein is shown as a gray ribbon model. Protein residues that form hydrophobic contacts and putative hydrogen bonds to the phospho-CTD peptide are shown in gray sticks.

(E) Scheme showing contacts and energetics between the NIM peptide and Nrd1 CID. Equilibrium binding experiments with both the protein and peptide mutants (in red and blue, respectively) were monitored by FA (for the binding isotherms, see Figure S5). Other residues involved in the canonical CTD-CID interface were mutated previously (Kubicek et al., 2012; Vasiljeva et al., 2008). L20D mutant disrupts the hydrophobic contact with Phe17 and impairs the overall geometry of the α1-α2 loop that contributes to the interaction with the upstream electronegative stretch of NIM.

(Figure 4) and is similar to the structure of Nrd1p CID in the apo form (Vasiljeva et al., 2008) or bound to the phosphorylated CTD (Kubicek et al., 2012). The subtle differences originate from the extended interaction surface with Trf4p NIM, which involves loop α1-α2 and helices α2, α4, and α7 of Nrd1p CID (Figures 4A and S4A–S4G). Interestingly, [$^{1}$H,$^{15}$N] heteronuclear single

quantum coherence (HSQC) titration experiments of Nrd1p CID revealed that the protein amide resonances are in fast or slow exchange regimes between their free and bound forms relative to NMR timescale, when titrated with the phosphorylated CTD or NIM, respectively (Figures S4C–S4E). This observation is in agreement with the fact that the two substrates differ in their

**Table 1. NMR and Refinement Statistics for the Nrd1p CID-Trf4p NIM Complex**

| Nrd1p CID-Trf4p NIM Complex | |
|---|---|
| **NMR Distance and Dihedral Constraints** | |
| Distance restraints | |
| Total NOEs | 2,440 |
| Intraresidue $|i-j| = 0$ | 602 |
| Sequential $|i-j| = 1$ | 661 |
| Medium range $1 < |i-j| < 5$ | 700 |
| Long range $|i-j| \geq 5$ | 477 |
| Hydrogen bonds | |
| Intermolecular distance restraints | 54 |
| Total dihedral angle restraints[a] | 222 |
| **Structure Statistics[b]** | |
| Violations (mean and SD) | |
| Number of distance restraint violations >0.5 Å | 0 |
| Number of dihedral angle restraint violations >15° | 0 |
| Maximum dihedral angle restraint violation (°) | 6.67 ± 1.89 |
| Maximum distance constraint violation (Å) | 0.34 ± 0.12 |
| Deviations from idealized geometry[b] | |
| Bond lengths (Å) | 0.0035 ± 0.0001 |
| Bond angles (°) | 1.6 ± 0.02 |
| Average pairwise r.m.s.d (Å)[b] | |
| Complex | |
| Heavy atoms | 1.17 ± 0.08 |
| Backbone atoms | 0.72 ± 0.10 |
| Ramachandran plot statistics[c] | |
| Residues in most-favored regions (%) | 72.6 |
| Residues in additionally allowed regions (%) | 26.2 |
| Residues in generously allowed regions (%) | 0.6 |
| Residues in disallowed regions (%) | 0.6 |

[a] α-helical dihedral angle restraints imposed for the backbone based on the CSI.
[b] Calculated for an ensemble of the 20 lowest-energy structures.
[c] Based on PROCHECK analysis (Laskowski et al., 1996).

binding affinities to the CID by two orders of magnitude as evidenced by the FA data.

The structure of the Nrd1p CID-Trf4p NIM complex shows that the specific recognition of the NIM is facilitated by a hydrophobic β turn in the C-terminal region and negatively charged residues in the N-terminal region of the peptide (Figure 4B). These two elements are conserved in budding yeast. Akin to the CTD, the NIM peptide adopts the β turn conformation at Asn580-Pro581-Tyr582-Thr583 (Figure 4C). Pro581 and Tyr582 of the NIM β turn, along with the preceding Tyr579, dock into a hydrophobic pocket of the Nrd1p CID that is formed by Ile29, Tyr67, Met126, Leu127, and Ile130 (Figures 4A and S4G). Tyr579 of the NIM also forms intramolecular stacking with Pro581, and the hydroxyl group of Tyr579 forms a hydrogen bond with a conserved Asp70 of Nrd1p CID (Figure 4A). Alanine substitution at position Tyr579 of the NIM strongly diminished the binding affinity for Nrd1p CID (Figures 4E and S5B), confirming the importance of the intricate interaction network of this residue.

Consistently, the Asp70Ala variant of CID displayed a significant drop in binding affinity for the NIM peptide (Figures 4E and S5A). Tyr582 of the NIM β turn stacks with the methyl groups of Ile130 of the CID, and the perturbation of this interaction yields a 20-fold decrease in binding affinity (Ile130Lys variant).

The N-terminal region of NIM contains a stretch of negatively charged residues (Asp573-Asp574-Asp575-Glu576-Asp577), which contains the putative Ser5P mimic (Glu576) and interacts with a positively charged pocket formed by Lys21, Ser25, and Arg28 in the α1-α2 loop of CID (Figure 4B). In particular, the carboxyl groups of Asp577 and Asp575 of NIM form a hydrogen bond with Ser25 and Arg28 of the CID, respectively (these hydrogen bonds are inferred from the final ensemble of structures, and they are indirectly defined by neighboring proton-proton Nuclear Overhauser Effect [NOE]). Similarly, Asp573 of the NIM contacts Lys21 of the CID via a hydrogen bond between the carboxyl group and the side-chain amino group of Lys21. The importance of these contacts was further tested in a quantitative in vitro binding assay using FA. Aspartate (charge swapping) or arginine substitutions at positions Leu20, Lys21, Ser25, and Arg28 significantly decreased the binding affinity for the NIM peptide (Figures 4A, 4E, and S5A). The equivalent mutations in the NIM also showed a decrease in the binding affinity for Nrd1p CID, albeit with a lower magnitude. Notably, mutation of Glu576 to alanine or to arginine partially affected binding, indicating that the putative phosphomimic is important, but not essential, for the interaction (Figures 4A, 4E, and S5B). This suggests that the flanking aspartates of the Asp-rich stretch may substitute one another in the single-point mutants. Altogether, the structural and binding data show that Trf4p NIM is recognized by Nrd1p CID through two elements, hydrophobic β turn and Asp-rich stretch, which is a recognition mechanism similar to that used for the recognition of phosphorylated CTD.

### The Architecture of Interactions between the NNS Complex, the TRAMP, and the Nuclear Exosome

Having established that Nrd1p interacts directly with Trf4p via the CID-NIM interaction, we undertook the characterization of the interactions linking the NNS complex and TRAMP with the nuclear exosome and Rrp6p. The strong and stable binding of TRAMP to the NNS complex (Figure 2A) contrasts with the weak, mostly RNA-dependent signal that we observed for Rrp6p and the core exosome in Nrd1p immunoprecipitates (Figure S3A). We considered the possibility that Rrp6p and the core exosome might be recruited to the NNS complex via the Nrd1p-TRAMP interaction. However, the molecular details of the interaction between TRAMP and the nuclear exosome/Rrp6p have not been elucidated. Therefore, we first performed pull-down assays using *E. coli* extracts expressing recombinant TRAMP components and halo-tagged Rrp6. RNase treatment was included to prevent detection of RNA-dependent interactions. These experiments revealed a clear and direct interaction between rTrf4 and rRrp6 irrespective of the presence of rMtr4 or rAir2 (Figure 5A). To assess whether TRAMP also interacts with the nuclear exosome independently of Rrp6p, we immunoprecipitated the core exosome using TAP-tagged Rrp41p from wild-type or *Δrrp6* cells. As shown in Figure 5B, we detected significant signals for Trf4p in purified core exosome from wild-type,
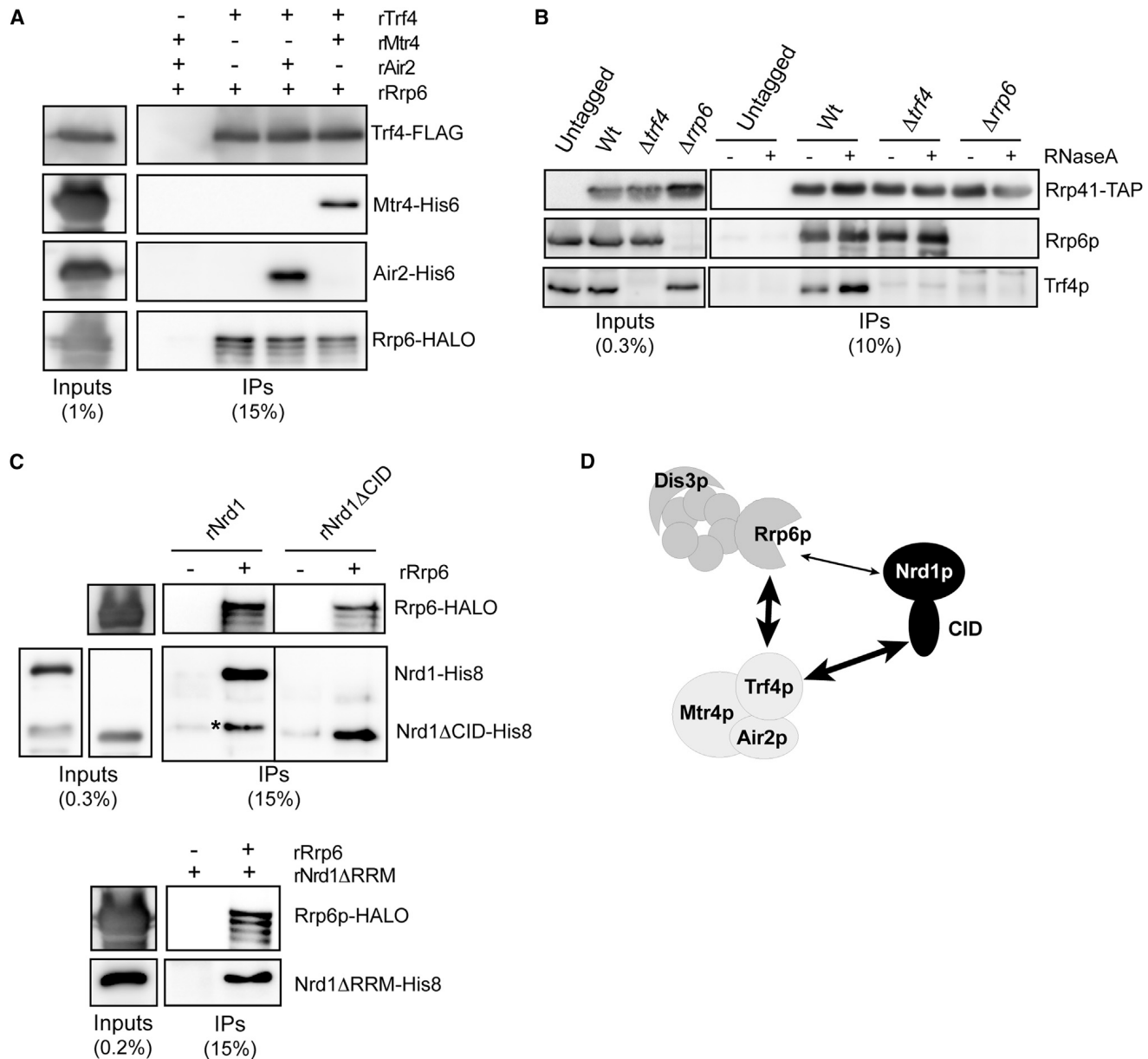
**Figure 5. Analysis of the Interactions between the Exosome, the TRAMP, and Nrd1p**

(A) Pull-down experiments as in Figure 3B, using Trf4-FLAG as a bait and recombinant Rrp6-Halo, or His-tagged Air2 and Mtr4. The fraction of extract and immunoprecipitated material that is loaded on the gel is indicated.

(B) Western blot analysis of factors associated with the core exosome. Rrp41p-TAP eluates purified from wild-type, *trf4Δ*, or *Δrrp6* cells (IPs) were probed with anti-Rrp6 (1:1,000 dilution) and anti-Trf4 antibodies, respectively.

(C) Pull-down experiment using recombinant Rrp6-Halo as bait and His-tagged Nrd1 variants as indicated. An asterisk indicates a proteolytic fragment of rNrd1 that lacks most of the CID domain.

(D) Schematic summarizing the protein-protein interactions identified in this work. A thinner arrow is used to indicate that the interaction between Rrp6p and Nrd1p cannot be detected in vivo.

but not from *Δrrp6* cells, strongly suggesting that Rrp6p bridges TRAMP and the core exosome.

The interaction we detected in vivo between Rrp6p and Nrd1p is strongly dependent on the presence of RNA (Figure S3A). However, we cannot completely exclude the existence of direct but weak or transient contacts in vivo that are not easily detected

under our assay conditions. Therefore, we decided to assess whether a direct interaction could be detected between recombinant Rrp6 and Nrd1. To address the RNA dependency of the interaction, we either treated our extracts with RNase A or used a variant of rNrd1 lacking the RNA binding domain (rNrd1-ΔRRM). As shown in Figure 5C, we observed an interaction

between rRrp6 and rNrd1 or rNrd1-ΔRRM. Importantly, this interaction was not mediated by the CID since it was also detected with rNrd1ΔCID. From these experiments, we conclude that Rrp6p interacts directly with Nrd1 in a CID-independent manner and with Trf4p. The latter interaction allows the association of TRAMP with the nuclear exosome (Figure 5D).

### Nrd1p Interaction with Trf4p Stimulates RNA Polyadenylation by TRAMP

The Trf4-Air2 heterodimer possesses a distributive poly(A) polymerase activity in vitro (LaCava et al., 2005; Vanácová et al., 2005; Wyers et al., 2005). Considering the high affinity of Nrd1p and Nab3p for their RNA targets (Carroll et al., 2007; Hobor et al., 2011; Porrua et al., 2012), and the strong interaction of Nrd1p with Trf4p, we surmised that the Nrd1-Nab3 heterodimer might stimulate polyadenylation of NNS substrates by Trf4p-Air2p, for instance by improving recruitment or by stabilizing the Trf4-Air2 complex on the RNA. Therefore, we assessed the effect of adding rNrd1 and rNab3 to in vitro polyadenylation assays with recombinant Trf4p-Air2p (Figures 6 and S6A). We first used a 40-mer RNA substrate containing Nrd1p and Nab3p binding sites that we have previously shown to efficiently bind the Nrd1-Nab3 heterodimer in vitro and elicit efficient NNS-dependent transcription termination in vivo (Porrua et al., 2012). As shown in Figure 6A, using limiting concentrations of rTrf4-rAir2 relative to the substrate, polyadenylation was markedly stimulated by the addition of purified recombinant Nrd1p-Nab3p, resulting in a longer length of the added poly(A) tails. We did not observe any substantial increase in the fraction of RNA that is polyadenylated in response to the addition of rNrd1-Nab3 (compare the levels of nonadenylated substrate in Figures 6A and 6B), suggesting that, at least in vitro, stimulation preferentially occurs on molecules that have already undergone a polyadenylation cycle.

Stimulation was dependent on the interaction between rNrd1 and rTrf4 because it was abolished by the use of rNrd1ΔCID instead of wild-type rNrd1 (Figure 6B). However, the addition of recombinant CID alone did not enhance the polyadenylation activity of rTrf4-rAir2, ruling out that stimulation could result from an allosteric activation of Trf4p by the CID. The whole rNrd1-rNab3 heterodimer was required because neither subunit alone could enhance polyadenylation; rather, we observed a mild inhibition of rTrf4p-rAir2p activity when adding only rNab3 to the reaction (Figure 6B). Finally, high-affinity binding of rNrd1-Nab3 to the RNA was required because no significant stimulation could be obtained when using a mutant substrate that binds the heterodimer with an affinity ≈20-fold lower than that of the wild-type version (Figures S6B and S6C). Taken together, our results strongly suggest that the simultaneous interaction of Nrd1p-Nab3p with Trf4p-Air2p and the RNA enhances polyadenylation, most likely by favoring or stabilizing the association of Trf4p-Air2p with its substrate.

### DISCUSSION

The Nrd1-Nab3-Sen1 complex is of major biological relevance because of its central role in the biogenesis of snRNAs and snoRNAs and in the control of pervasive transcription in connection with the exosome and TRAMP complexes (Schulz et al., 2013). The CID domain of Nrd1p has retained special attention because of its ability to bind the CTD of RNAPII. However, despite a number of biochemical and structural studies (Vasiljeva et al., 2008; Kubicek et al., 2012), its actual function in vivo has remained mysterious. In this work, we reexamine the role of the CID and demonstrate that it plays important roles in the efficiency of transcription termination and in RNA degradation.

### The Role of Nrd1p CID in Transcription Termination

The genome-wide occupancy of RNAPII upon deletion of the CID reported here clearly reveals a widespread role for this domain in transcription termination of CUTs, snoRNAs, and to some extent, small ORFs (Figures 1 and S1), which is consistent with the known landscape of NNS targets (Steinmetz et al., 2006b). While this work was in progress, Heo et al. (2013) reported similar findings using CID-swapped chimeric constructs. However, because deletion of the CID does not prevent NNS termination, the interaction between Nrd1p and the RNAPII CTD only impacts the efficiency of the process.

Deletion of the CID has a strong impact on Nrd1p recruitment, even in the presence (as in CUTs) of clusters of Nrd1 and Nab3 sites on the nascent RNA, which have been shown to be required for recruitment (Gudipati et al., 2008). This can be best explained if recruitment depends synergistically (and not redundantly) on the interaction of the Nrd1-Nab3 complex with both the CTD and the nascent RNA.

Termination of mRNA coding genes depends on the cleavage and polyadenylation factor/cleavage factor (CPF/CF) that also interacts with the nascent RNA and the CTD. We have previously shown that the CPF and the NNS complex have partially overlapping sequence requirements and that the same RNA sequence can be used as a terminator by either complex depending only on the distance from the transcriptional start (Gudipati et al., 2008; Porrua et al., 2012). It is likely that early recruitment of the NNS complex via the Nrd1p CID-CTD interaction kinetically favors the appropriate recognition of RNA binding sites that could otherwise be bound by the CPF complex, impairing termination by the NNS pathway.

### The CID Mediates the Connection between the NNS Complex and TRAMP

Our results strongly suggest that the function of the CID in RNA degradation relies on the interaction with a CTD-like motif in Trf4p. The strong association between the termination complex and the corresponding poly(A) polymerase mirrors the association of Pap1p with the CPF complex, suggesting that in both cases the appropriate poly(A) polymerase is brought in the proximity of the 3′-OH of the newly released (or cleaved) RNA, thus preventing or limiting spurious cross-processing events.

The two alternative forms of TRAMP containing either Trf4p or Trf5p have partially redundant functions, but the former predominates in the degradation of CUTs and the processing of snRNAs and snoRNAs (San Paolo et al., 2009). Because Trf5p does not contain a NIM (Figure S3D), this different target specificity can now also be explained by the interaction of the NNS complex with Trf4p.
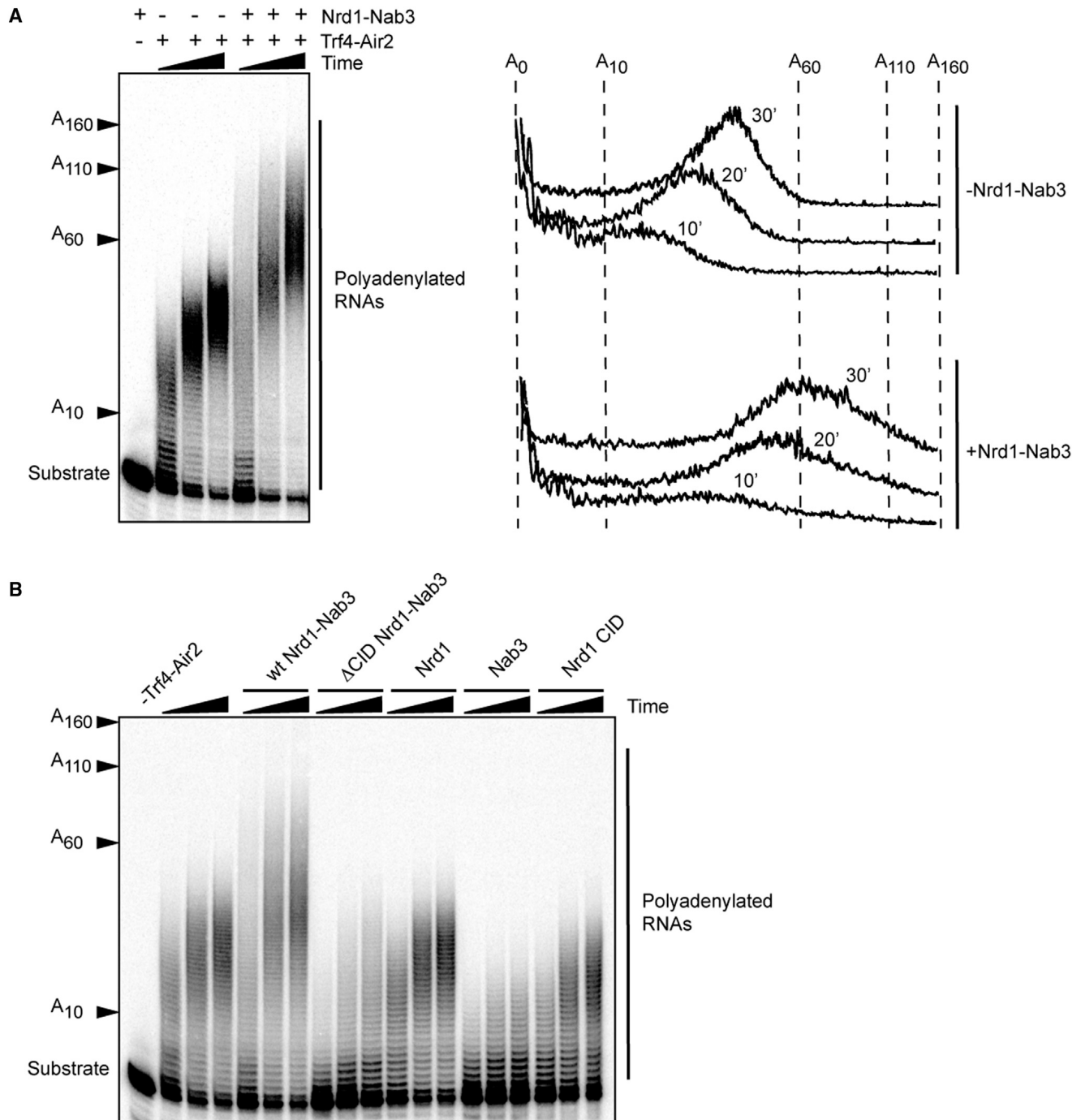
**Figure 6. Analysis of the Effect of rNrd1-Nab3 on the Polyadenylation Activity of rTrf4-rAir2 In Vitro**

(A) Polyadenylation assays with recombinant Trf4-Air2 in the absence or in the presence of recombinant Nrd1-Nab3. Left: PAGE analysis of polyadenylated species at 10, 20, and 30 min reaction time. The position of the substrate and the number of added As is indicated. Right: lane scans of the gel shown on the left.

(B) Polyadenylation assays as in (A) to individually assess the role of rNrd1, rNab3, and the Nrd1 CID domain in stimulating polyadenylation by rTrf4-rAir2. Reactions were performed with rTrf4-rAir2 in the presence of the indicated proteins or protein complex.

In contrast to the robust interaction of the NNS complex with TRAMP, we found the interaction of Rrp6p and the exosome with the NNS complex in vivo to be strongly RNA dependent (Figure S3A), which is in apparent discrepancy with a previous report (Vasiljeva and Buratowski, 2006). Nevertheless, we could show a direct interaction between Rrp6p and Nrd1p in vitro that is

independent from the CID, suggesting that this interaction is too weak to withstand our immunoprecipitation conditions or that it forms only transiently in vivo.

## Functional Significance of the NNS-TRAMP Interaction in Degradation

It is known that the RNA quality control factors target a vast repertoire of defective molecules that are sorted because of kinetic competition between RNA degradation and processing (Bousquet-Antonelli et al., 2000; Gudipati et al., 2012). However, when alternative routes to discarding are not desired, it is crucial to enforce degradation by the use of specific adaptors.

The NNS complex fulfils such a function by recruiting the TRAMP to its targets after transcription termination via the CID-NIM interaction. This could stimulate degradation because of enhanced polyadenylation of the substrates, but it is also possible that degradation is stimulated by a poly(A)-independent mechanism (Callahan and Butler, 2010; LaCava et al., 2005; Wyers et al., 2005), maybe by recruiting the exosome by virtue of the direct Trf4p-Rrp6p interaction (Figure 5).

We note that stimulation preferentially occurs on a fraction of already polyadenylated RNAs rather than on the nonadenylated substrate. It is possible that Nrd1p-Nab3p and Trf4p-Air2p compete for binding to the short RNA substrates we used in our assays, which is also suggested by the inhibition of the polyadenylation reaction when rNab3 alone or rNrd1ΔCID-Nab3 are used (Figure 6B). The emergence of a poly(A) tail likely provides a binding platform in the vicinity of the substrate 3′-OH that would be preferentially bound by Trf4p-Air2p.

It has previously been suggested that the RNA helicase Mtr4p modulates the activity of Trf4p-Air2p, restricting the addition of poly(A) tails after 3–4 nt in vitro (Jia et al., 2011). Although very short poly(A) tails (≈ 4 nt) have also been observed in vivo (Jia et al., 2011; Wlotzka et al., 2011), it must be noticed that these represent average steady-state lengths, resulting from an equilibrium between synthesis and degradation. When degradation is impaired, Trf4p-dependent poly(A) tails in vivo are longer (Wyers et al., 2005; D.L., unpublished data), most likely within the range required to allow threading of substrates through the exosome channel (i.e., roughly 30 nt). The antagonistic impact of the Nrd1-Nab3 complex and Mtr4p on Trf4p activity might imply a temporal regulation of polyadenylation, restricted by Mtr4p early after transcription to prevent the binding of Pab1p (Jia et al., 2011) and stimulated later on by Nrd1p-Nab3p to favor degradation. Alternatively, Mtr4p and Nrd1p-Nab3p might modulate Trf4p activity at different substrates.

## Recognition of Hydrophobic β Turn Hairpin and Electronegative Stretch by the CID

The CTD and NIM share a sequence element that can form a β turn (Figure 4). The binding mode of the NIM peptide at the β turn conformation resembles other previously determined structures of CTD bound to CIDs of Pcf11p, SCAF8, Rtt103p, and Nrd1p (Figures 4C and S7) (Meinhart and Cramer, 2004; Becker et al., 2008; Lunde et al., 2010; Kubicek et al., 2012). However, in contrast to these CID-CTD complexes, the β turn of the NIM peptide has more extensive hydrophobic contacts with Nrd1p due to the presence of Tyr582, the third residue of the β turn not present in any of the CTD repeats (Jasnovidova and Stefl, 2013). The stacking of Tyr582 with Ile130 of Nrd1p significantly contributes to the overall increase of binding affinity of NIM to Nrd1p in comparison to the phosphorylated CTD. The other region of the NIM peptide that contributes to the overall affinity is the aspartate-rich region located at the N terminus that intimately interacts with the electronegative pocket of Nrd1p (Figure 4B). To some extent, this mimics the recognition of phosphorylated CTD, but it involves more contacts, strengthening the overall binding affinity (Figure 4C). It is also likely that other proteins contain these two elements with the same arrangement and therefore could interact with CID-containing proteins in a similar manner.

## A Role for the CID in Coordinating Transcription Termination and RNA Degradation

Our structural data together with our competition assays demonstrate that the interactions of the CID with the CTD and the NIM are mutually exclusive (Figures 4 and 5). This implies the existence of at least two distinct forms of the NNS complex, one associated with the polymerase and the other associated with the TRAMP, which could represent pre- and posttermination forms of the NNS complex. Although we show that the affinity of the CID for the NIM is 100-fold higher than that for the CTD, the real balance between the two alternative complexes also depends on the number of RNAPII and Trf4p molecules available for interaction and, importantly, on the number of interaction targets, which is presumably higher for RNAPII (25 possible diheptapeptides in the CTD).

We propose that by virtue of its alternative interactions, the CID controls the handover of the NNS complex from RNAPII to the TRAMP, which would temporally coordinate the two functions of the complex. Regulation of NNS and TRAMP function might be reciprocal because interaction with the TRAMP might control the release of Nrd1p from RNAPII (Figure 7). This could be important for the downstream function of the NNS complex in processing/degradation, but also for making the complex available for the interaction with new elongating RNAPIIs.

Our results open up the interesting possibility that the dynamics of factors interacting with the CTD throughout the transcription cycle is regulated not only by the enzymes responsible for CTD modifications and proline isomerizations, but also by competitive interactions with proteins containing CTD-like motifs.

### EXPERIMENTAL PROCEDURES

#### Yeast Strains, Plasmids, and Standard Analyses

ChIP and ChIP-chip experiments were performed as previously described (Mayer et al., 2010; Rougemaille et al., 2008). More complete details for these experimental procedures as well as construction of plasmids, yeast strains, RNA analyses, and standard biochemical analyses can be found in the Supplemental Experimental Procedures.

#### ChIP-Chip Occupancy Profiling

ChIP-chip data analysis was performed essentially as described (Mayer et al., 2010). Briefly, we first performed quantile normalization between replicate measurements and averaged the signal for each probe over the replicate intensities. ChIP enrichments were obtained by dividing ChIP intensities by the corresponding input intensities. The normalized ChIP signal at each nucleotide was calculated as the median signal for all probes overlapping this
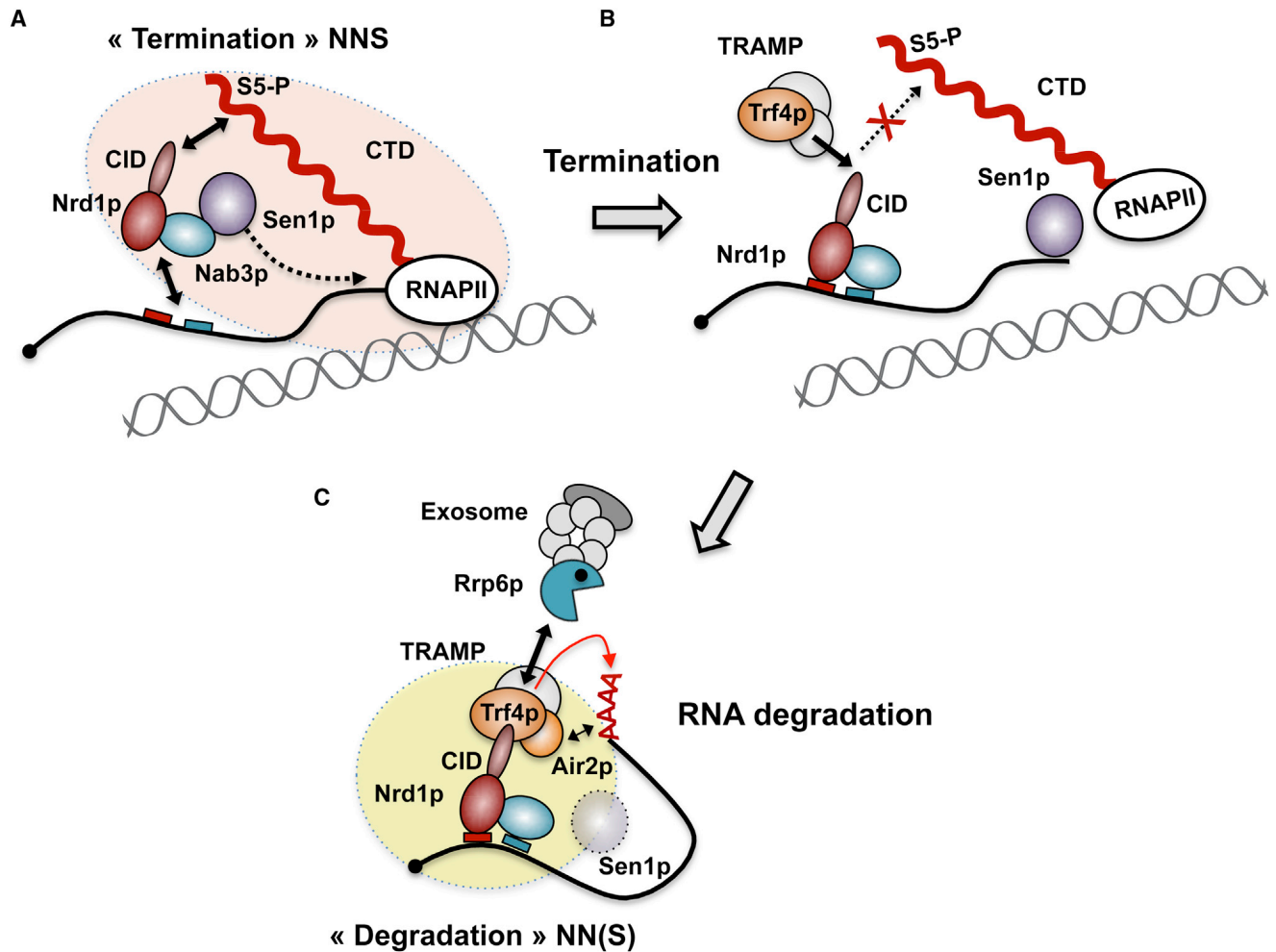
**Figure 7. Model for the Coordination of Transcription Termination with RNA Degradation at NNS Targets**

(A) The NNS complex interacts with the Ser5P form of the CTD via the CID of Nrd1p and with the nascent RNA (boxes indicate recognition motifs for the Nrd1-Nab3 heterodimer), which defines a termination form of the complex (orange shaded).

(B) Concomitantly with or subsequently to dissociation of the elongation complex by Sen1p, Trf4p interacts with the CID of Nrd1p, replacing the CID-CTD interaction and allowing the release of Nrd1p-Nab3p from RNAPII.

(C) Polyadenylation of the transcript by TRAMP (in the presence or absence of Mtr4p) is stimulated by the simultaneous interaction of Nrd1p-Nab3p with the RNA and with Trf4p. Trf4p also might recruit the exosome via the interaction with Rrp6p, possibly favoring subsequent degradation of the transcript in a polyadenylation- independent manner. The alternative interaction of Nrd1p with Trf4p (instead of the CTD) defines a degradation form of the NNS complex (yellow shaded) that might or might not contain Sen1p.

position. Profiles were smoothed using running median smoothing with a window half size of 75 bp. To average profiles over feature classes, features were aligned at their transcription start sites (TSSs) and transcription termination sites (TTSs) or poly(A) addition sites for protein-coding genes, scaled to the same length (i.e., the median length of all transcripts in the class), and averaged by calculating the 5% trimmed mean at each genomic position. For better comparison of CUTs and small-sized protein-coding genes, features of both classes with a length between 350 and 550 bp were selected and scaled to a length of 450 bp. Note that annotation of CUTs does not generally take into account the 3′ end heterogeneity that is characteristic of these transcripts. Medium-sized protein-coding genes were selected by taking the 50% most highly expressed genes (Dengl et al., 2009) that were at least 200 bp away from neighboring genes, with an ORF length of 1,238 ± 300 bp.

Note that because the size of CUTs and snoRNAs is generally similar to or lower than the resolution of the ChIP technique (200–300 nt), the increased downstream RNAPII signal also bleeds over the body of the metagene. The apparent RNAPII increase within CUTs is also due to the fact that these genes have multiple termination sites (Neil et al., 2009; Wyers et al., 2005) that fall within the coordinates of each annotation (and therefore of the metagene). The possibility that deletion of the CID generally affects transcription initiation at CUTs and snoRNAs is very unlikely because (i) the RNAPII increase in *nrd1ΔCID* relative to the WT is always minimal at the 5′ end and progressively increases toward the termination region (Figures 1A–1C and 1E-F) and (ii) we did not detect a significant increase in the levels of mature snRNAs (data not shown).

### NMR Spectroscopy

All NMR spectra for the backbone and side-chain assignments of 0.5–2.0 mM uniformly $^{15}$N,$^{13}$C-labeled Nrd1p CID in 50 mM sodium phosphate buffer (pH 8.0), 100 mM NaCl (90% $H_2O$/10% $D_2O$) were recorded on Bruker AVANCE 700 and 950 MHz spectrometers equipped with a cryoprobe at a

sample temperature of 20°C. The spectra were processed using NMRPipe package (Delaglio et al., 1995), and the protein resonances were assigned manually using Sparky software (T.G. Goddard and D.G. Kellner, University of California, San Francisco). The $^1$H, $^{13}$C, and $^{15}$N chemical shifts of the bound form of Nrd1p CID were assigned as described previously (Kubíček et al., 2011; Kubicek et al., 2012). All distance constraints were derived from the 3D $^{15}$N- and $^{13}$C-separated NOESYs and 2D $^1$H-$^1$H NOESY (with mixing time of 80 ms) collected on a 950 MHz spectrometer. Intermolecular distance constraints were derived from the 3D F$_1$-$^{13}$C/$^{15}$N-filtered NOESY-[$^{13}$C,$^1$H]-HSQC experiment (Peterson et al., 2004; Zwahlen et al., 1997), with a mixing time of 150 ms on a 950 MHz spectrometer. Intramolecular distance constraints of the bound Trf4p NIM peptide (unlabeled) were derived from a 2D F$_1$, F$_2$-$^{13}$C/$^{15}$N-filtered [$^1$H,$^1$H]-NOESY (t$_m$ = 150 ms) (Peterson et al., 2004; Zwahlen et al., 1997). The NOEs were semiquantitatively classified based on their intensities in the 2D and 3D NOESY spectra. The structure determination was performed as described previously (Kubicek et al., 2012).

### Polyadenylation Assays

Polyadenylation reactions were performed at 30°C in a final volume of 20 μl containing 2 nM 5′ end-labeled RNA substrate and 1 nM recombinant Trf4-Air2 in 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.5 mM MgCl$_2$, 10% glycerol, 0.01% nonidet P-40, and 1 mM dithiothreitol in the presence of RNase inhibitors. Reactions were started upon addition of 2 μl of an ATP-MgCl$_2$ mixture (20 mM each) and stopped at different time points by collecting 4 μl aliquots and mixing them with an equal volume of loading buffer (80% formamide, 0.05% w/v bromophenol blue, and 0.05% w/v xylen cyanol). RNAs were denatured for 5 min at 75°C and separated by 10% (w/v) denaturing PAGE. After electrophoresis, gels were dried and analyzed using a Phosphorimager scanner (GE Healthcare). To assess the effect of rNrd1 and rNab3 on the polyadenylation activity of rTrf4-Air2, individual proteins or the heterodimeric complex were added to the reaction at a 3 nM final concentration and incubated for 10 min at 30°C before starting the reaction.

### ACCESSION NUMBERS

The genome-wide ChIP-chip data are accessible with the number E-MTAB-2175.

The atomic coordinates and restraints for the Nrd1p CID-Trf4p NIM complex have been deposited in the Protein Data Bank under ID code 2MOW.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and seven figures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2014.05.031.

### AUTHOR CONTRIBUTIONS

A.T. and O.P. designed and performed molecular biology, genetic, and biochemistry experiments. T.K. performed biochemical and structural analyses. M.L. performed genome-wide ChIP-chip experiments and analyzed the data. K.K. carried out NMR experiments and contributed to NMR data analyses. A.F. performed ChIP experiments. F.L. constructed strains. S.V. designed experiments. P.C. designed experiments. R.S. designed experiments and contributed to structural analyses. D.L. designed experiments and analyzed the data. All authors discussed the results. R.S. directed the research for all the studies performed at Masaryk University, notably the determination of the solution structure and part of the biochemical analyses. O.P. directed part of the research performed in the Centre de Génétique Moléculaire. Specifically, she directed the work of A.T. and designed experimental strategies relating to the discovery and functional characterization of the NIM. D.L. partially directed the work performed in the Centre de Génétique Moléculaire, contributing to the definition of the general experimental strategies and to the coordination of the research performed in the two main sites. D.L. also directed the work on the genome-wide analyses of Nrd1ΔCID and RNAPII distribution. O.P., R.S., and D.L. wrote the manuscript.

### REFERENCES

Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Mol. Cell 23, 841–851.

Becker, R., Loll, B., and Meinhart, A. (2008). Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. J. Biol. Chem. 283, 22659–22669.

Bousquet-Antonelli, C., Presutti, C., and Tollervey, D. (2000). Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell 102, 765–775.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell 36, 541–546.

Callahan, K.P., and Butler, J.S. (2010). TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. J. Biol. Chem. 285, 3540–3547.

Carroll, K.L., Ghirlando, R., Ames, J.M., and Corden, J.L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. RNA 13, 361–373.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: the exosome. Biochim. Biophys. Acta 1829, 552–560.

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet. 7, e1002329.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277–293.

Dengl, S., Mayer, A., Sun, M., and Cramer, P. (2009). Structure and in vivo requirement of the yeast Spt6 SH2 domain. J. Mol. Biol. 389, 211–225.

Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat. Struct. Mol. Biol. 15, 786–794.

Gudipati, R.K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012). Extensive degradation of RNA precursors by the exosome in wild-type cells. Mol. Cell 48, 409–421.

Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. Mol. Cell *49*, 55–66.

Heo, D.H., Yoo, I., Kong, J., Lidschreiber, M., Mayer, A., Choi, B.-Y., Hahn, Y., Cramer, P., Buratowski, S., and Kim, M. (2013). The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing by the nuclear exosome. J. Biol. Chem. *288*, 36676–36690.

Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., and Stefl, R. (2011). Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. J. Biol. Chem. *286*, 3645–3657.

Jasnovidova, O., and Stefl, R. (2013). The CTD code of RNA polymerase II: a structural view. Wiley Interdiscip Rev RNA *4*, 1–16.

Jenks, M.H., O'Rourke, T.W., and Reines, D. (2008). Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. Mol. Cell. Biol. *28*, 3883–3893.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. Mol. Cell *52*, 473–484.

Jia, H., Wang, X., Liu, F., Guenther, U.-P., Srinivasan, S., Anderson, J.T., and Jankowsky, E. (2011). The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. Cell *145*, 890–901.

Kadaba, S., Krueger, A., Trice, T., Krecic, A.M., Hinnebusch, A.G., and Anderson, J. (2004). Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae. Genes Dev. *18*, 1227–1240.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat. Struct. Mol. Biol. *17*, 1279–1286.

Kubíček, K., Pasulka, J., Černá, H., Löhr, F., and Štefl, R. (2011). 1H, 13C, and 15N resonance assignments for the CTD-interacting domain of Nrd1 bound to Ser5-phosphorylated CTD of RNA polymerase II. Biomol. NMR Assign. *5*, 203–205.

Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. Genes Dev. *26*, 1891–1896.

Kuehner, J.N., and Brow, D.A. (2008). Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. Mol. Cell *31*, 201–211.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell *121*, 713–724.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*, 477–486.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. *17*, 1195–1201.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat. Struct. Mol. Biol. *17*, 1272–1278.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science *336*, 1723–1725.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature *430*, 223–226.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038–1042.

Peterson, R.D., Theimer, C.A., Wu, H., and Feigon, J. (2004). New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. J. Biomol. NMR *28*, 59–67.

Porrua, O., and Libri, D. (2013a). RNA quality control in the nucleus: the Angels' share of RNA. Biochim. Biophys. Acta *1829*, 604–611.

Porrua, O., and Libri, D. (2013b). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nat. Struct. Mol. Biol. *20*, 884–891.

Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R.K., Stefl, R., and Libri, D. (2012). In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. EMBO J. *31*, 3935–3948.

Rougemaille, M., Gudipati, R.K., Olesen, J.R., Thomsen, R., Seraphin, B., Libri, D., and Jensen, T.H. (2007). Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. EMBO J. *26*, 2317–2326.

Rougemaille, M., Dieppois, G., Kisseleva-Romanova, E., Gudipati, R.K., Lemoine, S., Blugeon, C., Boulay, J., Jensen, T.H., Stutz, F., Devaux, F., and Libri, D. (2008). THO/Sub2p functions to coordinate 3′-end processing with gene-nuclear pore association. Cell *135*, 308–321.

San Paolo, S., Vanacova, S., Schenk, L., Scherrer, T., Blank, D., Keller, W., and Gerber, A.P. (2009). Distinct roles of non-canonical poly(A) polymerases in RNA metabolism. PLoS Genet. *5*, e1000555.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell *155*, 1075–1087.

Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3′-end formation of RNA polymerase II transcripts. Nature *413*, 327–331.

Steinmetz, E.J., Ng, S.B., Cloute, J.P., and Brow, D.A. (2006a). cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. Mol. Cell. Biol. *26*, 2688–2696.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006b). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol. Cell *24*, 735–746.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol. Cell *23*, 853–864.

Thiebaut, M., Colin, J., Neil, H., Jacquier, A., Séraphin, B., Lacroute, F., and Libri, D. (2008). Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in S. cerevisiae. Mol. Cell *31*, 671–682.

Vanácová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol. *3*, e189.

Vasiljeva, L., and Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3′ processing of RNA polymerase II transcripts. Mol. Cell *21*, 239–248.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. *15*, 795–804.

Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J. *30*, 1790–1803.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725–737.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R., and Kay, L.E. (1997). Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage λN-peptide/boxB RNA complex. J Am Chem Soc. *119*, 6711–6721.

## 5.7 Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p

In this publication, the solution structure of the Rtt103p CTD-interacting domain (CID) bound to Thr4 phosphorylated CTD, a poorly understood letter of the CTD code, is presented. The structure reveals a direct recognition of the phospho-Thr4 mark by Rtt103p CID and extensive interactions involving residues from three repeats of the CTD heptad. Intriguingly, Rtt103p's CID binds equally well Thr4 and Ser2 phosphorylated CTD, while double phosphorylation of CTD at Ser2 and Thr4 diminishes its binding affinity towards RTT103 due to electrostatic repulsion. The structural data suggest that the recruitment of a CID-containing CTD-binding factor may be coded by more than one letter of the CTD code.

## Scientific Report

# Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p

Olga Jasnovidova[*] iD, Magdalena Krejcikova, Karel Kubicek & Richard Stefl[**] iD

## Abstract

Phosphorylation patterns of the C-terminal domain (CTD) of largest subunit of RNA polymerase II (called the CTD code) orchestrate the recruitment of RNA processing and transcription factors. Recent studies showed that not only serines and tyrosines but also threonines of the CTD can be phosphorylated with a number of functional consequences, including the interaction with yeast transcription termination factor, Rtt103p. Here, we report the solution structure of the Rtt103p CTD-interacting domain (CID) bound to Thr4 phosphorylated CTD, a poorly understood letter of the CTD code. The structure reveals a direct recognition of the phospho-Thr4 mark by Rtt103p CID and extensive interactions involving residues from three repeats of the CTD heptad. Intriguingly, Rtt103p's CID binds equally well Thr4 and Ser2 phosphorylated CTD. A doubly phosphorylated CTD at Ser2 and Thr4 diminishes its binding affinity due to electrostatic repulsion. Our structural data suggest that the recruitment of a CID-containing CTD-binding factor may be coded by more than one letter of the CTD code.

## Introduction

RNA polymerase II (RNAPII) utilizes a long and flexible carboxyl-terminal domain (CTD) of its largest subunit to specifically recruit protein/RNA-binding factors during transcription [1–5]. The CTD consists of tandem repeats with conserved consensus Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 that is repeated 26 times in yeast and 52 times in humans [6]. The CTD sequence is post-translationally phosphorylated at serines (Ser2, Ser5 and Ser7) and Tyr1 in a dynamic manner, yielding specific patterns that are recognized by appropriate factors in coordination with the transcription cycle events [3–5,7].

Additionally another highly conserved position, Thr4, was reported to be phosphorylated both in yeast and humans [8–12]. However, the levels of pThr4 in cells remain controversial based on two recent mass-spectrometry studies [10,11]. Substitution of Thr4 to Ala (T4A) or Val (T4V) is lethal for chicken and human cells [12–14]; however, the same mutants are viable in yeast [9,15,16]. In humans, genomewide studies revealed increasing levels of pThr4 throughout the gene body with the peak after the poly-A site [12]. In agreement with this, the T4A mutant showed defect in transcription elongation [12]. In yeast, the pThr4-mark is enriched along the whole gene body, similarly to the pTyr1-mark [17]. Both marks go down prior recruitment of transcription termination factors [17]. Therefore, it was suggested that the pThr4 mark along with the pTyr1-mark prevent binding of transcription termination factors during transcription elongation [17]. However, recent high resolution ChiP-nexus data suggested a different role for the pThr4 mark involved in transcription termination and post-transcriptional splicing [18].

It has been unclear for a long time what protein factors are recruited through the pThr4 signal. Interestingly, yeast transcription termination factor, Rtt103p, well known to be associated with the pSer2-mark [17,19,20], was identified as a part of the interactome of RNAPII phosphorylated at Thr4 [18]. Based on the overlay of NET-seq and ChIP-nexus profiles, Rtt103p coincides with the pThr4 mark after poly-A site. Both, deletion of the entire Rtt103p protein or expression of Rpb1 T4V CTD mutant, cause similar RNAPII pausing defect after poly-A site. The authors suggested a model, in which both pSer2 and pThr4 marks can contribute to the recruitment of Rtt103p to the poly-A site [18]. This concept is also supported by recent mass-spectrometry analyses of RNAPII CTD population pulled down by Rtt103p, which revealed simultaneous presence of pThr4 and pSer2 marks [11].

To understand the puzzling roles of the pSer2 and pThr4 marks in recruitment of transcription termination factor Rtt103p, we solved NMR structure of the pThr4 CTD peptide in complex with Rtt103p CTD-interacting domain (CID). Our structure reveals for the first time a direct readout of the pThr4 mark within the CTD. We also reveal significantly larger interaction area of Rtt103p with the CTD peptide than previously reported [20]. Next, we show that two adjacently positioned phosphorylations, pSer2 followed by pThr4, inhibit

---

CEITEC – Central European Institute of Technology, Masaryk University, Brno, Czech Republic
*Corresponding author. E-mail: olga.jasnovidova@ceitec.muni.cz
**Corresponding author. Tel: +420 549492436; E-mail: richard.stefl@ceitec.muni.cz

the binding of Rtt103p CID due to a charge–charge repulsion of the two closely positioned phosphate moieties. Finally, we propose that the CTD code is degenerated, as Rtt103p reads the pThr4 and pSer2 marks equally well using the same molecular mechanism.

## Results and Discussion

### Rtt103p CID binds equally well Thr4 and Ser2 phosphomarks

To test the binding affinity of Rtt103p CID towards pThr4-CTD *in vitro,* we performed an equilibrium-binding assay using

fluorescence anisotropy (FA) (Fig 1B). The experiment revealed that Rtt103p binds pThr4-CTD with a $K_D$ of $15 \pm 1$ μM, which is 2.5 times weaker binding than to the CTD with the pSer2 mark ($K_D = 6.0 \pm 0.2$ μM). This finding is in a good agreement with previous co-immunoprecipitation studies, where Rtt103p was pulled down by RNAPII with the pThr4 mark and successfully competed out by pSer2-CTD or pThr4-CTD antisera [18]. Doubly phosphorylated pThr4-CTD at both Thr4 displayed increased binding affinity due to avidity effects ($K_D = 6 \pm 0.2$ μM). Remarkably, if the pSer2 and pThr4 marks are positioned adjacently, binding affinity ($K_D = 43 \pm 2$ μM) is lowered almost to the level of non-phosphorylated CTD ($K_D = 64 \pm 2$ μM). The pSer5 phosphorylation mark was
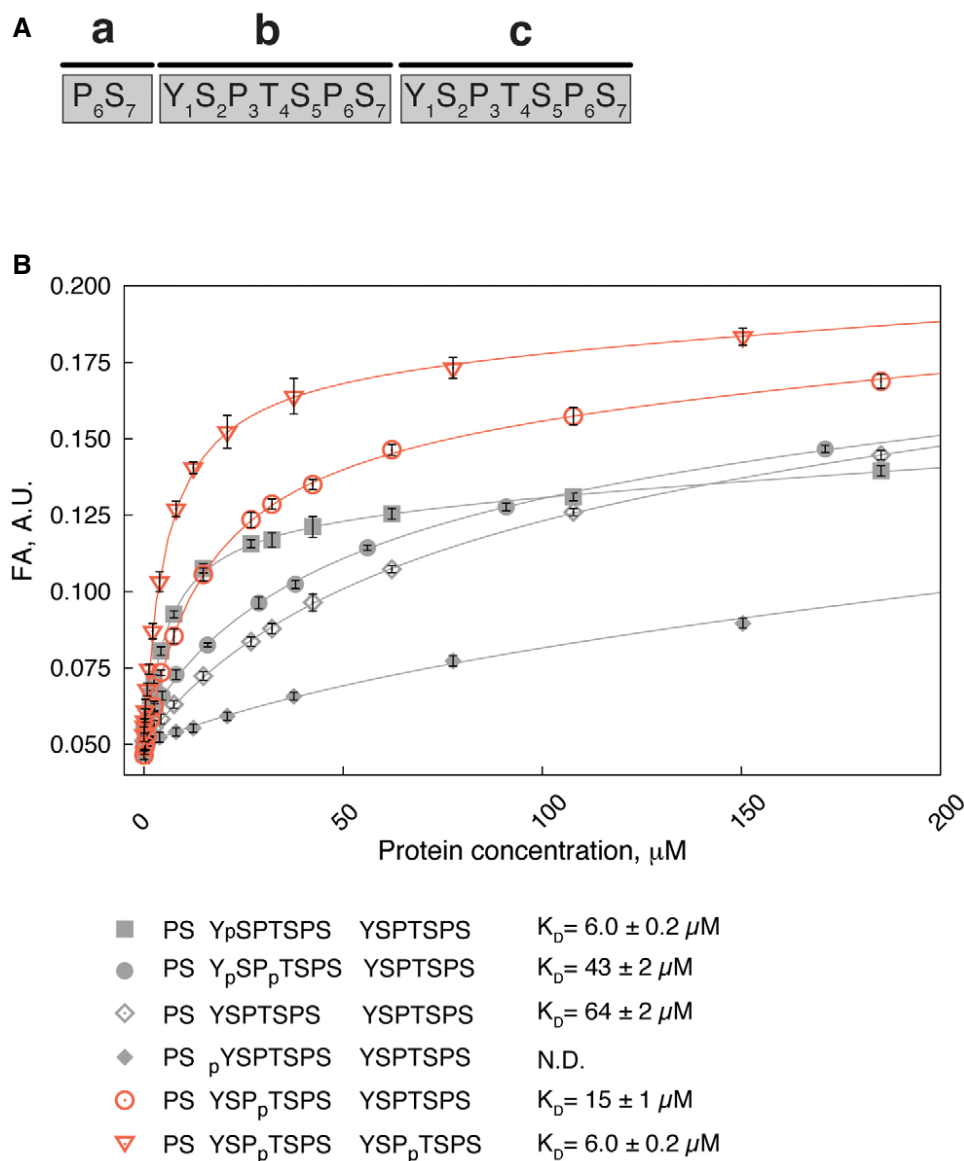


**Figure 1. How CTD phosphorylations modulate binding to Rtt103p CID.**

A   Numbering of residues and order of heptad repeats of the CTD peptide used throughout the study.

B   Equilibrium binding of Rtt103p CID with fluorescently labelled CTD peptides monitored by fluorescence anisotropy (FA). Rtt103p CID titrated into 10 nM FAM-labelled CTD peptides. Peptide sequences, corresponding binding isotherms and dissociation constant ($K_D$, $\pm$ standard deviation of the fit) are shown. FAM, 5,6-carboxyfluorescein. N.D., not determined.

also previously shown to abolish and lower the binding with Rtt103p or its close human homologue [20,21]. Next, we introduced the pTyr1 mark to the central heptad of the CTD peptide, which completely abolished the binding with Rtt103p (Fig 1B). This suggests that $Y_{1b}$ is accommodated in the hydrophobic pocket following the previously established binding model for CIDs [17].

## NMR structure of Rtt103p CID bound to CTD with phospho-threonine mark

To reveal the structural basis of pThr4 recognition, we solved solution structure of a reconstituted complex that harbours Rtt103p CID (3-131) and a 16-amino acid peptide, pThr4-CTD (PS YSP(pT)SPS
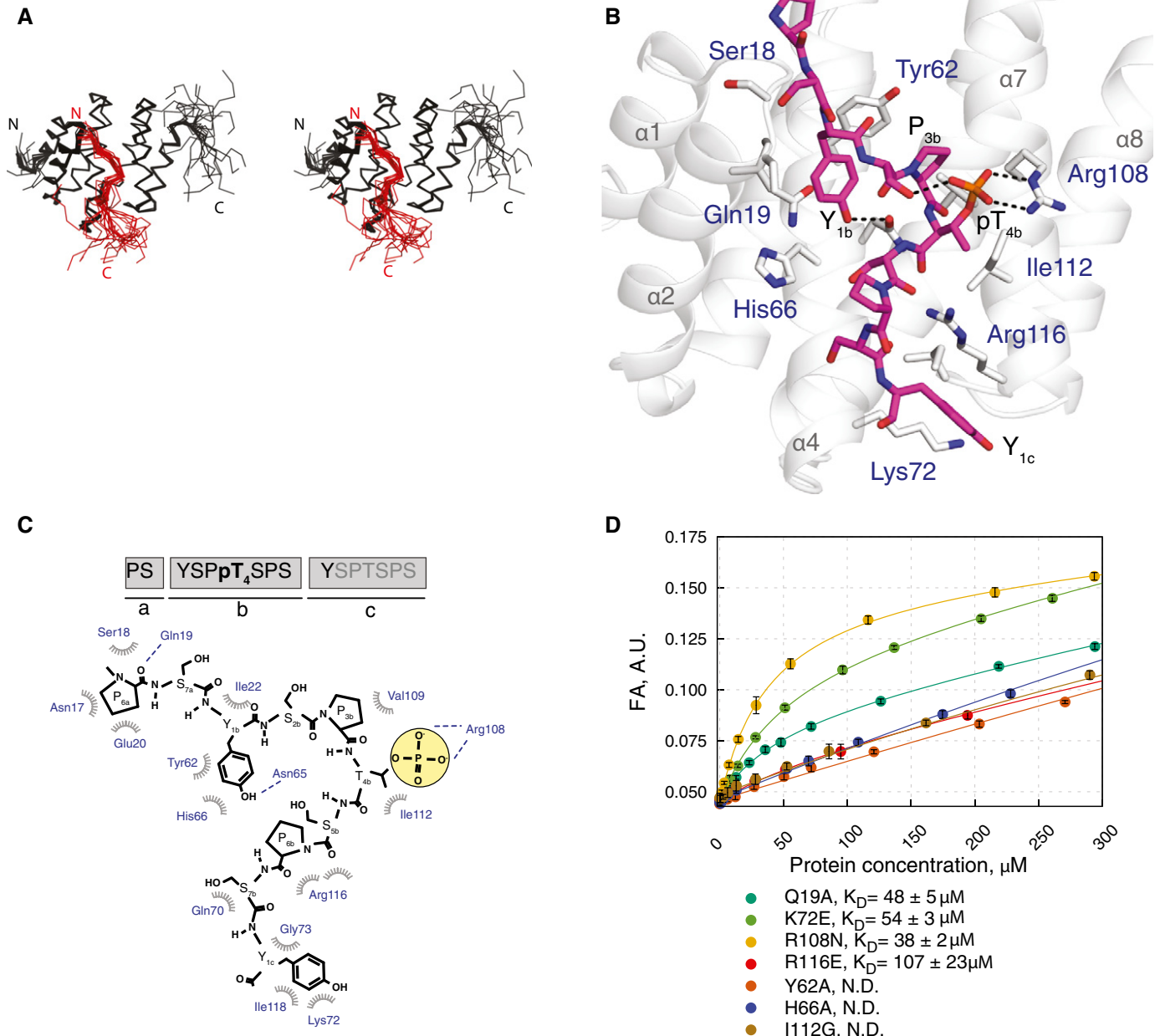


**Figure 2. Solution structure of Rtt103p CID in complex with pThr4-CTD.**

A   Overlay of the 20 lowest energy structures of Rtt103p CID (black ribbon) complexed with pThr4-CTD (red ribbon) shown in stereo. N- and C-termini of the protein and peptide are indicated.

B   Solution structure of Rtt103p CID (grey helices) bound to the pThr4-CTD peptide (magenta sticks). Highlighted Rtt103p CID residues (grey sticks, blue labels) form hydrophobic contacts and putative hydrogen bonds (dashed black lines) with pThr4-CTD peptide.

C   Schematic diagram of Rtt103p CID (blue) and pThr4-CTD (black) interactions (hydrophobic contacts, spoked arcs; hydrogen bonds, dashed lines).

D   Equilibrium binding of Rtt103p CID mutants with pThr4-CTD peptide monitored by fluorescence anisotropy (FA). Rtt103p CID mutants titrated into 10 nM FAM-labelled CTD peptides. Corresponding binding isotherms and $K_D$s (± standard deviation of the fit) are shown. FAM, 5,6-carboxyfluorescein. N.D., not determined.

YSPTSPS; Fig 2A–C; Table 1). We used this peptide with a single phosphorylation to avoid binding in multiple registers that would complicate NMR data analyses. The resulted structure of Rtt103p CID is formed by eight α-helices in a right-handed superhelical

**Table 1. NMR and refinement statistics for the Rtt103p CID-pThr4 CTD complex.**

| | Rtt103p CID–pThr4 CTD complex |
|---|---|
| **NMR distance & dihedral constraints** | |
| Distance restraints | |
| Total NOEs | 3,639 |
| Intra-residue | 843 |
| Inter-residue | 2,796 |
| Short | 1,691 |
| Medium | 1,104 |
| Long | 844 |
| Hydrogen bonds | 99 |
| Intermolecular distance restraints | 47 |
| Total dihedral angle restraints[a] | 198 |
| **Structure statistics[b]** | |
| Violations (mean and s.d.) | |
| Number of distance restraint violations > 0.5 Å | $0.10 \pm 0.31$ |
| Number of dihedral angle restraint violations > 15° | $13.8 \pm 2.33$ |
| Maximum dihedral angle restraint violation (°) | $39.57 \pm 9.55$ |
| Maximum distance constraint violation (Å) | $0.30 \pm 0.12$ |
| Deviations from idealized geometry[b] | |
| Bond lengths (Å) | $0.00355 \pm 0.00008$ |
| Bond angles (°) | $1.702 \pm 0.012$ |
| Average pairwise r.m.s.d. (Å)[b] | |
| Rtt103p CID (7–12; 19–31; 36–48; 54–73; 77–94; 100–116; 121–133) | |
| Heavy atoms | $0.91 \pm 0.13$ |
| Backbone atoms | $0.25 \pm 0.05$ |
| CTD (143–152) | |
| Heavy atoms | $2.06 \pm 0.42$ |
| Backbone atoms | $1.42 \pm 0.34$ |
| Complex | |
| All complex heavy atoms | $1.12 \pm 0.16$ |
| All complex backbone atoms | $0.61 \pm 0.14$ |
| Ramachandran plot statistics[c] | |
| Residues in most favoured regions (%) | 88.5 |
| Residues in additionally allowed regions (%) | 10.1 |
| Residues in generously allowed regions (%) | 0.9 |
| Residues in disallowed regions (%) | 0.5 |

[a]α-helical dihedral angle restraints imposed for the backbone based on the CSI.
[b]Calculated for an ensemble of the 20 lowest energy structures.
[c]Based on PROCHECK analysis [42].

arrangement (Fig 2A and B), out of which helices α2, α4 and α7 contact the pThr4-CTD peptide at residues $P_{6a}$, $S_{7a}$, $Y_{1b}$, $P_{3b}$, $pT_{4b}$, $S_{7b}$ and $Y_{1c}$ (Figs 2B and C, and EV1). This minimal CTD-binding moiety binds Rtt103p CID with a $K_D$ of $18 \pm 1$ μM (assayed by FA), which is almost identical as pThr4-CTD used for structural determination. The structure is similar to the one of Rtt103p CID–pSer2-CTD complex [20] in terms of the overall CID fold and the conformation of the N-terminal part of the CTD peptide, but entirely different for the C-terminal part of the CTD peptide (Figs 3, EV1, and EV2).

## Recognition of the phospho-threonine CTD by Rtt103p

The upstream part of the pThr4-CTD peptide adopts a β-turn conformation at $S_{2b}P_{3b}pT_{4b}S_{5b}$ and docks into a hydrophobic pocket of the Rtt103p CID that is formed by Ile22, Tyr62, His66, Val109 and Ile112, using $Y_{1b}$ and $P_{3b}$ residues (Fig 2B and C). The peptide conformation in the hydrophobic pocket is further stabilized by a hydrogen bond between hydroxyl of $Y_{1b}$ and the side-chain amide of Asn65. This hydrophobic pocket of Rtt103p is highly conserved, and mutations of residues Tyr62 and His66 (not affecting the structural integrity; Fig EV3) completely abolish the binding with pThr4-CTD (Figs 2D and EV4). $P_{3b}$ is inserted into the hydrophobic pocket next to Val109 and has a *trans* conformation of the $S_{2b}P_{3b}$ peptidyl-prolyl bond. As a result of this arrangement, both the $S_{2b}$ and $pT_{4b}$ side chains are positioned closely to each other in the solvent exposed area and form intramolecular hydrogen bond between the hydroxyl group of $S_{2b}$ and phospho-group of $T_{4b}$. The phospho-group of $pT_{4b}$ forms a hydrogen bond with the guanidinium group of Arg108. This is a critical interaction with the pThr4 mark, as confirmed by the affinity data for the Arg108Asn mutant (Fig 2D). Akin to Rtt103p, also other CID-containing proteins such as SCAF4/8 [22], RPRD1A/1B/2 [21] and CHERP [21] contain the equivalent arginine in the CID pocket. It will be interesting to see whether these human proteins really recognize pThr4-CTD as well and whether the pThr4 mark is relevant to their functions. Other CID-containing proteins in yeast, such as Nrd1p and Pcf11p, do not contain the equivalent arginine and these proteins were absent in the pThr4-CTD interactome [18].

Remarkably, we observed multiple strong intermolecular NOEs among the aromatics of $Y_{1c}$ in the downstream region of the CTD peptide and the C-terminal parts of helices α4 and α7 (Figs 2C, EV1, and EV5). The interaction of $Y_{1c}$ at the tip of helices α4 and α7 creates a second turn in the peptide at residues $pT_{4b}S_{5b}P_{6b}S_{7b}$, bringing two backbone carbonyl groups in close proximity and allows for their interaction with the guanidinium group of Arg116 (Figs 2B and C, and EV1). The side chain of $Y_{1c}$ forms numerous hydrophobic contacts with Lys72, Gly73, Ile118. Arg116Glu and Lys72Glu charge swapping mutants cause affinity drop of $K_D = 107 \pm 23$ μM and $K_D = 54 \pm 3$ μM, respectively (Fig 2D). The similar arrangement of the downstream region of the CTD was observed in the crystal structure of close human homologue of Rtt103p CID, RPRD1A, where the arginine forms a hydrogen bond with carbonyl of $T_{4b}$ and $P_{6b}$ [21]. The Arg116 position is conserved in RPRD1A/1B (Arg114) and RPRD2 (Arg130) (Fig EV4) [21]. Interestingly, the coordination of tyrosine from the third heptad repeat $Y_{1c}$ was not observed previously in the structure of Rtt103p CID bound to the CTD with Ser2 phosphorylation [20]. The previous study used a CTD peptide

lacking the complete binding moiety (PS YSPTSPS Y) that possibly precluded the accommodation of the downstream part of CTD peptide including the second tyrosine ($Y_{1c}$; Fig EV1). The comparison of chemical shift perturbations of Rtt103p upon binding to the singly phosphorylated pSer2-CTD and pThr4-CTD peptides with complete binding moiety suggests similar accommodation of downstream region of both peptides (Fig EV2).

### *Cis-trans* equilibrium of the Ser–Pro prolyl-peptidyl bond

We also tested as to whether two proximal phosphorylation marks (pSer2/pThr4) on the CTD peptide can alter the *cis-trans* equilibrium of the neighbouring prolyl-peptidyl bond (Fig EV6A). It has been shown that the *cis-trans* equilibrium of the CTD is critical for its recognition by cognate proteins [23–25] and the *trans* conformation of the Ser–Pro prolyl-peptidyl bond is required for the β-turn formation [20,23,26,27]. To exclude the possibility that a highly populated *cis* conformer would attenuate the binding of the CTD peptide with two phospho-marks, we assayed the conformational population of mono- and diphosphorylated peptides using the $[^{1}H,^{13}C]$-HSQC spectra of PS Y(pS)$^{13C}$P(pT)SPS YS and PS YS$^{13C}$P(pT)SPS YS peptides, where all $P_{3b}$ carbons were $^{13}C$-isotopically labelled (Fig EV6B). In case of pThr4-CTD, we observed 6.6% of the *cis* conformer. We obtained virtually identical number for the pSer2pThr4-CTD peptide, where the *cis* conformation was populated at 7.8%. Our data suggest that the double phosphorylation at pSer2/pThr4 of the CTD does not influence the ratio of *cis-trans* conformers. Next, we titrated the PS YS$^{13C}$P(pT)SPS YS peptide with Rtt103p CID and monitored the titration by $[^{1}H,^{13}C]$-HSQC experiment (Fig EV6C). The spectra show the disappearance of peaks that correspond to the *cis* conformation during titration, indicating a shift in the *cis-trans* equilibrium towards the *trans* conformation of the $S_{2b}$–$P_{3b}$ prolyl-peptidyl bond that is required for the β-turn formation. The peaks corresponding to the *trans* conformation of $P_{3b}$ moved upon titration with protein, reflecting the accommodation of the proline in the hydrophobic pocket of Rtt103p CID.

### CTD code degeneration

The complex of Rtt103p CID–pThr4-CTD reported here represents the first structure capturing the recognition of the CTD phosphorylated at threonine and explains the structural basis of why Rtt103p can be a part of the pSer2- and pThr4-CTD interactomes. Previous reports suggested that Thr4 phosphorylations could interfere with CTD binding by destabilizing the β-turn conformation that is required for CTD binding [17,26]. However, our structure shows that the pThr4 mark is directly recognized by Rtt103p and also that the phosphate group of pThr4 forms intramolecular hydrogen bond stabilizing the bound CTD conformation. This conformation involves the β-turn at $S_{2b}P_{3b}pT_{4b}S_{5b}$ that is a prerequisite for an effective docking into the hydrophobic pocket of Rtt103p CID (Figs 2B and C, and 3). Interestingly, the intramolecular hydrogen bond that stabilizes the β-turn mirrors the one of the Ser2 phosphorylated CTD bound to Rtt103p (Fig 3B). The Rtt103p Arg108Asn mutant has also a similar drop in affinity for pThr4-CTD and pSer2-CTD, $K_D = 38 \pm 2$ μM and $K_D = 44 \pm 2$ μM, respectively. These observations suggest that CTD modifications preventing intramolecular stabilization of the β-turn should negatively affect CTD binding. Indeed, we observed that doubly phosphorylated CTD at Ser2 and Thr4 binds to Rtt103p as weak as unmodified CTD (Fig 1). Electrostatic repulsion between closely arranged phosphates of pSer2 and pThr4 interfere with the formation of the bound CTD conformation and the peptide with pSer2/pThr4 marks cannot be accommodated in the binding pocket of Rtt103p (Fig 3). In support of this, the coexistence of the pSer2/pThr4 marks in the same repeat has not been detected by recent mass-spectrometry analysis of RNAPII CTD population pulled down by Rtt103p [11]. Our structure also explains lethality of the Thr4Glu CTD mutant in yeast [18]. Permanent substitution for glutamate mimics Thr4 phosphorylation that interferes with Ser2 phosphorylation, which consequently prevents binding of the CTD to cognate proteins as described above.

The individual letters of the CTD code have so far been associated with unique information translated to stimulation or inhibition of recruitment of CTD readers. Comparison of Rtt103p
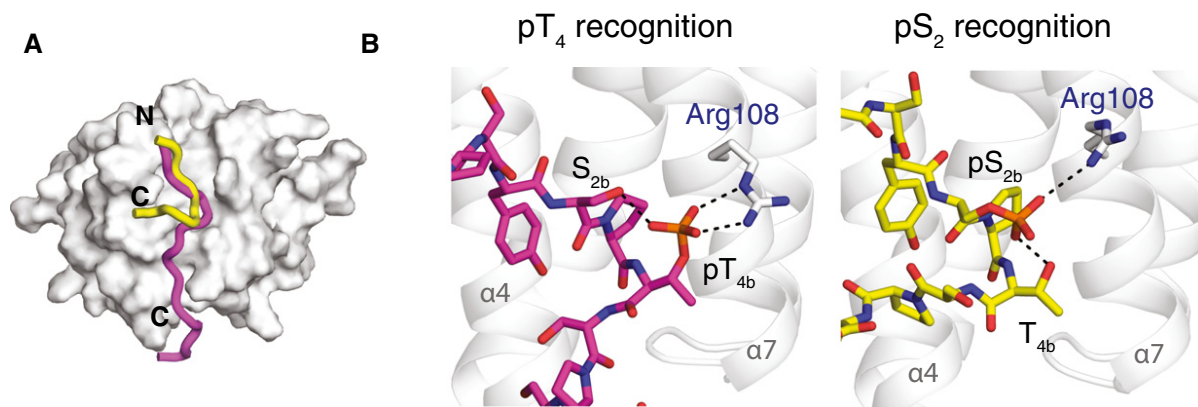


**Figure 3.  Degeneracy of the CTD code.**

A   Superposition of pThr4-CTD (magenta; PDB ID: 5LVF) and pSer2-CTD (yellow; PDB ID: 2L0I) peptides on the Rtt103p CID surface (grey). N- and C-termini of the peptides are indicated.

B   Close view on the phospho-recognition site of Rtt103p CID. Interaction of pThr4- (magenta sticks, left) and pSer2-CTD (yellow sticks, right) peptides with Arg108 (grey). Hydrogen bonds of the phospho-groups are indicated with black dashed lines.

CID–pThr4-CTD structure with the Rtt103p CID–pSer2-CTD complex shows fascinating feature that the same interaction pocket of Rtt103p can read two different phosphorylation patterns of the CTD (pSer2 and pThr4) using the same mechanism and involves the same residues (mainly Arg108; Fig 3). Based on our structural findings, we suggest that the CTD code can be degenerated when read by CID-containing proteins. In other words, the recruitment of a single CTD-binding factor may be coded by more than one letter of the CTD code. As a consequence of this redundancy, CID-containing CTD-binding factors can be recruited to the poorly conserved heptad repeats of the CTD (e.g. the CTD of fruit fly) or they can tolerate some errors or imperfections in phosphorylation of the CTD [1,3,28].

## Materials and Methods

### Cloning and protein purification

pET28b-Rtt103p CID was a gift from B. Lunde [20]. Rtt103p CID point mutants were obtained by QuikChange site-directed mutagenesis kit (Stratagene). Resulting constructs were verified by DNA sequencing and then transformed into *E. coli* BL21-Codon Plus (DE3)-RIPL cells (Stratagene). Rtt103p CID (3-131–6xHIS) was expressed and purified as previously described [20].

### NMR measurements and structure determination

All NMR spectra for the backbone and side-chain assignments were recorded on Bruker AVANCE III HD 950, 850 and 700 MHz spectrometers equipped with cryoprobes at a sample temperature of 20°C using 1 mM uniformly $^{15}$N,$^{13}$C-labelled Rtt103p CID in 35 mM KH$_2$PO$_4$, 100 mM KCl, pH 6.8 (20°C) (90% H$_2$O/10% D$_2$O). Initial backbone resonance frequency assignment was transferred from BMRB entries 17044 and 16411 and confirmed by HNCA spectrum. The spectra were processed using TOPSPIN 3.2 (Bruker Biospin), and the protein resonances were assigned manually using Sparky software (Goddard T.G. and Kneller D.G., University of California, San Francisco). For the assignment of the side-chain proton and carbon resonances, 4D version of HCCH TOCSY [29] was measured with a non-uniform sampling. Acquired data were processed and analysed analogously as described previously [30,31].

All distance constraints were derived from the three-dimensional $^{15}$N- and $^{13}$C-edited NOESYs collected on a 950 MHz spectrometer. Additionally, intermolecular distance constraints were obtained from the three-dimensional F$_1$-$^{13}$C/$^{15}$N-filtered NOESY-[$^{13}$C,$^{1}$H]-HSQC experiment [32,33], with a mixing time of 150 ms on a 950 MHz spectrometer. The NOEs were semi-quantitatively classified based on their intensities in the 3D NOESY spectra. The initial structure determinations of the Rtt103p-CTD complex were performed with the automated NOE assignment module implemented in the CYANA 3.97 program [34]. Then, the CYANA-generated restraints along with manually assigned protein-CTD intermolecular restraints were used for further refinement of the preliminary structures with AMBER16 software [35]. These calculations employed a modified version (AMBER ff14SB) of the force field [36], using a protocol described previously [37,38]. The 20 lowest energy conformers were selected (out of 50 calculated) to form the final ensemble of structures. The atomic coordinates for the NMR ensemble of the Rtt103p CID-pThr4-CTD complex have been deposited in the Protein Data Bank under ID code 5LVF and in Biological Magnetic Resonance Bank under ID code 34041. Molecular graphics were generated using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

### Fluorescence anisotropy

The equilibrium binding of Rtt103p CID constructs to differently phosphorylated CTD was analysed by fluorescence anisotropy. The CTD peptides were N-terminally labelled with the 5,6-carboxyfluorescein (FAM). The measurements were conducted on a FluoroLog-3 spectrofluorometer (Horiba Jobin-Yvon Edison, NJ). The instrument was equipped with a thermostatted cell holder with a Neslab RTE7 water bath (Thermo Scientific). Samples were excited with vertically polarized light at 467 nm, and both vertical and horizontal emissions were recorded at 516 nm. All measurements were conducted at 10°C in 35 mM KH$_2$PO$_4$, 100 mM KCl (pH 6.8). Each data point is an average of three measurements. The experimental binding isotherms were analysed by DynaFit using 1:1 model with non-specific binding [39].

### *Cis-trans* population estimation

For the estimation of the *cis-trans* population of conformers around the Ser-Pro peptide bond, aliphatic [$^{13}$C,$^{1}$H]-HSQC was collected using 1 mM sample of peptide PSYS$^{13}$CP(pT)SPSYS or PSY(pS)$^{13}$CP(pT)SPSYS in 35 mM KH$_2$PO$_4$, 100 mM KCl (pH 6.8) in 90% H$_2$O/10% D$_2$O at 20°C on Bruker AVANCE III HD 700 MHz spectrometer. For the [$^{13}$C,$^{1}$H]-HSQC titration experiment, 0.2 mM PSYS$^{13}$CP(pT)SPSYS peptide was used and 1.2 mM Rtt103p CID stock was added. $^{1}$H-$^{13}$C$_\gamma$ and $^{1}$H-$^{13}$C$_\beta$ peaks were integrated using Sparky routine (Goddard T.G. and Kneller D.G., University of California, San Francisco). Population was estimated as a ratio of the peak volume of a given conformation to the sum of volumes of all conformations.

### Determination of chemical shift perturbation (CSP) value

Chemical shift perturbation (CSP) value is defined as the normalized length of a vector $E_j$, whose components are differences $\delta_{ji}$ between observed chemical shifts (bound form) and chemical shifts from a reference experiment (free form). Index $j$ represents the amino acid type within the primary sequence of the protein. Weight factors for each atom type $w_H = 1$ and $w_N = 0.15$ were used.

$$|E_j| = \sqrt{\sum_{i=H,N} w_i \delta_{ji}^2}$$

### Peptides used in the study

The following peptides were synthesized by JPT (Berlin, DE) and Clonestar (Brno, CZ): FAM-PSY(pS)PTSPSYSPTSPS; FAM-PSYSP(pT)SPSYSPTSPS; FAM-PSY(pS)P(pT)SPSYSPTSPS; FAM-PSYSPTSPSYSPTSPS; FAM-PSYSP(pT)SPSYSP(pT)SPS; FAM-PSYSP(pT)SPSYS; FAM-PS(pY)SPTSPSYSPTSPS; PSYSP(pT)SPSYSPTSPS; PSYS$^{13}$CP(pT)SPSYS; PS Y(pS)$^{13}$CP(pT)SPSYS.

## Author contributions

OJ designed the experiments, prepared protein and peptide samples, measured and analysed FA, assigned spectra, calculated and refined structure, and wrote the manuscript; MK collected and processed 4D HCCH TOCSY spectra, and assisted with structure refinement; KK collected and processed NMR spectra, and assisted with structure calculation and refinement; RS designed the experiments, assisted with structure calculation and refinement, and wrote the manuscript.

## Conflict of interest

The authors declare that they have no conflict of interest.

# References

1.  Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. *Genes Dev* 19: 1401–1415

2.  Jasnovidova O, Stefl R (2013) The CTD code of RNA polymerase II: a structural view. *Wiley Interdiscip Rev RNA* 4: 1–16

3.  Eick D, Geyer M (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* 113: 8456–8490

4.  Zaborowska J, Egloff S, Murphy S (2016) The pol II CTD: new twists in the tail. *Nat Struct Mol Biol* 23: 771–777

5.  Jeronimo C, Collin P, Robert F (2016) The RNA polymerase II CTD: the increasing complexity of a low-complexity protein domain. *J Mol Biol* 428: 2607–2622

6.  Chapman RD, Heidemann M, Hintermair C, Eick D (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet* 24: 289–296

7.  Jeronimo C, Bataille AR, Robert F (2013) The writers, readers, and functions of the RNA polymerase ii c-terminal domain code. *Chem Rev* 113: 8491–8522

8.  Sakurai H, Ishihama A (2002) Level of the RNA polymerase II in the fission yeast stays constant but phosphorylation of its carboxyl terminal domain varies depending on the phase and rate of cell growth. *Genes Cells* 7: 273–284

9.  Rosonina E, Yurko N, Li W, Hoque M, Tian B, Manley JL (2014) Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. *Proc Natl Acad Sci* 111: 11924–11931

10. Schüller R, Forné I, Straub T, Schreieck A, Texier Y, Shah N, Decker T-M, Cramer P, Imhof A, Eick D (2016) Heptad-specific phosphorylation of RNA polymerase II CTD. *Mol Cell* 61: 305–314

11. Suh H, Ficarro SB, Kang U-B, Chun Y, Marto JA, Buratowski S (2016) Direct analysis of phosphorylation sites on the Rpb1 C-terminal domain of RNA polymerase II. *Mol Cell* 61: 297–304

12. Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R, Ferrier P, Flatley A, Kremmer E *et al* (2012) Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J* 31: 2784–2797

13. Hsin J-P, Sheth A, Manley JL (2011) RNAP II CTD phosphorylated on Threonine-4 is required for histone mRNA 3′ end processing. *Science* 334: 683–686

14. Hintermair C, Voß K, Forné I, Heidemann M, Flatley A, Kremmer E, Imhof A, Eick D (2016) Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression. *Sci Rep* 6: 27401

15. Stiller JW, McConaughy BL, Hall BD (2000) Evolutionary complementation for polymerase II CTD function. *Yeast* 16: 57–64

16. Schwer B, Shuman S (2011) Deciphering the RNA polymerase II CTD code in fission yeast. *Mol Cell* 43: 311–318

17. Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336: 1723–1725

18. Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS (2016) Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep* 15: 2147–2158

19. Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432: 517–522

20. Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 17: 1195–1201

21. Ni Z, Xu C, Guo X, Hunter GO, Kuznetsova OV, Tempel W, Marcon E, Zhong G, Guo H, Kuo W-HW *et al* (2014) RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nat Struct Mol Biol* 21: 686–695

22. Becker R, Loll B, Meinhart A (2008) Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* 283: 22659–22669

23. Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, Hofr C, Vanacova S, Stefl R (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* 26: 1891–1896

24. Xiang K, Nagaike T, Xiang S, Kilic T, Beh MM, Manley JL, Tong L (2010) Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* 467: 729–733

25. Werner-Allen JW, Lee C-J, Liu P, Nicely NI, Wang S, Greenleaf AL, Zhou P (2011) cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. *J Biol Chem* 286: 5717–5726

26.  Meinhart A, Cramer P (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 430: 223−226

27.  Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A (2005) Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* 12: 144−151

28.  Buratowski S (2003) The CTD code. *Nat Struct Mol Biol* 10: 679−680

29.  Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD (1993) A gradient-enhanced HCCH TOCSY experiment for recording side-chain H-1 and C-13 correlations in H2O samples of proteins. *J Magn Reson Ser* 101: 333−337

30.  Nováček J, Haba NY, Chill JH, Žídek L, Sklenář V (2012) 4D Non-uniformly sampled HCBCACON and 1J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *J Biomol NMR* 53: 139−148

31.  Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Žídek L, Šanderová H, Krásný L, Koźmiński W, Sklenář V (2011) 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR* 50: 1−11

32.  Peterson RD, Theimer CA, Wu H, Feigon J (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J Biomol NMR* 28: 59−67

33.  Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-Peptide/boxB RNA complex. *J Am Chem Soc* 119: 6711−6721

34.  Güntert P, Buchner L (2015) Combined automated NOE assignment and structure calculation with CYANA. *J Biomol NMR* 62: 453−471

35.  Case DA, Betz RM, Botello-Smith W, Cerutti DS, Cheatham TE, Darden TA, Duke RE, Giese TJ (2016) *AMBER 2016*. San Francisco: University of California

36.  Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11: 3696−3713

37.  Stefl R, Oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L, Maris C, Peng L, Hofr C, Emeson RB *et al* (2010) The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* 143: 225−237

38.  Hobor F, Pergoli R, Kubicek K, Hrossova D, Bacikova V, Zimmermann M, Pasulka J, Hofr C, Vanacova S, Stefl R (2011) Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J Biol Chem* 286: 3645−3657

39.  Kuzmic P (2009) DynaFit—a software package for enzymology. *Methods Enzymol* 467: 247−280

40.  Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605−1612

41.  Schubert M, Labudde D, Oschkinat H, Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. *J Biomol NMR* 24: 149−154

42.  Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8: 477−486

## 5.8   Structure and dynamics of the RNAPII CTDsome with Rtt103

In this work, a hybrid/integrative approach was used to visualize the architecture of the full-length CTD in complex with the transcription termination factor Rtt103. Specifically, first solved the structures of the isolated subcomplexes at high resolution and then these structures were arranged into the overall envelopes determined at low resolution by small-angle X-ray scattering (SAXS). The reconstructed overall architecture of the Rtt103–CTD complex reveals how Rtt103 decorates the CTD platform.

# Structure and dynamics of the RNAPII CTDsome with Rtt103

**Olga Jasnovidova[a], Tomas Klumpler[a], Karel Kubicek[a], Sergei Kalynych[a], Pavel Plevka[a], and Richard Stefl[a,1]**

[a]CEITEC–Central European Institute of Technology, Masaryk University, CZ-62500 Brno, Czech Republic

**RNA polymerase II contains a long C-terminal domain (CTD) that regulates interactions at the site of transcription. The CTD architecture remains poorly understood due to its low sequence complexity, dynamic phosphorylation patterns, and structural variability. We used integrative structural biology to visualize the architecture of the CTD in complex with Rtt103, a 3′-end RNA-processing and transcription termination factor. Rtt103 forms homodimers via its long coiled-coil domain and associates densely on the repetitive sequence of the phosphorylated CTD via its N-terminal CTD-interacting domain. The CTD–Rtt103 association opens the compact random coil structure of the CTD, leading to a beads-on-a-string topology in which the long rod-shaped Rtt103 dimers define the topological and mobility restraints of the entire assembly. These findings underpin the importance of the structural plasticity of the CTD, which is templated by a particular set of CTD-binding proteins.**

RNA polymerase II | CTD | structural biology | transcription | Rtt103

The C-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNAPII) consists of multiple tandem repeats (26 in yeast, 52 in humans) of the heptapeptide consensus Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, which is highly conserved from yeast to human (1–3). The CTD serves as a binding platform for many RNA/protein-binding factors involved in the regulation of the transcription cycle (1, 3). Yeast are inviable if the CTD is trimmed to less than 11 repeats of the heptapeptide consensus (4) or if the periodicity of two repeats is perturbed (5), suggesting the importance of both the CTD length and its repetitiveness.

The CTD interaction network is regulated by posttranslational modifications of the CTD, which yield specific phosphorylation and subsequent factor-binding patterns in coordination with the transcription cycle (the "CTD code") (1, 6–11). Phosphorylations at $Y_1$, $S_2$, $T_4$, $S_5$, and $S_7$ are the most common and well-studied posttranslational modifications of the CTD (12). Mass spectrometry studies of the CTD showed that the CTD heptads are homogeneously phosphorylated along the entire length of the domain in proliferating yeast and human cells (13, 14). Major phosphorylation sites are $S_2$ and $S_5$, whereas $Y_1$, $T_4$, and $S_7$ are minor phosphorylation sites (13, 14), but all sites are important for transcription regulation and proper functioning of the cell. On average, each CTD heptad is phosphorylated once and the occurrence of two phosphorylations per repeat is a rare event (13, 14). The coimmunoprecipitation of specific CTD phosphoisoforms revealed distinct functional sets of factors (CTD-interactome) related to each CTD phosphoisoform (15).

The CTD has no well-defined 3D structure and, therefore, is not observed in the crystal structures of RNAPII (16–19) and forms fuzzy densities on electron microscopy images (20, 21). Nevertheless, the first structural information of the unbound CTD has recently been reported in the fruit fly (22, 23), where it was shown that the CTD forms a compact random coil and that its phosphorylation induces a modest extension and stiffening of the CTD (22, 23).

Current structural knowledge of interactions between the CTD and its recognition factors is based on short peptides mimicking the CTD bound to given CTD binding factors (1, 19, 24). However, the atomic-level structural architecture of the full-length CTD modulated by associated factors remains unknown. Several studies attempted to propose a structural model for the full-length CTD. For example, in the complex of the CTD peptide with the CTD-interacting domain (CID) of Pcf11, a subunit of cleavage factor IA (25), the CTD heptad was found to adopt a β-turn conformation (26). Therefore, a compact, left-handed, β-spiral model of the CTD was proposed (26). A β-spiral conformation would allow the CTD chain with a length of 100 Å to fold into a compact structure, which corresponds to the observed densities in low-resolution electron microscopy images of RNAPII (20). The heterodimer composed of the human proteins RPRD1A and RPRD1B was found to bind the CTD, thereby stimulating the recruitment and phosphatase activity of RPAP2 ($pS_5$-CTD-phosphatase) (27). These findings led to the proposal of a model in which the CTD and accessory molecules form a high-order arrangement dubbed the "CTDsome" (27).

To probe the CTDsome architecture experimentally, we set out to apply integrative structural biology methods and investigate how the termination factor Rtt103 decorates the sequence of the CTD. First, we independently solved high-resolution structures of stable subunits by solution NMR spectroscopy (NMR) and X-ray crystallography. Then, we corroborated the obtained structural information with small-angle X-ray scattering (SAXS) data to reconstruct the overall architecture of the Rtt103–CTD complex. We show that Rtt103 contains a coiled-coil domain that mediates Rtt103 dimerization and uses its

## Significance

RNA polymerase II (RNAPII) not only transcribes protein coding genes and many noncoding RNA, but also coordinates transcription and RNA processing. This coordination is mediated by a long C-terminal domain (CTD) of the largest RNAPII subunit, which serves as a binding platform for many RNA/protein-binding factors involved in transcription regulation. In this work, we used a hybrid approach to visualize the architecture of the full-length CTD in complex with the transcription termination factor Rtt103. Specifically, we first solved the structures of the isolated subcomplexes at high resolution and then arranged them into the overall envelopes determined at low resolution by small-angle X-ray scattering. The reconstructed overall architecture of the Rtt103–CTD complex reveals how Rtt103 decorates the CTD platform.

N-terminal CID to read adjacent repetitive phosphorylation marks on the CTD independently of one other. Our reconstruction demonstrates how Rtt103 explores the repetitiveness and length of the CTD sequence while keeping the entire arrangement partially flexible.

## Results

**Limited Proteolysis of Rtt103 Reveals a Coiled-Coil Domain That Mediates Dimerization.** In our divide-and-conquer approach, we first identified the overall domain organization of Rtt103. Trypsin digestion of the full-length Rtt103 coupled with mass spectrometry revealed that the protein fragment harboring amino acid residues 1–246 (Rtt103$_{1–246}$) is protected from proteolytic cleavage (Fig. 1$A$ and Fig. S1). The remaining C-terminal part of Rtt103 (amino acid residues 247–409) was efficiently digested by trypsin, suggesting the absence of additional structured domains (Fig. 1$A$ and Fig. S1). Subsequent biochemical characterization of the identified stable constructs revealed that Rtt103$_{141–246}$ and Rtt103$_{1–246}$ form homodimers (Fig. S2 $A$ and $B$). Subsequent crystallization screens of the Rtt103$_{141–246}$ and Rtt103$_{1–246}$ constructs showed that only the Rtt103$_{141–246}$ construct formed well-diffracting crystals. The structure of Rtt103$_{141–246}$ was determined to a resolution of 2.6 Å
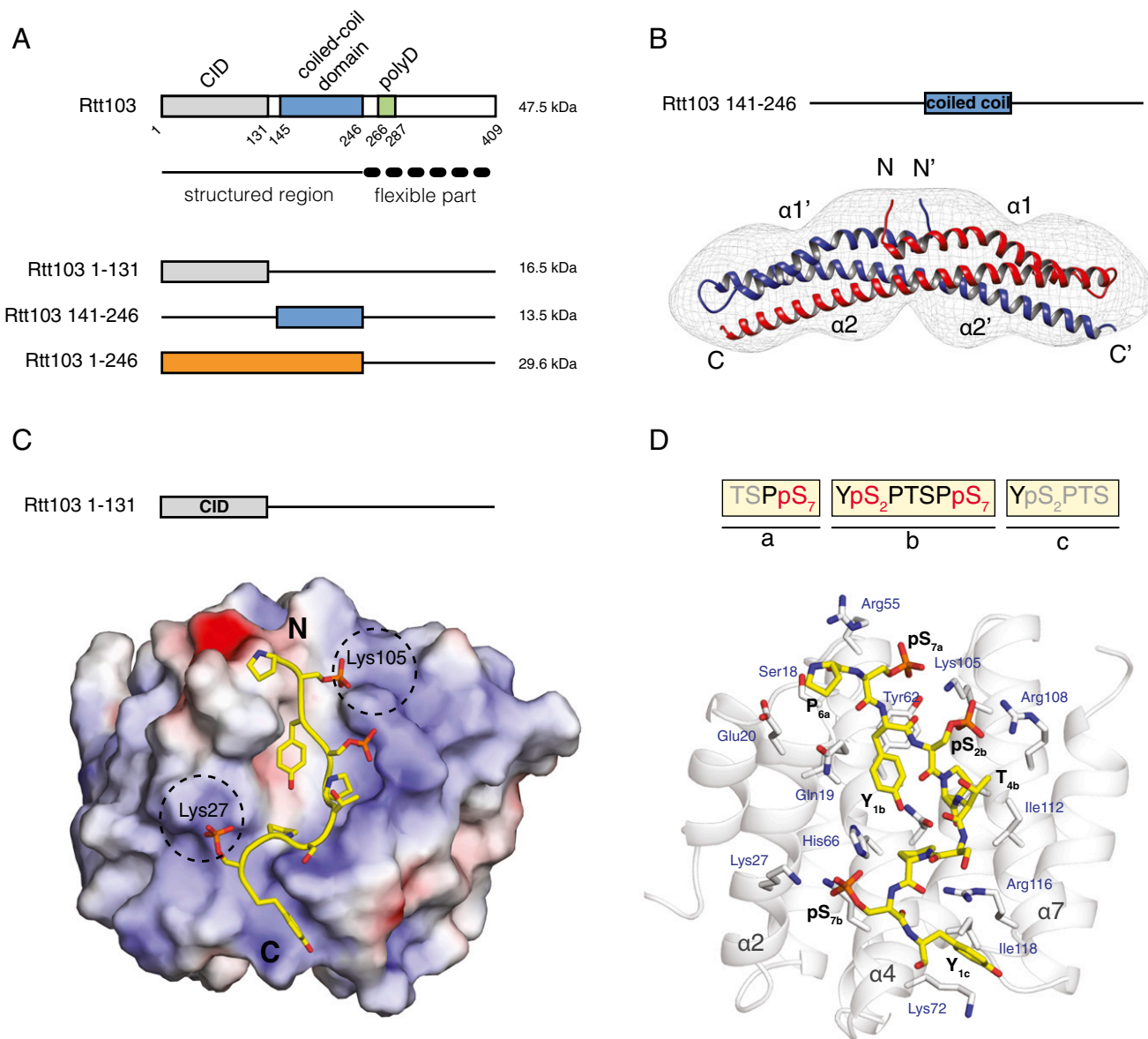


**Fig. 1.** Dimerization and RNAPII CTD recognition by Rtt103. (*A*) Scheme of Rtt103 domain organization (Upper). The numbers below the scheme represent borders of the amino acid segments. Structured and flexible regions were determined based on the limited proteolysis study (Fig. S1). The recombinant protein constructs used in the study, along with their respective molecular masses, are shown (*Lower*). CID, CTD-interacting domain; polyD, polyaspartate stretch. (*B*) Crystal structure of the Rtt103$_{141–246}$ coiled-coil domain shown superimposed with an ab initio model (gray mesh) derived using DAMMIN (40) from SAXS scattering data. The two different polypeptide chains of the coiled-coil dimer are shown in red and blue; their respective N- and C-termini, as well as α-helices, are indicated. (*C*) Electrostatic surface representation of the Rtt103 CID (electropositive in blue, electronegative in red, neutral in white) in complex with the pS$_2$pS$_7$-CTD peptide (yellow sticks; PDB ID code: 5M9D). The N- and C-termini of the peptide are indicated. Dashed black circles indicate electropositive areas that accommodate pS$_7$ residues. (*D*) Detailed view of the Rtt103 CID (gray cartoons) bound to the pS$_2$pS$_7$-CTD peptide (yellow sticks). Highlighted Rtt103 CID residues (gray sticks, blue labels) form hydrophobic contacts and putative hydrogen bonds with the pS$_2$pS$_7$-CTD peptide (yellow sticks, black labels). The sequence of the peptide used for structure determination is indicated above; residues that showed interaction with the Rtt103 CID are shown in black and red.

(Tables S1 and S2). We found that each Rtt103$_{141-246}$ subunit consists of two α-helices, namely the α1-helix (Gln146-Glu177) with a small bend in the middle and a long α2-helix of ~105 Å in length (Val184-Asp246) (Fig. 1B). In the crystal, two protein chains form a dimer where the α2-helices are arranged in an antiparallel fashion (Fig. 1B). This architecture of the dimer is in agreement with findings from gel filtration experiments (Fig. S2A) and SAXS data (Fig. S2B). Importantly, the central region of the α2-helix (Lys200-Ile238) contains a coiled-coil signature, which is arranged in trans in the antiparallel dimer assembly of the two Rtt103 molecules. The coiled-coil domain contains a characteristic knobs-into-holes packing with mixed "a" and "d" layers (Fig. S2C), with an average pitch of 172 Å (defined by CCCP; ref. 28). The dimer structure is also stabilized by multiple intermolecular (Asp149-Lys152, Asp153-Lys216, Lys168-Asp172, Asp223-Arg226) and intramolecular (Lys200-Glu239, Glu231-Arg210-Glu224) salt bridges. Altogether, the key findings regarding Rtt103 architecture are as follows: (i) Rtt103 contains a coiled-coil domain that follows the CID and (ii) the C-terminal half of the Rtt103 is disordered.

**The Rtt103 CID Binds the Extended pS$_2$-pS$_7$-CTD Peptide.** The structure of Rtt103$_{1-131}$ (or CTD-interacting domain; CID) bound to a short Ser2-phosphorylated CTD moiety has previously been reported (29). Here, we used NMR to determine the structure of Rtt103$_{1-131}$ bound to a longer CTD substrate (Ser2/7-phosphorylated; Tables S3 and S4), which revealed that the recognition interface of the CID is, in fact, larger than previously reported (29). Our NMR structure of Rtt103$_{1-131}$ bound to the extended TSPpS$_7$ YpS$_2$PTSPpS$_7$ YpS$_2$PTS peptide (termed pS$_2$pS$_7$-CTD) confirmed the previously reported observation regarding the recognition of the upstream region of pS$_2$pS$_7$-CTD, and further revealed information regarding the recognition of the downstream region of pS$_2$pS$_7$-CTD (Fig. 1 C and D). The structure of Rtt103 CID is formed by eight α-helices in a right-handed superhelical arrangement. The NMR data show that the pS$_2$pS$_7$-CTD peptide binds at the conserved surface formed by helices α2, α4, and α7 of the Rtt103 CID (Fig. 1D and Fig. S3). NMR revealed intermolecular contacts between Rtt103$_{1-131}$ and the residues P$_{6a}$, pS$_{7a}$, Y$_{1b}$, pS$_{2b}$, P$_{3b}$, T$_{4b}$, S$_{5b}$, and Y$_{1c}$ of the pS$_2$pS$_7$-CTD peptide. Specifically, P$_{6a}$ lies in the proximity of the hydrophobic area formed by the N-terminal tip of the α2-helix, being involved in multiple intermolecular contacts with Ser18, Gln19, and Glu20. Residue Y$_{1b}$ is also docked into a hydrophobic pocket (Ile22, Tyr62) and stabilized by a hydrogen bond between its hydroxyl group and the side-chain amide group of Asn65. Residue P$_{3b}$ forms hydrophobic interactions with Val109 and Ile112. Residues pS$_{2b}$P$_{3b}$T$_{4b}$S$_{5b}$ form a β-turn stabilized by hydrogen bonds between the pS$_{2b}$ carbonyl and the S$_{5b}$ amide, between the pS$_{2b}$ γ-oxygen and T$_{4b}$ amide, and between the pS$_{2b}$ phosphate and T$_{4b}$ hydroxyl. Perturbation of the above-described hydrophobic pocket (not affecting the structural integrity; Fig. S4C and refs. 27 and 30) caused a drop of 30- to 50-fold in the affinity between pS$_2$pS$_7$-CTD and Rtt103$_{1-131}$ ($K_D$ = 33 ± 1.2 μM for Ile112Ala, $K_D$ = 80 ± 11 μM for Ile112Gly) (Fig. S4). In agreement with previous structural observations (29), we noted that the phosphorylation of S$_{2b}$ is recognized by the side chain of Arg108. Interestingly, we observed multiple close contacts between Y$_{1c}$ and the C-terminal parts of helices α4 and α7. The positioning of Y$_{1c}$ near the tip of helices α4 and α7 induces a second sharp turn in the pS$_2$pS$_7$-CTD peptide. The side chain of Y$_{1c}$ forms a broad range of hydrophobic contacts with Lys72, Gly73, and Ile118, whereas the guanidinium group of Arg116 coordinates the backbone of pS$_2$pS$_7$-CTD. We found that charge-swapping mutations at the interacting sites of Rtt103 (not affecting the structural integrity; Fig. S4C) resulted in pronounced affinity decrease between Rtt103 and pS$_2$pS$_7$-CTD ($K_D$ = 9.7 ± 0.7, 51 ± 2.2, and 65 ± 8.9 μM for Lys72Glu, Arg116Glu, and Lys72Glu/Arg116Glu, respectively), highlighting the importance

of the CTD backbone interactions with Arg116. A large area of the Rtt103 CID surface is positively charged and enriched in residues that could stabilize interaction with negatively charged sites of the phosphorylated CTD peptide (Fig. 1C). Although our data did not indicate the presence of intermolecular contacts for the pS$_7$ residues, the positions of these residues are indirectly defined by the nuclear Overhauser effects from the neighboring residues. Therefore, residues pS$_7$ are likely involved in charge–charge interactions with Lys27 and Lys105 (which is part of the poly-Lys tract Lys103-Lys104-Lys105) (Fig. 1C). We found that the Lys27Glu mutant (perturbation of one of the pS$_7$ binding pockets; Fig. 1C) showed lower binding only for the pS$_7$ containing peptide ($K_D$ of 13.2 ± 0.3 μM and 28.5 ± 1 μM for wild type and Lys27Glu, respectively) but not for the pS$_2$-containing peptide ($K_D$ of 1.6 ± 0.07 μM and 2 ± 0.8 μM for wild type and Lys27Glu, respectively). Altogether, the key finding is that the Rtt103 CID interacts with pS$_2$pS$_7$-CTD via a larger area than previously reported (29), specifically recognizing the downstream region of the CTD peptide. Our structure reveals that P$_{6a}$pS$_{7a}$Y$_{1b}$pS$_{2b}$P$_{3b}$T$_{4b}$S$_{5b}$P$_{6b}$pS$_{7b}$Y$_{1c}$ is the minimal CTD-binding moiety recognized by Rtt103.

**Two CIDs Tethered by a Coiled-Coil Domain Tumble Independently.** As a result of the antiparallel arrangement of the coiled coils, the Rtt103 CIDs are attached by a linker of 15 amino acids to the middle region of the coiled-coil domain. NMR investigations of the Rtt103$_{1-246}$ and CID constructs showed that the CID structure is not influenced by the presence of the coiled-coil domain and that the CIDs are likely to tumble independently (Fig. S5). To visualize the arrangement of the Rtt103 CIDs relative to the coiled-coil domains, we analyzed SAXS scattering data of purified Rtt103$_{1-246}$ using available atomic structures (PDB ID codes: 2KM4, 5M48) by ensemble-optimization method (EOM 2.0) (31). This approach enables deconvolving the conformational averaging into the contribution of individual conformers. The obtained models suggest that the coiled-coil domain defines the length of the protein (~105 Å) (Fig. 2), and the linker allows the CIDs to reach all across the 105-Å-long coiled-coil domain. Thus, the CIDs could be positioned relatively close to each other but are also able to sample a large surrounding space to recognize the substrate (Movie S1). Next, we tested whether the coiled-coil–mediated dimerization of Rtt103 affects the binding to the CTD using fluorescence anisotropy (FA). We measured the binding affinity for the minimal CTD-binding moiety (SPS YpSPTSPpS YS) and a long CTD substrate (harboring two minimal CTD-binding moieties connected with a spacer; SPS YpSPTSPpS YSPTSPS YpSPTSPpS YS) with Rtt103$_{1-246}$ and with the CID. We found that Rtt103$_{1-246}$ binds to the minimal CTD-binding moiety with a $K_D$ of 3.3 ± 0.06 μM, and to the long CTD substrate with a $K_D$ of 0.3 ± 0.01 μM. The isolated CID binds to the minimal CTD-binding moiety with a $K_D$ of 1.65 ± 0.13 μM, and to the long CTD substrate with a $K_D$ of 0.5 ± 0.01 μM. This suggests that the dimerization increases the local concentration of CIDs that are available for the CTD binding. Altogether, the key finding is that dimerization of the Rtt103 coiled-coil domains does not promote formation of a rigid structure between the Rtt103 CIDs, but in fact helps the CIDs sample multiple conformations restricted only by their tethering to the flexible linker.

**The Rtt103 Coiled-Coil Domain Restricts the Variability of the CTD–CIDs Assembly.** Next, we asked whether the coiled-coil-mediated dimerization of Rtt103 affects the overall fashion in which the repetitive CTD sequence is decorated with Rtt103 CIDs. The complex between Rtt103$_{1-246}$ and pS$_2$E-CTD {pSer2-CTD mimic [SPEFTCEPTSPS-(YEPTSPS)$_{13}$-YEPAAADYKDDDDK]; Fig. S6} was prepared by mixing individual proteins with molar excess of Rtt103$_{1-246}$, followed by size-exclusion chromatography and SAXS data collection. The estimation of the molecular mass
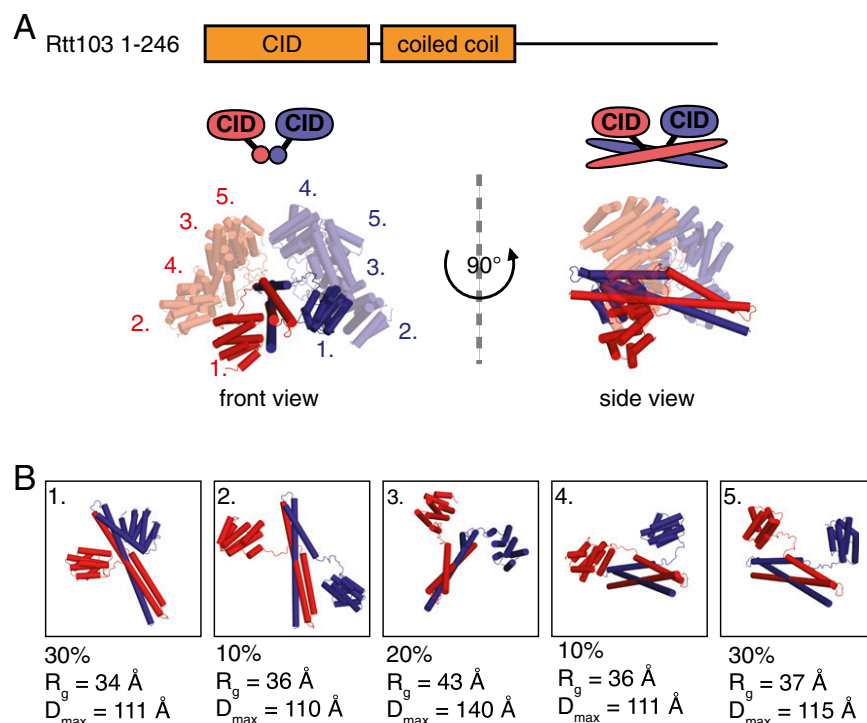
**Fig. 2.** The two Rtt103 CIDs are tethered by a coiled-coil domain but tumble independently. (*A*) Overlay of individual conformations from the ensemble of free Rtt103$_{1-246}$ structures derived using the ensemble-optimization method (EOM 2.0) (31) ($\chi^2$ = 1.001); front and side views are provided. Conformations are superimposed based on structure of the coiled-coil domain. Conformations 2–5 are shown with 60% opacity. The two different polypeptide chains are shown in red and blue. (*B*) Individual conformations from the Rtt103$_{1-246}$ ensemble derived by EOM 2.0; the fraction (%), radius of gyration ($R_g$), and maximum intraparticle distance ($D_{max}$) are indicated for each conformation.

(MM) of the complex was done using DAMMIF (32), which yielded a MM of 200 ± 5 kDa, that is close to the theoretical MM of the Rtt103$_{1-246}$:CTD complex with a ratio of 6:1 (190 kDa). In terms of molecular architecture, it suggests that three Rtt103 dimers bind to a 13-repeat-long CTD upon saturation. The interpretation of the scattering data was performed using the CORAL software (33). The structures of the Rtt103 CID (PDB ID code: 5M9D) and coiled-coil domain (PDB ID code: 5M48) were combined together with the distance constraints between the CTD and Rtt103 CID and fitted against the experimental SAXS data. The calculation was repeated 10 times for each interaction scenario with a ratio of 6:1 for Rtt103$_{1-246}$:CTD, which provided the best fit to the experimental data (Fig. 3 and Fig. S7). All reconstituted complexes displayed a similar elongated architecture characterized by a $D_{max}$ value of 180–250 Å (Fig. S7*B*). One Rtt103 dimer is accommodated on four CTD repeats (**PS YEPTSPS Y**EPTSPS **YEPTSPS Y**EPTS; CID recognition sites are shown in bold). In this architecture, the coiled-coil domains surround the individual Rtt103 CIDs accommodated on the CTD (Fig. 3). The shielding provided by the coiled coils restricts the flexibility of CIDs on the CTD to some extent, but promotes the stretching of the compact random coil structure of the CTD (22, 23). Interestingly, we obtained a similar quality fit to the experimental data without including the constraint that two CIDs of the dimer must bind neighboring CTD epitopes (Fig. S7). The obtained models contain CIDs accommodated in different areas of the CTD, supporting the hypothesis of residual flexibility in the core of the CTDsome shielded from the outside by coiled coils. Altogether, the key finding is that the CTDsome architecture is dynamic and allows for optimal recognition of available phosphorylation signals in the CTD. This variability is essential as the CTD contains some poorly conserved heptads (2) whose recognition is promoted by dimerization in which the

coiled-coil domains prime the sampling of the CTD epitopes (Fig. 3 and Fig. S7).

## Discussion

Assembly and reassembly of the CTDsome during transcription by RNAPII is important for regulation of transcription and RNA processing. However, due to the structural complexity and dynamics of the CTDsome, the mechanistic aspects of this process remain poorly understood. Here, we report an experimentally based structural model of the CTDsome, which has been derived using a combination of X-ray, NMR, and SAXS data (Fig. 3).

First, we determined that Rtt103 is capable of dimerizing in free form via a coiled-coil domain. Several CID-containing proteins are known to have multimerization regions such as the Nrd1-Nab3 heterodimerization region (34), coiled-coil region in Pcf11 (35), and coiled-coil regions in RPRD1A, RPRD1B, and RPRD2 (27, 36). The RPRD1A-RPRD1B heterodimer binds to multiple pS$_2$-CTD repeats and exposes the pS$_5$ sites on the CTD, which stimulates the activity of RPAP2 pS$_5$-CTD phosphatase. It is likely that the Rtt103 scaffold is used to recruit other factors or enzymes (e.g., Rat1-Rai1) that act on the CTD. In contrast to RPRD1A and RPRD1B, which include only the CID and dimerization domain, Rtt103 has a long unstructured C-terminal part that occupies half of the protein and, therefore, could greatly impact multisubunit architectures and interactions with other RNA/protein-binding factors.

Here, we determined that the Rtt103 CID binds across three CTD heptads and that the minimal CTD binding moiety consists of the P$_{6a}$pS$_{7a}$Y$_{1b}$pS$_{2b}$P$_{3b}$T$_{4b}$S$_{5b}$P$_{6b}$pS$_{7b}$Y$_{1c}$ sequence. These findings indicate that the Rtt103 CID binds a longer CTD stretch than previously reported (29). Accommodation of the core P$_{6a}$S$_{7a}$Y$_{1b}$S$_{2b}$P$_{3b}$T$_{4b}$S$_{5b}$ stretch of the CTD in the CID binding pocket is highly conserved among CID–CTD peptide complexes (26, 27, 37, 38). In contrast, the conformation of the upstream

and especially downstream region of the CTD peptide on the CID surface varies among CID–CTD peptide complexes (26, 27, 37, 38). In our solution NMR structure, the $pS_2pS_7$-CTD peptide makes multiple contacts with a conserved Arg116 residue and exits the binding pocket between helices α4 and α7, thereby occupying almost the whole conserved surface of the CID. Similarly, the solution structure of $pT_4$-CTD–Rtt103-CID and crystal structure of RPRD1B/RPRD1A also exhibit an elongated conformation of the CTD (27, 30). In these structures, the binding of additional residues at the C-terminal of the β-turn stretch of the CTD significantly changes the conformation of bound CTD, which could be important for the higher order arrangement of CIDs and exposure of the nonbound CTD residues to other factors (27). Additionally, the extended interaction surface of the CIDs in Rtt103 and RPRD1B/RPRD1A may partially
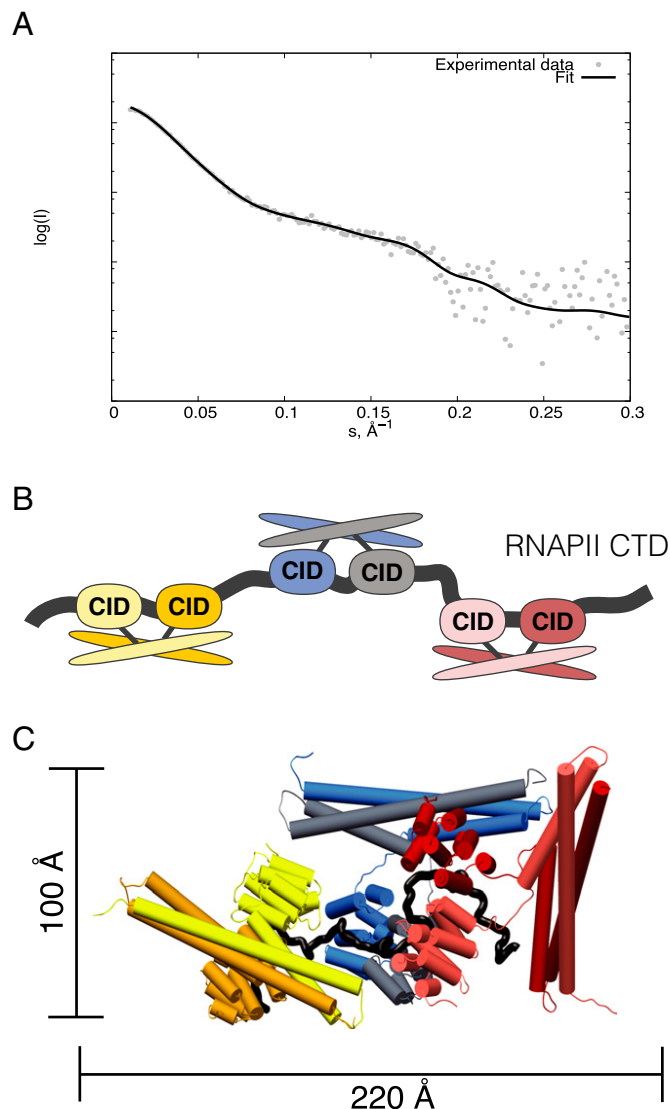


**Fig. 4.** Model of the Rtt103-CTDsome assembly involving the full-length RNAPII CTD. The model of the RNAPII with the full-length CTD is decorated with six dimers of Rtt103$_{1-246}$ (Movie S2). The structure of RNAPII (PDB ID code: 5F12) is combined with two CORAL models of the Rtt103$_{1-246}$–CTD complex, where C-interacting domains (CIDs) are arranged in a consecutive manner.

explain their higher affinity toward the CTD compared with the affinity of the CIDs in Nrd1 and Pcf11 (38, 39).

The regulation of transcription requires a complex interplay involving fast and dynamic exchange of multiple RNA/protein-binding factors. This network is largely maintained and balanced by means of a structurally adaptable CTD which increases the local concentration of factors and allosterically regulates their activity of transcription and RNA processing factors near the emerging nascent transcript. Our reconstruction of the CTD in complex with Rtt103 shows that Rtt103 can fully explore the repetitiveness and length of the CTD sequence by occupying CTD in a repetitive manner ("beads-on-a-string") while keeping the entire arrangement flexible and dynamic. Rtt103 dimerization creates topological and mobility restraints, which, in turn, tune the protein's affinity toward the CTD by increasing the local concentration of CIDs, and further governs the exposure of the CTD sequence to other protein-binding factors. We suggest that CTD code readers, such as Rtt103, and other CTD effector molecules form a high-order structure that is essential for the conception and interpretation of the CTD code (Fig. 4 and Movie S2). The tail-like architecture allows for quick exchange of binding factors and coordinates the regulatory networks necessary for efficient gene regulation. Interestingly, the structure of the CTD tail decorated with Rtt103 dimers appears to be fully extended and protrudes away from the invariant core of the RNAPII (Fig. 4 and Movie S2). The structural model of the Rtt103–CTDsome demonstrates how the CTD allows forming diverse and tuneable protein assemblies around the invariant core of the RNAPII, supporting the complex networks necessary for efficient gene regulation.

## Methods

A full description of the methods for protein expression, purification, and fluorescence anisotropy measurements as well as NMR, X-ray, and SAXS data collection and analysis is provided in *SI Methods* and Tables S5 and S6. SAXS data are deposited in Small Angle Scattering Biological Data Bank (SASBDB ID: SASDCZ2). The model and the diffraction data containing phase information was deposited to Protein Data Bank, PDB ID code: 5M48. The atomic coordinates and restraints for the NMR ensemble of the Rtt103-CID—$pS_2pS_7$-CTD complex have been deposited in the Protein Data Bank, PDB ID code: 5M9D.

**Fig. 3.** Multisubunit arrangement of Rtt103$_{1-246}$ across half the length of the CTD. (*A*) Comparison of the theoretical scattering (black) derived using CORAL based on the Rtt103$_{1-246}$-CTD model against the experimental scattering data (gray). (*B*) Scheme showing the arrangement of the individual Rtt103$_{1-246}$ molecules across half the length of the CTD (13 heptad repeats), which is the CTD construct employed during modeling using CORAL (consecutive interaction, "scenario 0," see Fig. S7). (*C*) Representative model obtained from the CORAL calculation for the consecutive interaction scenario. Color coding is according to the scheme in *B*.
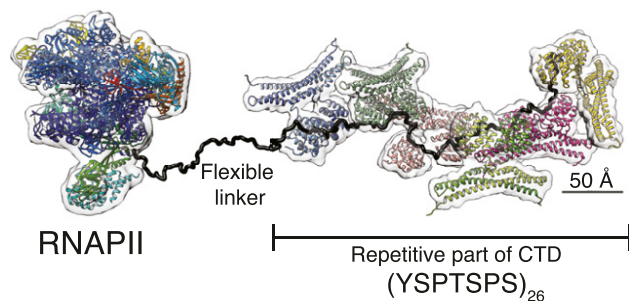
BIOCHEMISTRY

1. Eick D, Geyer M (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* 113:8456–8490.
2. Chapman RD, Heidemann M, Hintermair C, Eick D (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet* 24:289–296.
3. Harlen KM, Churchman LS (2017) The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* 18:263–273.
4. West ML, Corden JL (1995) Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140:1223–1233.
5. Liu P, Kenney JM, Stiller JW, Greenleaf AL (2010) Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol Biol Evol* 27:2628–2641.
6. Buratowski S (2003) The CTD code. *Nat Struct Biol* 10:679–680.
7. Mayer A, et al. (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17:1272–1278.
8. Mayer A, et al. (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336:1723–1725.
9. Bataille AR, et al. (2012) A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* 45:158–170.
10. Kim H, et al. (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* 17:1279–1286.
11. Tietjen JR, et al. (2010) Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* 17:1154–1161.
12. Heidemann M, Hintermair C, Voß K, Eick D (2013) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* 1829:55–62.
13. Suh H, et al. (2016) Direct analysis of phosphorylation sites on the Rpb1 C-terminal domain of RNA polymerase II. *Mol Cell* 61:297–304.
14. Schüller R, et al. (2016) Heptad-specific phosphorylation of RNA polymerase II CTD. *Mol Cell* 61:305–314.
15. Harlen KM, et al. (2016) Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep* 15:2147–2158.
16. Cramer P, et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288:640–649.
17. Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292:1863–1876.
18. Spåhr H, Calero G, Bushnell DA, Kornberg RD (2009) Schizosacharomyces pombe RNA polymerase II at 3.6-Å resolution. *Proc Natl Acad Sci USA* 106:9185–9190.
19. Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. *Genes Dev* 19:1401–1415.
20. Meredith GD, et al. (1996) The C-terminal domain revealed in the structure of RNA polymerase II. *J Mol Biol* 258:413–419.
21. Tsai K-L, et al. (2013) A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nat Struct Mol Biol* 20:611–619.
22. Portz B, et al. (2017) Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat Commun* 8:15231.
23. Gibbs EB, et al. (2017) Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat Commun* 8:15233.
24. Jasnovidova O, Stefl R (2013) The CTD code of RNA polymerase II: A structural view. *Wiley Interdiscip Rev RNA* 4:1–16.
25. Barillà D, Lee BA, Proudfoot NJ (2001) Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* 98:445–450.
26. Meinhart A, Cramer P (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature* 430:223–226.
27. Ni Z, et al. (2014) RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nat Struct Mol Biol* 21:686–695.
28. Grigoryan G, Degrado WF (2011) Probing designability via a generalized model of helical bundle geometry. *J Mol Biol* 405:1079–1100.
29. Lunde BM, et al. (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 17:1195–1201.
30. Jasnovidova O, Krejcikova M, Kubicek K, Stefl R (2017) Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO Rep* 18:906–913.
31. Tria G, Mertens HDT, Kachala M, Svergun DI (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* 2:207–217.
32. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 42:342–346.
33. Petoukhov MV, et al. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45:342–350.
34. Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 15:795–804.
35. Xu X, Pérébaskine N, Minvielle-Sébastia L, Fribourg S, Mackereth CD (2015) Chemical shift assignments of a new folded domain from yeast Pcf11. *Biomol NMR Assign* 9:421–425.
36. Mei K, et al. (2014) Structural basis for the recognition of RNA polymerase II C-terminal domain by CREPT and p15RS. *Sci China Life Sci* 57:97–106.
37. Becker R, Loll B, Meinhart A (2008) Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* 283:22659–22669.
38. Kubicek K, et al. (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* 26:1891–1896.
39. Noble CG, et al. (2005) Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* 12:144–151.
40. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76:2879–2886.
41. Mossessova E, Lima CD (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* 5:865–876.
42. Svergun D, Barberato C, Koch MHJ (1995) CRYSOL–A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773.
43. Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 34:33–41.
44. Kuzmic P (2009) DynaFit–A software package for enzymology. *Methods Enzymol* 467:247–280.
45. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66:125–132.
46. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62:72–82.
47. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–242.
48. Adams PD, et al. (2011) The Phenix software for automated determination of macromolecular structures. *Methods* 55:94–106.
49. McCoy AJ, Storoni LC, Read RJ (2004) Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallogr D Biol Crystallogr* 60:1220–1228.
50. Grosse-Kunstleve RW, Adams PD (2003) Substructure search procedures for macromolecular structures. *Acta Crystallogr D Biol Crystallogr* 59:1966–1973.
51. Terwilliger T (2004) SOLVE and RESOLVE: Automated structure solution, density modification and model building. *J Synchrotron Radiat* 11:49–52.
52. Terwilliger TC, et al. (2009) Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65:582–601.
53. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62:1002–1011.
54. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
55. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.
56. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68:352–367.
57. Pettersen EF, et al. (2004) UCSF Chimera–A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
58. Walshaw J, Woolfson DN (2001) Socket: A program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol* 307:1427–1450.
59. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.

# 6   List of Author's Publications

Throughout my scientific career I have authored or co-authored the below-listed publications. The list shows my focus on studies of biomolecular structures using NMR in their free or complexed form as well as on preparation of protein or peptide samples to study these with various biophysical methods complementary to NMR.

1) Brázda, P., Šedo, O., **Kubíček, K.**, Štefl, R.: Efficient and robust preparation of tyrosine phosphorylated intrinsically disordered proteins. *Biotechniques.* **2019** 67(1):16–22.

2) Hemzal, D., Kang, Y.R., Dvořák, J., Kabzinski, T., **Kubíček, K.**, Kim, Y.D., Humlíček, J.: Treatment of Surface Plasmon Resonance (SPR) Background in Total Internal Reflection Ellipsometry: Characterization of RNA Polymerase II Film Formation. *Appl Spectrosc.* **2019** 73(3):261–270.

3) Jasnovidova, O., Klumpler, T., **Kubicek, K.**, Kalynych, S., Plevka, P., Stefl, R.: Structure and dynamics of the RNAPII CTDsome with Rtt103. *Proc Natl Acad Sci U S A.* **2017** 114(42):11133–11138.

4) Jasnovidova, O., Krejcikova, M., **Kubicek, K.**, Stefl, R.: Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO Rep. 2017* 18(6):906–913.

5) Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., **Kubíček, K.**, Fořtová, A., Lacroute, F., Vaňáčová, Š., Cramer, P., Štefl, R., Libri, D.: Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell.* **2014** 55(3): 467–81.

6) Bačíková, V., Pasulka, J., **Kubíček, K.**, Štefl, R.: Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res.* **2014** 42(12): 8024–38.

7) Šikorský, T., Hobor, F., Križanová, E., Pasulka, J., **Kubíček, K.**, Štefl, R.: Recognition of asymmetrically dimethylated arginine by TDRD3. *Nucleic Acids Res.* **2012** 40(22): 11748–55.

8) Porrua, O., Hobor, F., Boulay, J., **Kubíček, K.**, D'Aubenton-Carafa, Y., Gudipati, R.K., Štefl, R., Libri, D.: *In vivo* SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.* **2012** 31(19): 3935–48.

9) **Kubíček, K.**, Černá, H., Holub, P., Pasulka, J., Hroššová, D., Löhr, F., Hofr, C., Vaňáčová, Š., Štefl, R.: Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* **2012** 26(17): 1891–6.

10) Suchánková, T., **Kubíček, K.**, Kašpárková, J., Brabec, V., Kozelka, J.: Platinum-DNA interstrand crosslinks: molecular determinants of bending and unwinding of the double helix. *J. Inorg. Biochem.* **2012** 108: 69–79.

11) **Kubíček, K.**, Pasulka, J., Černá, H., Löhr, F., Štefl, R.: $^{1}$H, $^{13}$C, and $^{15}$ resonance assignments for the CTD-interacting domain of Nrd1 bound to Ser5-phosphorylated CTD of RNA polymerase II. *Biomol. NMR Assign.* **2011** 5(2): 203–5.

12) Hobor, F., Pergoli, R., **Kubíček, K.**, Hroššová, D., Bačíková, V., Zimmermann, M., Pasulka, J., Hofr, C., Vaňáčová, Š., Štefl, R.: Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J. Biol. Chem.* **2011** 286(5): 3645–57.

13) **Kubíček, K.**, Grimm, S.K., Orts, J., Sasse, F., Carlomagno, T.: The tubulin-bound structure of the antimitotic drug tubulysin. *Angew. Chem. Int. Ed.* **2010** 49(28):4809–12.

14) Pergoli, R., **Kubíček, K.**, Hobor, F., Pasulka, J., Štefl, R.: $^{1}$H, $^{13}$C, and $^{15}$N chemical shift assignments for the RNA recognition motif of Nab3. *Biomol. NMR Assign.* **2010** 4(1): 119–21.

15) **Kubíček, K.**, Monnet, J., Scintilla, S., Kopečná, J., Arnesano, F., Trantírek, L., Chopard, C., Natile, G., Kozelka, J.: Unusual interstrand Pt(S,S-diaminocyclohexane)-GG crosslink formed by rearrangement of a classical intrastrand crosslink within a DNA duplex. *Chem. Asian. J.* **2010** 5(2): 244–7.

16) Reese, M., Sanchez-Pedregal, V.M., **Kubíček, K.**, Meiler, J., Blommers, M.J., Griesinger, C., Carlomagno, T.: Structural basis of the activity of the microtubule-stabilizing agent epothilone a studied by NMR spectroscopy in solution. *Angew. Chem. Int. Ed.* **2007** 46(11): 1864–8.

17) Sanchez-Pedregal, V.M., **Kubíček, K.**, Meiler, J., Lyothier, I., Paterson, I., Carlomagno, T.: The tubulin-bound conformation of discodermolide derived by NMR studies in solution supports a common pharmacophore

model for epothilone and discodermolide. *Angew. Chem. Int. Ed.* **2006** 45(44): 7388–94.

18) Banci, L., Bertini, I., Ciofi-Baffoni, S., Katsari, E., Katsaros, N., **Kubíček, K.**, Mangani, S.: A copper(I) protein possibly involved in the assembly of $Cu_A$ center of bacterial cytochrome $c$ oxidase. *Proc. Natl. Acad. Sci.* **2005** 102(11): 3994–9.

19) Banci, L., Bertini, I., Felli, I.C., Krippahl, L., **Kubíček, K.**, Moura, J.J., Rosato, A.: A further investigation of the cytochrome $b_5$–cytochrome $c$ complex. *J. Biol. Inorg. Chem.* **2003** 8(7): 777–86

# 7 Summary

Progress in structural biology would never be possible without development in all fields of science that are necessary for structure determination. Structural biology is multidisciplinary technique and the first step of research is obtaining the sample. Homogeneous, high purity, large amounts, longevity stability etc. The biological technologies of biomolecular sample preparation made enormous progress so we are not limited to small colourful proteins like cytochromes which can be detected on chromatographic columns with naked eyes but we are studying huge biomolecules called "molecular machines"[54] (e.g. RNA Polymerase I, II, and III ) and if these are unstable, we stabilize them with crosslinking and still obtain biologically relevant structural information.

Through my research projects and here included scientific papers I was witnessing also the progress of hardware development which would not be possible without technological and engineering advancement. At the beginning of my carrier, 700-800 MHz (17.6-18.8 T) NMR spectrometers were the cutting edge models, while recently 1.2 GHz (25.9 T) spectrometers were presented and are opening new methodological possibilities.

In electron microscopy, we can see 300 kV microscopes with the capability to acquire data for structures with resolution below 2 Angstr. And Volta phase plate (VPP) allows to scan molecules as small as 40 kDa [55].

Improvement in resolution, scanning rates and possibilities to measure more complex samples requires larger and faster data storage devices and faster computers. And here again, development in software was needed so the computers can handle such large datasets.

To sum up the above paragraphs once again – development in each and every part of the process of structure determination was needed to arrive to the point where the structural biology field is nowadays. And still there are three facts to be pointed out: 1) In electron microscopy – 85-952) There is about 151 thousands of structures deposited on the Protein DataBank (http://www.rcsb.org; as of 30.4.2019) with yearly increase about 12 thousands structures plateauing within last 4 years. Nevertheless, among these structures there is no new fold discovered since 2012 (PIC). Major part of the protein folds has been discovered between 1993 and 2008. Since than we can be rather sure that the newly determined structures will contain some of the previously known structural motif. 3) Looking at the statistics, one can also see, that NMR was never major contributor to the resolved structures and it has been always outcompeted by X-ray crystallography. Since approximately 2014-5, X-ray's position as a main source of high resolution structures starts to be challenged by electron microscopy.

What does this is message I want to emphasize? Structural biologist are in very comfort situation where they are not limited by methods or availability of measurement time but rather by their scientific goals and capabilities of preparing samples good for measurement in any of these methods.

While listing here the impressive high number of all known protein structures available and the fact that since last 7 years there was no new protein fold discovered, the logical question would be – how many more structures and how

much more computational power do we need to tell the three dimensional protein structure from its amino-acid sequence (or in other words: resolve all the aspects of Anfinsen dogma)? I have a very good argument about this with my dear colleague – prof. Robert Vácha. We both agree at the point that computers (and artificial intelligence, AI) are still far from determine protein structure (rather unique fold) which has not been discovered so far. Presuming there are all protein folds known, this is not the limiting step anymore. We have a large spectrum of techniques that I have mentioned within this thesis (NMR, cryoEM, SAXS, X-ray, MS) that can provide experimental data with various quality and resolution content so there has to be a way how to quickly predict 3D protein structure from its amino-acid sequence. Prof. Vácha says – if you'd have a library with 150 thousands of books would that be sufficient to predict, what will be the next one about despite you know all the words? I am saying that there has to be a way and we will find the way sooner or later and programs like AlphaFold [56] are giving us some hint how this may be achieved.

However, I think, that the goal of structural biology is not to solve maximum structures in minimum time with highest resolution of the structures but, as I have shown in my works – help to explain biological processes through understanding structural features of the biological partners of a given process.

The development and the ultimate goal of structural biology is to study the processes in a real-time and in situ. Therefore we can see much effort in structure determination in cells either by NMR or by cryoEM on FIB milled lamellas . This is the direction which will enable us understand and suggest efficient cures for diseases and unravel the fundamental principles of life in cells.

# References

[1] Helen M. Berman, Buvaneswari Coimbatore Narayanan, Luigi Di Costanzo, Shuchismita Dutta, Sutapa Ghosh, Brian P. Hudson, Catherine L. Lawson, Ezra Peisach, Andreas Prlić, Peter W. Rose, Chenghua Shao, Huanwang Yang, Jasmine Young, and Christine Zardecki. Trendspotting in the Protein Data Bank. *FEBS Letters*, 587(8):1036–1045, apr 2013.

[2] Lu-Yun Lian. NMR studies of weak protein–protein interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 71:59–72, may 2013.

[3] Paul J. Barrett, Jiang Chen, Min Kyu Cho, Ji Hun Kim, Zhenwei Lu, Sijo Mathew, Dungeng Peng, Yuanli Song, Wade D. Van Horn, Tiandi Zhuang, Frank D. Sönnichsen, and Charles R. Sanders. The quiet renaissance of protein nuclear magnetic resonance. *Biochemistry*, 52(8):1303–1320, 2013.

[4] Wolfgang Bermel, Ivano Bertini, Isabella C. Felli, Yong Min Lee, Claudio Luchinat, and Roberta Pierattelli. Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *Journal of the American Chemical Society*, 128(12):3918–3919, 2006.

[5] Marc Boehning, Claire Dugast-Darzacq, Marija Rankovic, Anders S. Hansen, Taekyung Yu, Herve Marie-Nelly, David T. McSwiggen, Goran Kokic, Gina M. Dailey, Patrick Cramer, Xavier Darzacq, and Markus Zweckstetter. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature Structural & Molecular Biology*, 25(9):833–840, sep 2018.

[6] Eva Nogales, Sharon G. Wolf, and Kenneth H. Downing. Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature*, 391(6663):199–203, 1998.

[7] Benoît Gigant, Chunguang Wang, Raimond B.G. Ravelli, Fanny Roussi, Michel O. Steinmetz, Patrick A. Curmi, André Sobel, and Marcel Knossow. Structural basis for the regulation of tubulin by vinblastine. *Nature*, 435(7041):519–522, 2005.

[8] Francis J. McNally. Microtubule dynamics: Controlling split ends. *Current Biology*, 9(8):R274–R276, apr 1999.

[9] E Nogales, M Whittaker, R A Milligan, and K H Downing. High-resolution model of the microtubule. *Cell*, 96(1):79–88, jan 1999.

[10] Bonnie Howell, Niklas Larsson, Martin Gullberg, and Lynne Cassimeris. Dissociation of the Tubulin-sequestering and Microtubule Catastrophe-promoting Activities of Oncoprotein 18/Stathmin. *Molecular Biology of the Cell*, 10(1):105–118, jan 1999.

[11] Axel Sandmann, Florenz Sasse, and Rolf Müller. Identification and Analysis of the Core Biosynthetic Machinery of Tubulysin, a Potent Cytotoxin with Potential Anticancer Activity. *Chemistry & Biology*, 11(8):1071–1079, aug 2004.

[12] Karel Kubicek, S.Kaspar Grimm, Julien Orts, Florenz Sasse, and Teresa Carlomagno. The Tubulin-Bound Structure of the Antimitotic Drug Tubulysin. *Angewandte Chemie International Edition*, 49(28):4809–4812, jun 2010.

[13] Jun Xu, Indrajit Lahiri, Wei Wang, Adam Wier, Michael A. Cianfrocco, Jenny Chong, Alissa A. Hare, Peter B. Dervan, Frank DiMaio, Andres E. Leschziner, and Dong Wang. Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature*, 551(7682):653–657, nov 2017.

[14] Alan C M Cheung and Patrick Cramer. A movie of RNA polymerase II transcription. *Cell*, 149(7):1431–7, jun 2012.

[15] Dirk Eick and Matthias Geyer. The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chemical Reviews*, 113(11):8456–8490, nov 2013.

[16] Sylvain Egloff, Martin Dienstbier, and Shona Murphy. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends in Genetics*, 28(7):333–341, jul 2012.

[17] S. Yogesha, Joshua Mayfield, and Yan Zhang. Cross-Talk of Phosphorylation and Prolyl Isomerization of the C-terminal Domain of RNA Polymerase II. *Molecules*, 19(2):1481–1511, jan 2014.

[18] Olga Jasnovidova and Richard Stefl. The CTD code of RNA polymerase II: a structural view. *Wiley Interdisciplinary Reviews: RNA*, 4(1):1–16, jan 2013.

[19] Karel Kubicek, Hana Cerna, Peter Holub, Josef Pasulka, Dominika Hrossova, Frank Loehr, Ctirad Hofr, Stepanka Vanacova, and Richard Stefl. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes & development*, 26(17):1891–6, sep 2012.

[20] Manuel J. Muñoz, Manuel de la Mata, and Alberto R. Kornblihtt. The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends in Biochemical Sciences*, 35(9):497–504, sep 2010.

[21] Beate Schwer, Ana M Sanchez, and Stewart Shuman. Punctuation and syntax of the RNA polymerase II CTD code in fission yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44):18024–9, oct 2012.

[22] Roland Schüller, Ignasi Forné, Tobias Straub, Amelie Schreieck, Yves Texier, Nilay Shah, Tim-Michael Decker, Patrick Cramer, Axel Imhof, and Dirk Eick. Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Molecular Cell*, 61(2):305–314, jan 2016.

[23] Rakesh Srivastava and Seong Hoon Ahn. Modifications of RNA polymerase II CTD: Connections to the histone code and cellular function. *Biotechnology Advances*, 33(6):856–872, nov 2015.

[24] Kevin M Harlen and L Stirling Churchman. Subgenic Pol II interactomes identify region-specific transcription elongation regulators. *Molecular Systems Biology*, 13(1):900, jan 2017.

[25] D. Barilla, B. A. Lee, and N. J. Proudfoot. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 98(2):445–450, jan 2001.

[26] Zuyao Ni, Jonathan B. Olsen, Xinghua Guo, Guoqing Zhong, Eric Dongliang Ruan, Edyta Marcon, Peter Young, Hongbo Guo, Joyce Li, Jason Moffat, Andrew Emili, and Jack F. Greenblatt. Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B. *Transcription*, 2(5):201–206, 2011.

[27] Christian G Noble, David Hollingworth, Stephen R Martin, Valerie Ennis-Adeniran, Stephen J Smerdon, Geoff Kelly, Ian A Taylor, and Andres Ramos. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nature Structural & Molecular Biology*, 12(2):144–151, feb 2005.

[28] Roland Becker, Bernhard Loll, and Anton Meinhart. Snapshots of the RNA Processing Factor SCAF8 Bound to Different Phosphorylated Forms of the Carboxyl-terminal Domain of RNA Polymerase II. *Journal of Biological Chemistry*, 283(33):22659–22669, aug 2008.

[29] Lidia Vasiljeva, Minkyu Kim, Hannes Mutschler, Stephen Buratowski, and Anton Meinhart. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature Structural and Molecular Biology*, 15(8):795–804, 2008.

[30] Wolfgang Mühlbacher, Andreas Mayer, Mai Sun, Michael Remmert, Alan C. M. Cheung, Jürgen Niesser, Johannes Soeding, and Patrick Cramer. Structure of Ctk3, a subunit of the RNA polymerase II CTD kinase complex, reveals a noncanonical CTD-interacting domain fold. *Proteins: Structure, Function, and Bioinformatics*, 83(10):1849–1858, oct 2015.

[31] Lloyd Kelland. The resurgence of platinum-based cancer chemotherapy. *Nature Reviews Cancer*, 7(8):573–584, 2007.

[32] Ernest Wong and Christen M. Giandornenico. Current status of platinum-based antitumor drugs. *Chemical Reviews*, 99(9):2451–2466, 1999.

[33] J H Burchenal, K Kalaher, K Dew, and L Lokys. Rationale for development of platinum analogs. *Cancer treatment reports*, 63(9-10):1493–8.

[34] Bernhard Lippert. *Cisplatin: Chemistry and biochemistry of a leading anticancer drug.* 2006.

[35] Rebecca A. Alderden, Matthew D. Hall, and Trevor W. Hambley. The discovery and development of cisplatin. *Journal of Chemical Education*, 83(5):728–734, 2006.

[36] D Lebwohl and R Canetta. Clinical development of platinum complexes in cancer therapy: an historical perspective and an update. *European Journal of Cancer*, 34(10):1522–1534, sep 1998.

[37] Teni Boulikas. Clinical overview on Lipoplatin$^{TM}$: A successful liposomal formulation of cisplatin. *Expert Opinion on Investigational Drugs*, 18(8):1197–1218, 2009.

[38] Ming Wang, Laurent Ogé, Maria Dolores Perez-Garcia, Latifa Hamama, and Soulaiman Sakr. The PUF protein family: Overview on PUF RNA targets, biological functions, and post transcriptional regulation. *International Journal of Molecular Sciences*, 19(2):1–13, 2018.

[39] Christophe Maris, Cyril Dominguez, and Frédéric H.T. Allain. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal*, 272(9):2118–2131, 2005.

[40] Gideon Dreyfuss, Maurice S. Swanson, and Serafin Piñol-Roma. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends in Biochemical Sciences*, 13(3):86–91, mar 1988.

[41] Cyprian D. Cukier, David Hollingworth, Stephen R. Martin, Geoff Kelly, Irene Díaz-Moreno, and Andres Ramos. Molecular basis of FIR-mediated c-myc transcriptional control. *Nature Structural and Molecular Biology*, 17(9):1058–1064, 2010.

[42] Gerrit M Daubner, Antoine Cléry, Sandrine Jayne, James Stevenin, and Frédéric H-T Allain. A *syn* - *anti* conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *The EMBO Journal*, 31(1):162–174, jan 2012.

[43] Jiří Nováček, Lubomír Janda, Radka Dopitová, Lukáš Žídek, and Vladimír Sklenář. Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. *Journal of Biomolecular NMR*, 56(4):291–301, aug 2013.

[44] Michael Sattler, Jürgen Schleucher, and Christian Griesinger. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34(2):93–158, 1999.

[45] Arthur G. Palmer, John Williams, and Ann McDermott. Nuclear magnetic resonance studies of biopolymer dynamics. *Journal of Physical Chemistry*, 100(31):13293–13310, 1996.

[46] D.M. Korzhnev, M. Billeter, A.S. Arseniev, and V.Y. Orekhov. NMR studies of Brownian tumbling and internal motions in proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 38(3):197–266, apr 2001.

[47] L. Y. Lian, I. L. Barsukov, M. J. Sutcliffe, K. H. Sze, and G. C. K. Roberts. Protein-Ligand Interactions: Exchange Processes and Determination of Ligand Conformation and Protein-Ligand Contacts. *Methods Enzymol.*, 239:657–700, 1994.

[48] J. Feeney, J. G. Batchelor, J. P. Albrand, and G. C. K. Roberts. The Effects of Intermediate Exchange Processes on the Estimation of Equilibrium Constnants by NMR. *J. Mag. Res.*, 33:519–529, 1979.

[49] A. Kannt, S. Young, and D. S. Bendall. The role of acidic residues of plastocyanin in its interaction with cytochrome *f*. *Biochim. Biophys. Acta*, 1277:115–126, 1996.

[50] M. J. Rodríguez-Marañon, F. Qiu, R. E. Stark, S. P. White, X. Zhang, S. I. Foundling, V. Rodríguez, C. L. Schilling III, R. A. Bunce, and M. Rivera. $^{13}$C NMR Spectroscopic and X-ray Crystallographic Study of the Role Played by Mitochondrial Cytochrome $b_5$ Heme Propionates in the Electrostatic Binding to Cytochrome *c*. *Biochemistry*, 35:16378–16390, 1996.

[51] Worrall J. A., W. Reinle, R. Bernhardt, and Ubbink M. Transient Protein Interactions Studied by NMR Spectroscopy: The Case of Cytochrome *c* and Adrenodoxin. *Biochemistry*, 42:7068–7076, 2003.

[52] S. Grzesiek, A. Bax, G. M. Clore, A. M. Gronenborn, J. S. Hu, J. Kaufman, I. Palmer, S. J. Stahl, and P. T. Wingfield. The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. *Nat. Struct. Biol.*, 3:340–345, 1996.

[53] W. Shao, S. C. Im, E. R. Zuiderweg, and L. Waskell. Mapping the Binding Interface of the Cytochrome $b_5$-Cytochrome *c* Complex by Nuclear Magnetic Resonance. *Biochemistry*, 42:14774–14784, 2003.

[54] Matthew D Liptak, Robert D Fagerlund, Elizabeth C Ledgerwood, Sigurd M Wilbanks, and Kara L Bren. The proapoptotic G41S mutation to human cytochrome c alters the heme electronic structure and increases the electron self-exchange rate. *Journal of the American Chemical Society*, 133(5):1153–5, feb 2011.

[55] Kaiming Zhang, Shanshan Li, Kalli Kappel, Grigore Pintilie, Zhaoming Su, Tung-Chung Mou, Michael F. Schmid, Rhiju Das, and Wah Chiu. Cryo-EM

structure of a 40 kDa SAM-IV riboswitch RNA at 3.7Å resolution. *Nature Communications*, 10(1):5511, dec 2019.

[56] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins: Structure, Function, and Bioinformatics*, 87(12):1141–1148, dec 2019.

# 8 Curriculum Vitae - Karel Kubíček

## Personal data

| | |
|---|---|
| Date of Birth: | 19.9.1977 |
| Place of Birth: | Valtice, Czech Republic |
| Status: | Married |
| Children: | 2 kids |
| Nationality: | Czech |

| | |
|---|---|
| Address: | Department of Condensed Matter Physics |
| | Masaryk University |
| | Kotlářská 2 |
| | 611 37 Brno |
| | Czech Republic |
| | |
| | CEITEC-Central European Institute of Technology |
| | Structural Biology of Gene Regulation |
| | Kamenice 753/5 |
| | 625 00 Brno |
| | Czech Republic |

| | |
|---|---|
| **T:** | +420 549 49 3253 |
| **M:** | +420 731 28 7616 |
| **F:** | +420 541 21 1214 |
| **E:** | karelk@physics.muni.cz |
| | karel.kubicek@ceitec.muni.cz |

## Employment - Position

Assistant Professor
Researcher @Laboratory of Structural Biology of Gene Regulation (assoc. prof. Richard Štefl)

## Education & Academic Qualifications

| | |
|---|---|
| 2018 (5 months): | Scientific exchange within Twinning project (TWINFUSYON) Johannes Kepler University, Linz, Austria (prof. Kurt Hingerl) |
| 2008–present: | Researcher in the laboratory of prof. Richard Štefl |
| 2007–present: | Assistant Professor for Biophysics at the Dept. Condensed Matter Physics, Faculty of Science, Masaryk University, Brno, Czech Republic |

| | |
|---|---|
| 2005–2007: | Postdoctoral researcher at the Max-Planck-Institute for Biophysical Chemistry in Göttingen, Department of NMR-based Structural Biology; studies of ligand–target interactions by solution NMR in the group of Dr. Teresa Carlomagno and Prof. Christian Griesinger.) |
| 2001–2004: | International doctorate in structural biology, Magnetic Resonance Center (CERM) in Florence, Italy. Degree awarded by University of Florence, Italy. Topic of Thesis: NMR Characterization of Protein–Protein and Protein–Cofactor Interactions (Supervisor Prof. Lucia Banci) |
| 1996–2001: | Masaryk University, Faculty of Science, Department of Theoretical and Physical Chemistry, final exams passed in June 2001. Topic of Thesis: Application of Residual Dipolar Interactions for Structure Determination of Biomolecules (Supervisor Prof. Vladimír Sklenář) |
| 1992–1996: | Technical College of Chemistry in Brno |

## Grants

| | |
|---|---|
| 2015–2018 | GAČR standard grant |
| 2011–2013 | OP Vzdělávání pro konkurenceschopnost: Reforma a rozvoj výuky Biofyziky pro potřeby 21. století |
| 2008–2010 | GAČR postdoctoral grant |

## Teaching Activities

| | |
|---|---|
| 2013–present | Organisation of Biophysical competition (http://physics.muni.cz/biophys/souteze.shtml) |
| 2008–present | Mentoring students |
| 2008–present | Teaching activities (biophysics, practical and theoretical NMR spectroscopy, experimental techniques in biophysical research) |

## Scientific & Research Activities

| | |
|---|---|
| | NMR spectroscopy, AFM microscopy, protein-protein interactions, drug-design, protein-ligand interactions, SAXS, fluorescence anisotropy, isothermal titration calorimetry, molecular dynamics, electron microscopy . . . |

*University Activities*

| | |
|---|---|
| 2018–2020 | Academic Senate of the Masaryk University |
| 2015–2017 | Academic Senate of the Masaryk University |
| 2008–2011 | Academic Senate of the Faculty of Science of the Masaryk University |
| 1998–2001 | Academic Senate of the Masaryk University |
| 1996–2001 | Academic Senate of the Faculty of Science of the Masaryk University |

*Activities Outside University*

| | |
|---|---|
| | The Board of Biophysics @University of South Bohemia |

*Awards Related to Science & Research*

| | |
|---|---|
| 2006–2007 | DAAD scholarship within the program "Modern Applications of Biotechnology" |
| 2001 | Special prize of the Czech chemical society for the "Best presentation" on the Student scientific competition |
| 2001 | Chemistry section prize |
| 2000 | I. prize in the Student scientific competition |
| 2000 | Chemistry section prize |

1 JASNOVIDOVA, Olga, Magdaléna KREJČÍKOVÁ, **Karel KUBÍČEK** & Richard ŠTEFL. Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO reports*, Hoboken: Wiley-Blackwell, **2017**, roč. 18, č. 6, s. 906-913. ISSN 1469-221X. doi:10.15252/embr.201643723.

2 JASNOVIDOVA, Olga, Tomáš KLUMPLER, **Karel KUBÍČEK**, Sergei KALYNYCH, Pavel PLEVKA & Richard ŠTEFL. Structure and dynamics of the RNAPII CTDsome with Rtt103. *Proceeding of the National Academy of Sciences of the US*A, The Academy, **2017**, roč. 114, č. 42, s. 11133-11138. ISSN 0027-8424. doi:10.1073/pnas.1712450114.

3 BAČÍKOVÁ, Veronika, Josef PASULKA, **Karel KUBÍČEK** & Richard ŠTEFL. Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Research*, United Kingdom: Oxford University Press, **2014**, roč. 42, č. 12, s. 8024-8038. ISSN 0305-1048. doi:10.1093/nar/gku446.

4 TUDEK, Agnieszka, Odil Porrua FUERTE, Tomasz KABZINSKI, Michael LIDSCHREIBER, **Karel KUBÍČEK**, Andrea FOŘTOVÁ, Francoise

LACROUTE, Štěpánka VAŇÁČOVÁ, Patrick CRAMER, Richard ŠTEFL & Domenico LIBRI. Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Molecular Cell*, CAMBRIDGE: CELL PRESS, **2014**, roč. 55, č. 3, s. 467-481. ISSN 1097-2765.

5  SUCHÁNKOVÁ, Tereza, **Karel KUBÍČEK**, Jana KAŠPÁRKOVÁ, Viktor BRABEC & Jiří KOZELKA. Platinum-DNA interstrand crosslinks: Molecular determinants of bending and unwinding of the double helix. *Journal of Inorganic Biochemistry*, New York, USA: Elsevier Science Inc, **2012**, roč. 108, č. 1, s. 69-79. ISSN 0162-0134.

6  **KUBÍČEK, Karel**, Hana ČERNÁ, Peter HOLUB, Josef PASULKA, Dominika HROŠŠOVÁ, Frank LOEHR, Ctirad HOFR, Štěpánka VAŇÁČOVÁ & Richard ŠTEFL. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes & Development*, New York: Cold Spring Harbor Laboratory Press, **2012**, roč. 26, č. 17, s. 1891-1896. ISSN 0890-9369. doi:10.1101/gad.192781.112.

7  PORRUA-FUERTE, Odil, Fruzsina HÓBOR, Jocelyne BOULAY, **Karel KUBÍČEK**, Yves D'AUBENTON-CARAFA, Rajani Kanth GUDIPATI, Richard ŠTEFL & Domenico LIBRI. In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1 dependent transcription termination. *EMBO Journal*, NEW YORK: NATURE PUBLISHING GROUP, **2012**, roč. 31, č. 19, s. 3935-3948. ISSN 0261-4189. doi:10.1038/emboj.2012.237.

8  ŠIKORSKÝ, Tomáš, Fruzsina HÓBOR, Eva KRIŽANOVÁ, Josef PASULKA, **Karel KUBÍČEK** & Richard ŠTEFL. Recognition of asymmetrically dimethylated arginine by TDRD3. *Nucleic Acids Research*, Oxford: Oxford University Press, **2012**, roč. 40, č. 22, s. 11748-11755. ISSN 0305-1048. doi:10.1093/nar/gks929.

9  HÓBOR, Fruzsina, Roberto PERGOLI, **Karel KUBÍČEK**, Dominika HROŠŠOVÁ, Veronika BAČÍKOVÁ, Michal ZIMMERMANN, Josef PASULKA, Ctirad HOFR, Štěpánka VAŇÁČOVÁ & Richard ŠTEFL. Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (Nab)3 protein. *The Journal of Biological Chemistry*, Am. Soc. for Biochem. and Mol. Biol., **2011**, roč. 286, č. 5, s. 3645-3657. ISSN 0021-9258. doi:10.1074/jbc.M110.158774.

10  **KUBÍČEK, Karel**, Jordan MONNET, Simone SCINTILLA, Jana KOPEČNÁ, Fabio ARNESANO, Lukáš TRANTÍREK, Claude CHOPARD, Giovanni NATILE & Jiří KOZELKA. Unusual Interstrand Pt(S,S-diaminocyclohexane)-GG Crosslink Formed by Rearrangement of a Classical Intrastrand Crosslink Within a DNA Duplex. *Chemistry - An Asian Journal*, WILEY-VCH VERLAG GMBH, **2010**, roč. 5, č. 2, s. 244-247. ISSN 1861-4728.

11  **KUBICEK, Karel**, S. Kaspar GRIMM, Julien ORTS, Florenz SASSE &
    Teresa CARLOMAGNO. The Tubulin-Bound Structure of the Antimitotic
    Drug Tubulysin. *Angewandte Chemie International Edition*, Weinheim,
    Germany: Wiley-VCH Verlag GmbH & Co., **2010**, roč. 49, č. 28, s.
    4809-4812. ISSN 1433-7851.

12  REESE, Marcel, Victor M SANCHEZ-PEDREGAL, **Karel KUBÍČEK**,
    Jens MEILER, Marcel J BLOMMERS, Christian GRIESINGER & Teresa
    CARLOMAGNO. Structural basis of the activity of the microtubule-
    stabilizing agent epothilone a studied by NMR spectroscopy in solution.
    *Angewandte Chemie International Edition*, Weinheim: Wiley-VCH Verlag
    GmbH & Co., **2007**, roč. 46, č. 11, s. 1864-1868. ISSN 1433-7851.

13  SANCHEZ-PEDREGAL, Victor M, **Karel KUBÍČEK**, Jens MEILER,
    Isabelle LYOTHIER, Ian PATERSON & Teresa CARLOMAGNO. The
    tubulin-bound conformation of discodermolide derived by NMR studies
    in solution supports a common pharmacophore model for epothilone and
    discodermolide. *Angewandte Chemie International Edition*, Weinheim:
    Wiley-VCH Verlag GmbH & Co., **2006**, roč. 45, č. 44, s. 7388-7394.
    ISSN 1433-7851.

14  BANCI, Lucia, Ivano BERTINI, Simone CIOFI-BAFFONI, Efthalia KAT-
    SARI, Nikolaos KATSAROS, **Karel KUBÍČEK** & Stefano MANGANI.
    A copper(I) protein possibly involved in the assembly of CuA center of
    bacterial cytochrome c oxidase. *Proceedings of the National Academy of
    Sciences of the USA*, Washington, **2005**, roč. 102, č. 11, s. 3994-3999.
    ISSN 0027-8424.

15  BANCI, Lucia, Ivano BERTINI, Isabella C FELLI, Ludwig KRIPPAHL,
    **Karel KUBÍČEK**, Jose J MOURA & Antonio ROSATO. A further
    investigation of the cytochrome b5-cytochrome c complex. *Journal of
    Biological Inorganic Chemistry*, Heidelberg: Springer Verlag, **2003**, roč.
    8, č. 7, s. 777-786. ISSN 0949-8257.