University of South Bohemia

Faculty of Science

Binding of Cyanine Fluorescent Probes to DNA

Master Thesis

Bc. Ingrid Romancová

Supervisor: Mgr. Zdeněk Chval, Ph.D.

České Budějovice 2013

Romancová I., 2013: Binding of Cyanine Fluorescent Probes to DNA. Mgr. Thesis, in English - 84 p., Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

This master thesis is focused on theoretical study of the Cy3 and Cy5 dyes and their interactions with DNA. The main aim was to find the most populated conformations of the Cy3-DNA and Cy5-DNA complexes. A comparison with the experimental structure was also done and the influence of the cyanine dyes on the conformational changes of the DNA chain was evaluated.

Prohlašuji, že svoji diplomovou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 15. 4. 2013

Podpis studenta

Aknowledgments

I thank a number of people for helpful comments on my master thesis. I thank my supervisor Mgr. Zdeněk Chval, Ph.D. for helpful advice and discussions, for the correction of this thesis. Then I would like to thank my lecturers at Charles university in Prague for their study programme Modeling *of Chemical Properties of Nano- and Biostructures*. Last but not least I would like to thank to prof. Šíp for a financial support (grant GC204/09/J010).

Content

1 Introduction

Cyanines are fluorescent dyes with many applications in life sciences, particularly in biomedical imaging. Our research was focused on cyanines' interactions with DNA due to their use in DNA-microarrays (also known as DNA chips). Microarrays consist of tens to tens of thousands microscopic spots containing single-stranded deoxyribonucleotides (capture probes) attached to a solid surface (such as a membrane, a polymer, or glass). They are used for simultaneous analysis of sample solutions containing fragments of nucleic acids. Oligonucleotides in individual spots are identical but they differ across the spots to match various complementary DNA sequences present in a given sample. The extent of binding of a complementary fragment to the surface-attached oligonucleotide is detected mainly fluorometrically using cyanine dyes.

The sequences of capture oligonucleotides should not allow formation of internal structures such as hairpins. They should be sensitive to sequence variations and bind only to complementary strands. The hybridization on an array requires a similar melting temperature for all capture probes. There are many software tools available for the probe design that are used during the process of microarray development.[1,2] However these methods are based on standard hybridization conditions, i.e. nucleic acids in solution, with no cyanine dye bound and not being attached to a surface.

In this thesis a conformational behavior of the Cy3 and Cy5 cyanine dyes attached covalently to DNA is studied. It is the first step towards answering the question: what is the influence of the attached dye on hybridization conditions? This knowledge would help us to understand better the process of hybridization on an array and to improve the design of DNA chips.

This work is a part of a wider project “Modeling of DNA-chip hybridization in real conditions” which was supported by the Czech Science Foundation (grant No. 204/09/J010).

1.1 Molecular Simulations

Molecular simulations allow us to study the properties of many-body systems. They can be branch out into two separated methods of many-particle calculations: molecular dynamics and Monte Carlo (random sampling). The main differences between the methods are based on the evolution in time and some other features.

Monte Carlo methods use the deterministic algorithm (Repeated random sampling) to give us the structural information about the system, such as conformations of a molecule.

Molecular dynamics, the method used in this thesis, is based on the solution of equations of motion for a system composed of atom particles. These equations provide information about the dynamics of the system in time. If we consider the Newton's equations of motion, then we talk about classical molecular dynamics. The another type represents the quantum molecular dynamics, because the equations of motion are solved by Schrödinger equation for nuclei.

1.1.1 Settings of the MD Simulations

In MD simulations it is very important to set good initial conditions of the simulated system.

At the first step an initial geometry of the system has to be given according to either experimental data (from NMR, X-ray diffraction etc.) or some theoretical model. In our simulations the combination of experimental and theoretical models was used (the conformation of a fluorescent dye was taken from the experiment and the structure of DNA decamer was built by the NAB program).

At the second step the initial velocities of all atoms were set up using the Maxwell-Boltzmann distribution of velocities v_{ix} of the atoms with mass m_i in given direction (in our case x) and at given temperature:

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_b T} \right)^2 \exp\left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_b T} \right]$$

Forces acting on the atoms at each step of MD simulations are calculated from equations which describe energy dependence on the values of bonds, angles, torsions and on non-bonded interactions (see 1.2.1 bellow).

The first step of MD simulation is equilibration. The aim of this step is to obtain the system in a thermodynamic equilibrium since in the initial configuration one can easily encounter problems with close van der Waals interactions between the atoms or with inappropriate electrostatic interactions.

1.1.2 The Periodic Boundary Conditions

Size of the simulated system is limited by complexity of the potential and by computer performance. Typically the simulated systems contain 10^3 - 10^5 atoms. This may be enough for isolated systems but not for more realistic models of systems in solution that would require much larger number of atoms. But even after an increase of the number of atoms in a simple box one may not obtain precise results because a large number of atoms will be at the edges of the box surrounded

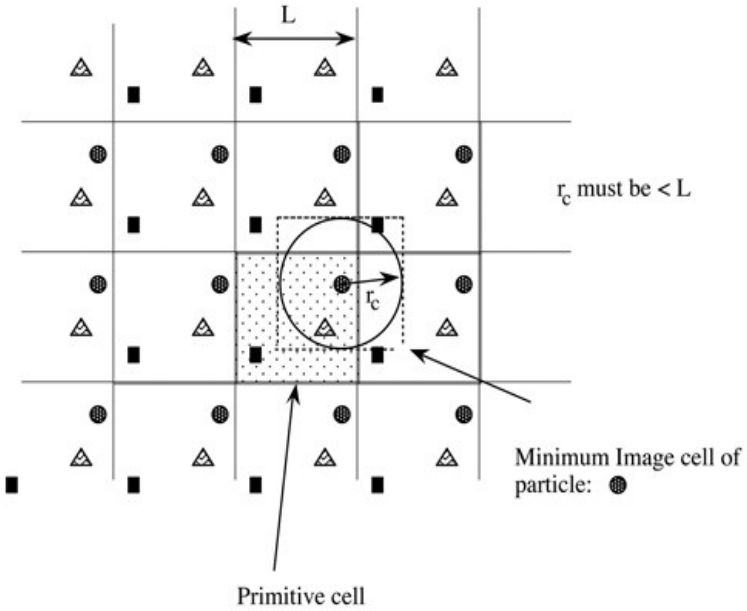


Fig. 1 *Orthogonal periodic boundary conditions - http://epress.anu.edu.au/sm/html/ch06.html*

by vacuum. To avoid this effect one can replicate this simple box in all three dimensions (see Figure 1). All particles may move freely and a constant number of particles in the box is ensured since particles that leave the box from one side in the same time enter to the same box from the opposite side. This arrangement of the simulation is called the periodic boundary conditions (PBC). The great advantage of this method is that the model is realistic, unwanted surface effects are eliminated but in the same time the system is finite not increasing a computational cost. Note that the size of the box has to be large enough to prevent the (macro)molecules to interact with its own image in the neighbor box.

1.1.3 Interaction Cutoff

Behavior of N-atoms in the central box can be described by $N^2/2$-interactions. Thus for a large number of atoms N, the calculation is computationally demanding. To reduce the number of interactions, that will be calculated at each step, an approximation called the interaction cutoff was introduced. Only interactions between the atoms whose distance r fulfill the condition $r \leq r_{max}$ ($r_{max} \leq L/2$ (L is the smallest dimension of the periodic cell)) contribute to potential $V(x, y, z)$. The interaction cutoff can be used only for short range interactions because then the error caused by the approximation is negligible (see Figure 2). The Lennard-Jones potential between two non-bonded atoms is then written in the form:

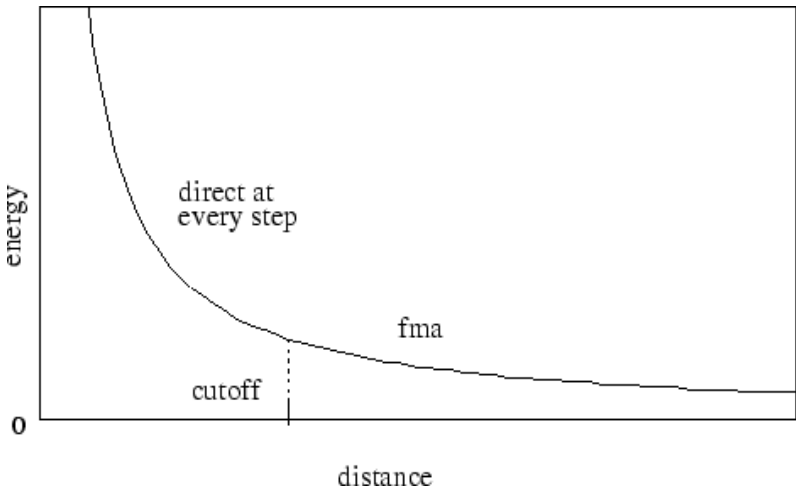


Fig. 2 *Graph showing an electrostatic potential when full electrostatics are used within NAMD, with one curve portion calculated directly and the other calculated using DPMTA, see http://www.ks.uiuc.edu/Research/namd/2.7b4/ug/node24.html*

$$V(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad for\ r \le r_{max}$$
$$V(r) = 0 \quad for\ r > r_{max}$$

The interaction cutoff may lead to larger energy fluctuations of the system. This approximation is not appropriate for electrostatic interactions due to their long range nature. The cutoff would neglect a significant part of the potential and it would cause serious errors. Instead, the Ewald sum with PBC or reaction field method can be used.

1.2 The Amber Software

Amber is a package of molecular simulation programs, the version Amber 11[3] was used for our simulations. AMBER acronym stands for "Assisted Model Building with Energy Refinement". When possible the software package uses the force field with the same name "Amber" as default.[4]

1.2.1 The Amber Force Field

In the mechanical molecular model the atoms are considered as the spheres. The interactions between chemically bound (not bound) atoms are described by special bonding (non-bonding) interaction terms. The energy of a system at a given conformation is calculated from the following equation:

$$E_{tot} = E_{str} + E_{bend} + E_{tor} + E_{non\text{-}bond}, \text{ where}$$

E_{str} ... Stretching Energy

E_{bend} … Bending Energy

E_{tor} … Torsion Energy

$E_{non\text{-}bond}$ … Non-bonded Interaction energy

Non-bonded atoms interact through van der Waals attractions, electrostatic attractions and/or repulsions, and steric repulsions.

Total energy of the system is calculated from the following functional form of the force field:

$$V(\vec{R}) = V_{bonded} - V_{nonbonded}$$

$$V_{bonded}(\vec{R}) = \sum_{bonds} \frac{1}{2} K_r (r - r_{eq})^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_{eq})_2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)]$$

$$V_{nonbonded}(\vec{R}) = \sum_{i<j} \left\{ \epsilon_{i,j} \left[\left(\frac{r_{eqi,j}}{r_{i,j}} \right)^{12} - 2 \left(\frac{r_{eqi,j}}{r_{i,j}} \right)^{6} \right] + \frac{q_i q_j}{4\pi \epsilon_0 r_{i,j}} \right\}$$

1.2.2 The General Amber Force Field

Amber force field was developed mainly for proteins and nucleic acids and thus it contains only a limited number of parameters for organic molecules. Organic molecules and fragments are usually treated by the General Amber Force Field (GAFF). It contains parameters for organic molecules that consist of H, C, N, O, S, P, and halogen atoms, and it is compatible with Amber Force Field for nucleic acids.[5]

A simple functional form is called the „Class I“ model and it is adequate to describe organic systems.

$$E_{pair} = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{v_n}{2} \times [1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right]$$

Here, r_{eq} and θ_{eq} are equilibrium structural parameters; K_r, K_θ, V_n are force constants; n is multiplicity and γ is the phase angle for the torsional angle parameters. A, B are parameters which characterize the non-bonded potentials and q_i , q_j are partial atomic charges.[5]

The GAFF parametrization was done by the following procedure:

*V*an der Waals parameters were taken from the Amber force field; partial charges were calculated by the RESP model. Force constants were estimated empirically and optimized to reproduce vibrational frequencies. Torsional angle parameters are determined to reproduce rotational energy profiles and the energy differences between various possible conformations. All these data including the structural parameters can be obtained from experiments and/or *ab-initio* calculations.

1.2.2.1 Generation of Parameters For the GAFF Force Field

Atom types

Chemical environments are usually described by atom types in molecular mechanics force fields. GAFF contains 35 basic atom types: five for carbon, eight for nitrogen, three for oxygen, five for sulfur, four for phosphorus, six for hydrogen, and atom types for halogens. [5]

For example the five carbon atom types are: c3 (sp3 carbon), c1 (sp1 carbon), c (sp2 carbon in C=O and C=S groups), ca (sp2 carbon in aromatic structures), c2 (sp2 carbon in aliphatic structures).

In addition to the basic atom types another 22 special atom types are included in GAFF being divided into four groups. [5]

Group I contains five atom types for hydrogen atoms that are attached to carbons in different environments.

Group II includes cx, cy (sp3 carbons in three- and four-membered rings) and cu, cv (sp2 carbons in three- and four-membered rings) atom types.

Group III contains cc (cd), nc (nd), and pc (pd) atom types which represent inner sp2 atoms

in conjugated ring systems; ce (cf), ne (nf), and pe (pf) represent inner sp2 atoms in conjugated chain systems; and cg (ch) describe inner sp carbons in conjugated systems.

In **Group IV**, cp (cq) represent the bridge atoms of two aromatic rings.

Charges

The default scheme of charge parametrization in GAFF includes the empirical approach based on the Gasteiger charges.[5]

Lennard-Jones Parameters

The van der Waals parameters are simply taken from Amber parm94 and parm99 force fields.[5]

Bond and Angle Parameters

There are three sources for equilibrium bond length r_{eq}: Amber force field, *ab-initio* calculations (geometries optimized by the MP2/6-31G* method), and crystal structures.[5]

Torsional Angle Parameters

The GAFF force field may not describe properly all torsional angle parameters in a specific organic system. In such cases one has to calculate the rotational profiles by *ab-initio* or DFT calculations.[5]

1.2.3 The Sander Program

Sander is a part of the Amber program package. It is able to minimize the energy of the system and perform molecular dynamics calculations. For minimization it uses the energy gradient method. Molecular dynamics is introduced by Newtonian equations of motion.

As an input the simulation program needs:

1. **Cartesian coordinates** of all atoms of the solute molecules and the surrounding environment. The initial coordinates can be obtained from X-ray crystallography, NMR spectroscopy or from theoretical models.

2. **Topology** gives information about atom and residue names, atom types, their charges and the atom connectivity. The topology of the system can be determined by the Antechamber program.

3. **Force field** parameters has to be given for all the bonds, angles, dihedral angles, and atom types in the system. For DNA we used the standard ff99bsc0 force field parameters[4]. Organic dyes Cy3 and Cy5 were described by the modified GAFF force field (see Chapter 2.2.1).

1.2.4 AmberTools

AmberTools is the program package that we used for a system preparation and for data analysis.

We used the following programs:

NAB – a molecular builder: in our case it was used for the construction of the LBDNA decamer

Antechamber – determines the atom types and atomic charges of organic molecules

ptraj – analyzes trajectories from the output of Sander; in our case we also used *ptraj* for a creation of *pdb* files from *mdcrd* or binary *binpos* files.

tleap – creates input files for Sander in the text mode
xleap – creates input files for Sander

1.3 Rigid Body Parameters

1.3.1 Base-Pair Identification

Base pairs are identified by the X3DNA program[6] which was modified to use the following restrictions for their mutual position:

- the distance between the origins of base reference frames is ≤ 15Å;
- the vertical offset (Stagger) of the base planes to values of magnitude is ≤ 2.5Å;
- the smaller of the two angles between the normals of the base planes is ≤ 65°;
- the distance between glycosidic base atoms, i.e., purine N9 and pyrimidine N1, is ≥ 4.5Å.

1.3.2 Base-pair Step Parameters

The six inter base pair parameters are *Shift*, *Slide*, *Rise*, *Roll*, *Tilt* and *Twist* (see Figure 3).

Twist, *Toll* and *Tilt* are rotational components, the three translational components are *Rise*, *Shift* and *Slide*.

Roll is the rotation of one base pair with respect to its neighbor, along the long axis of the base pair. A positive *Roll* indicates, that base pair step opens towards the minor groove, a negative roll means, that the base pair opens towards the major groove.

Tilt is the corresponding rotation of one base pair along the short axis of its neighbor.

Twist is a rotation around the helical axis.

Slide is a mutual displacement of the base pairs along their long axes.

Rise is the translation in the direction of the helical axis

Shift is a mutual displacement of the base pairs along their short axes.

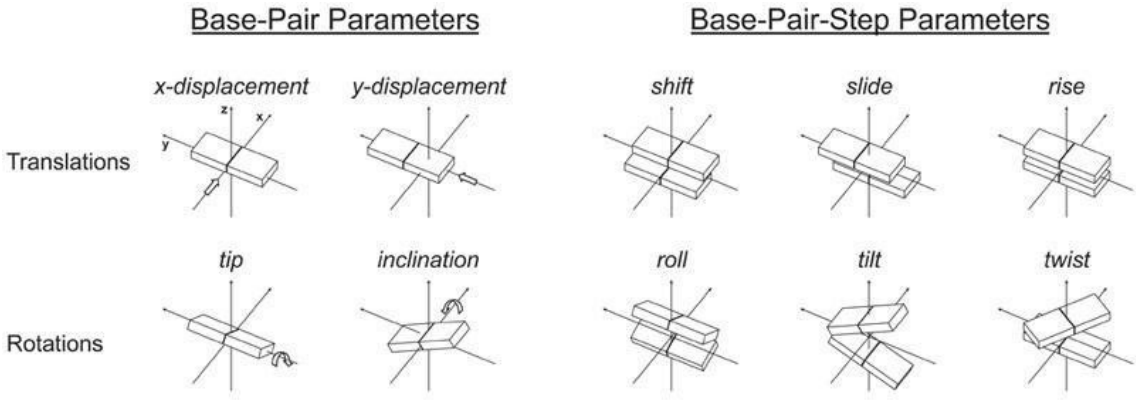


Fig. 3 *Schematic diagrams of the base-pair and base-pair-step parameters of the DNA duble helix - http://www.nature.com/nature/journal/v423/n6936/extref/nature01595-s1.jpg*

1.3.3 Base-Pair Parameters

Six base-pair parameters are commonly used to identify the geometry of a base pair: *Shear*, *Stretch*, *Stagger*, *Buckle*, *Propeller* and *Opening*. The Figure 4 illustrates these six parameters. The first group includes three translational parameters along the *x*-, *y*-, and *z*-axes: *Shear*, *Stretch* and *Stagger*. The second group contains the three rotational parameters about the *x*-, *y*-, *z*-axes: *Buckle*, *Propeller*, and *Opening*.

The major parameters are *Shear*, *Stretch* and *Opening*. They determine the relative offset of origins of the two bases in the mean base-pair plane and the angle formed by the two *x*-axes along the normal to the mean base-pair plane[7].

The minor parameters are *Stagger*, *Buckle*, and *Propeller* which mostly contribute to the fluctuations.

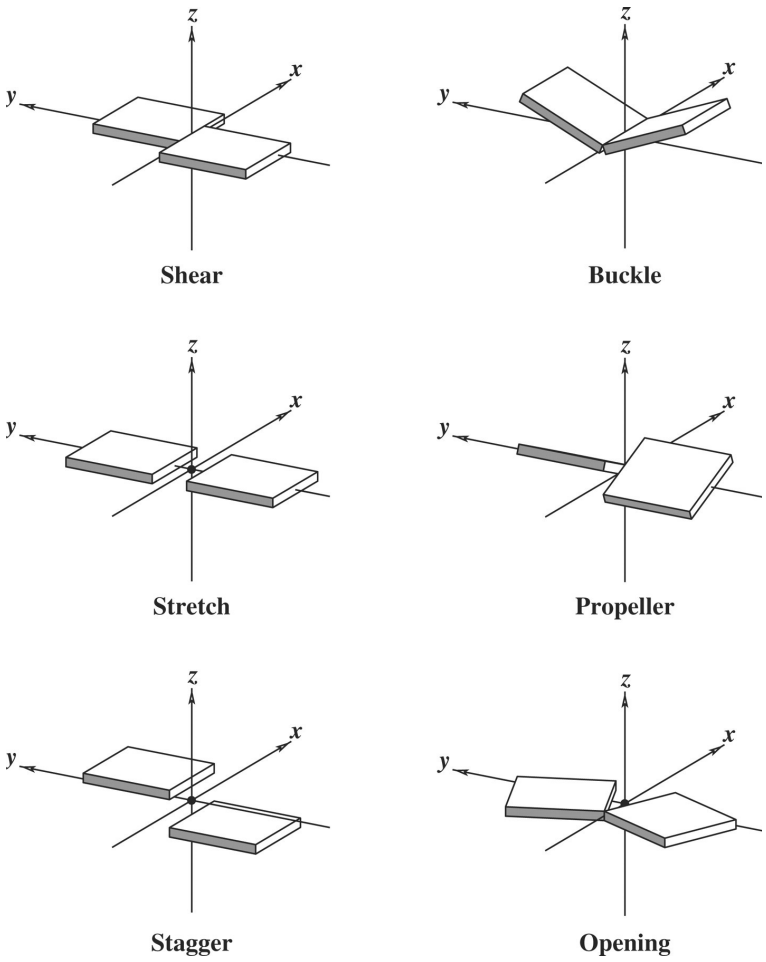


Fig. 4 *The six rigid-body parameters: three components of displacement called Shear, Stretch and Stagger and three angular parameters termed Buckle, Propeller and Opening - http://nar.oxfordjournals.org/content/37/suppl_1/D83.full*

1.3.4 Sugar Puckering and Glycosidic Torsion Angles

The five-member (deoxy)ribose ring is non-planar but has a limited conformational flexibility. The conformation of a (deoxy)ribose ring can be defined by the five endocyclic ribose torsion angles ν_0 - ν_4. The ring puckering depends on non-bonded interactions between substituents at four ring carbon atoms. The conformational space can be described by two parameters: Pseudorotation phase angle *P*, and puckering amplitude ν_M: [8,9]

$$tgP = \frac{[(\nu_4 + \nu_1) - (\nu_3 + \nu_0)]}{2\nu_2(\sin 36° + \sin 72°)} \qquad \nu_M = \frac{\nu_2}{\cos P}$$

The pseudorotation phase angle can take any value between 0° and 360°. If ν_2 is a negative value, then 180° is added to the value of P. Phase angle P correlates with the backbone torsion angle δ, C5'-C4'-C3'-O3':

$$\delta = 40° \cos(P° + 144°) + 120°$$

Several distinct (deoxy)ribose ring pucker geometries have been observed by X-ray crystallography and NMR techniques. Two atoms deviate from the ring plane, the direction of their displacement divides the ring pucker into the two families termed *endo* (the major displacement is on the same side as the base and C4'-C5' bond) and *exo* (on the opposite side). The puckers can be best described as a result of a twist conformation. [8,9]

The glycosidic bond between a (deoxy)ribose and a base is C1'-N9 bond for purines and C1'-N1 for pyrimidines. The torsion angle χ around this single bond is defined as O4'-C1'-N9-C4 for purines and O4'-C1'-N1-C2 for pyrimidines. According to the value of χ two orientations were recognized: *anti* as the most common orientation, with χ between 180° and 300°, and *syn* orientation with χ between 30° and 90°. [8,9]

1.3.5 Backbone Torsion Angles in Nucleic Acid Structures

Conformation of the phosphodiester backbone can be described by six variable torsion angles α, β, γ, δ, ε, and ζ. In polynucleotides the values of these angles are not independent but are correlated and restricted to discrete intervals as it is shown in a conformational wheel (see Figure 5). Together with glycosidic angle and sugar pucker flexibilities this leads to a large number of low-energy conformations for single stranded polynucleotides. Their number decreases substantially when the strand is the part of the double helix. [10]

All the atoms along the phosphodiester backbone are bound by single bonds. Dihedral angles can be assigned as gauche$^+$ (g^+) for the dihedral values of ~ +60°, gauche$^-$ (g^-) (for -60°), and

trans (t) (~180°). The torsion angles α (O3'-P-O5'-C5') can be found in the g^+, g^-, and t conformations. [10]

For example the two torsion angles α and ζ can be found in the tg-, g-g-, and g+g+ conformations, however, only g-g- conformation can be observed for DNA and RNA double helices.[8]

The torsion angle β (P-O5' – C5'-C4') is always in the trans conformation. The torsion angle γ (C4' – C5') can be observed in all three conformations.[10]

The last torsion angle δ (C3'-C4'-C5'-O5') is influenced by the internal torsion angle τ_3, that determines a sugar puckering. For C2'-endo puckers the τ_3 and δ angles have the values of about 35° and 150°, while for C3'-endo they adopt values around 40° and 75°, respectively.[10]

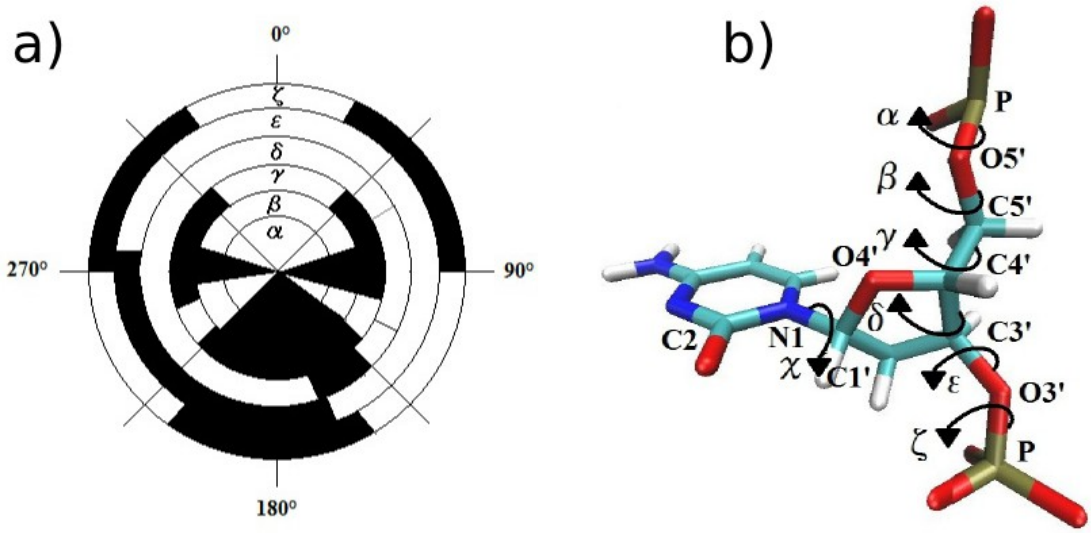


Fig. 5 a) *Conformational wheel – the allowed conformations for torsion angles for nucleosides, nucleotides, deoxyoligonucleotides, and deoxypolynucleotides.* ***b)*** *The backbone torsion angles in a unit nucleotide.*

1.4 Cyanine Dyes Used in Biotechnology

1.4.1 Cyanine Dyes

1.4.1.1 Structure of Cyanine Dyes

Cy3 and Cy5 dyes consist from two indole rings (designated as proximal and distal indole rings) which are connected by planar trimethine and pentamethine linkers, respectively. There is also a free hydroxypropyl side chain attached to the N9 atom of the distal indole ring. The dyes can be covalently connected via a three-carbon linker attached to N9' to the 5'-phosphate terminus of the DNA (see Figure 6). Both are water-soluble fluorescent dyes.

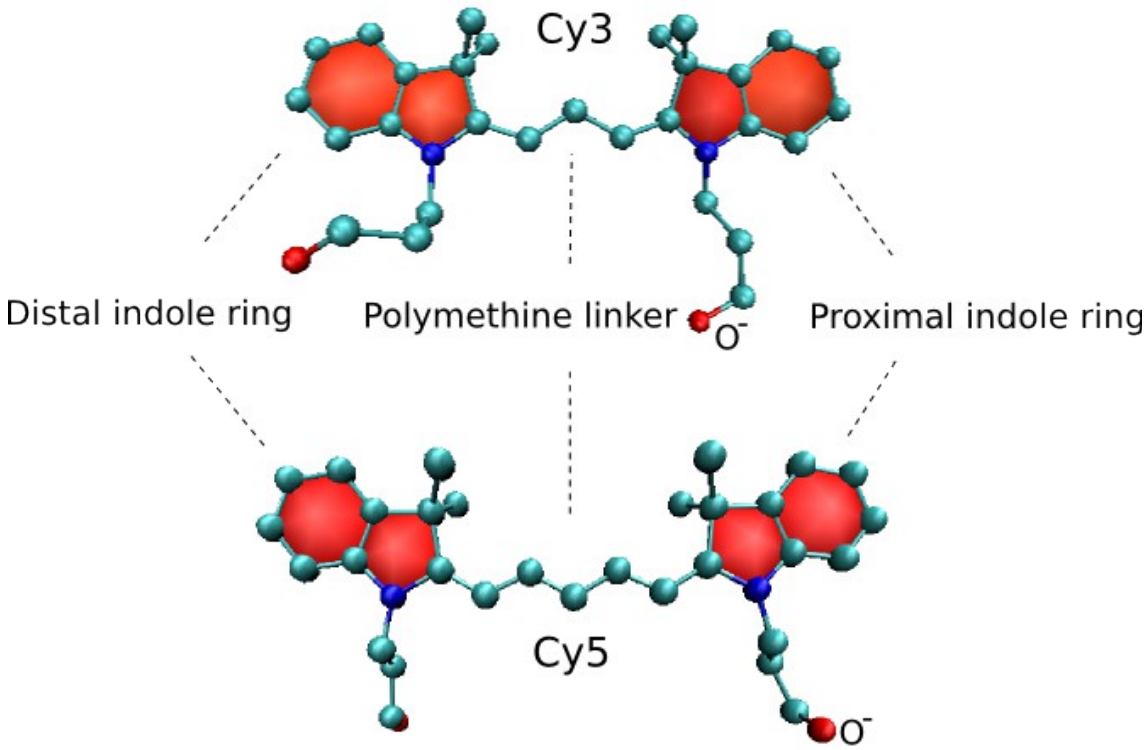


Fig. 6 *The structure of Cy3 and Cy5 dyes. Only heavy atoms are shown for clarity.*

1.4.1.2 The Fluorescence Spectra of Cyanine Dyes

Absorption and emission spectra of the cyanine dyes can be tuned by variation of the length of the polymethine chain joining the two heads of the cyanine dye.[11] Cy3 conjugates have maximum excitation at 550 nm, with peak fluorescent emission at 570 nm in the green region of the visible spectrum. Cy5 conjugates are excited maximally at 650 nm and fluoresce maximally at 670 nm in the red region of the visible spectrum. They are brighter than other fluorescent dyes and offer greater photostability (see Figure 7).

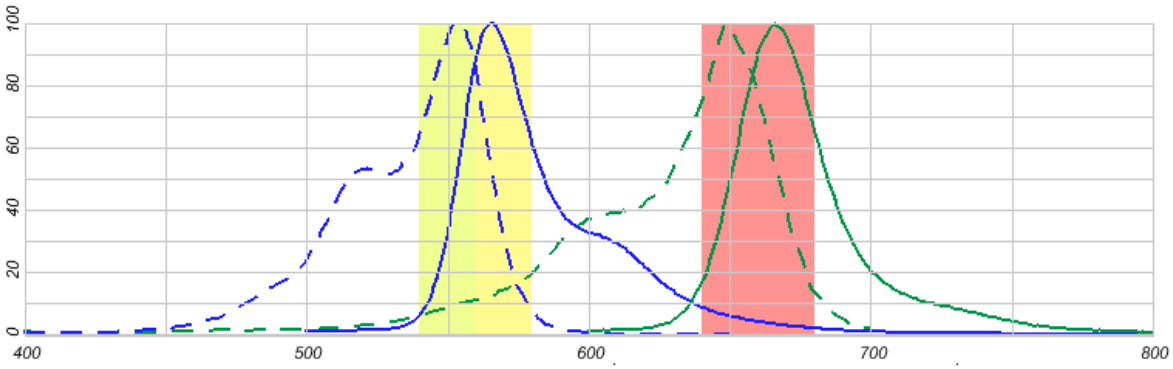


Fig. 7 *Fluorescence excitation (…) and emission (—) spectra of a Cy3 (blue) and Cy5 (green) dye-labeled IgG antibody in pH 7.2 buffer - http://www.invitrogen.com/site/us/en/home/support/Research-Tools/Fluorescence-SpectraViewer.html*

1.4.1.3 Spectral Properties of Cyanine Dyes Attached to DNA

Spectral properties of cyanine dyes are useful in many technical applications as laser disks, solar cells, fluorescent tags in DNA sequencing, fluorescent labeling agents for proteins, infra red laser dyes or dyes for polymers.[12]

These dyes change the fluorescence intensity after binding to DNA. Moreover if the cyanine dye is bonded to DNA, the absorption and emission bands are red shifted comparing to the free isolated dyes.[13] Fluorescence can be applied for understanding of DNA conformation and dynamics.

After a photon absorption, the molecule of the dye isomerizes to a non-fluorescent

photoisomer. Cy3 interacts with DNA in a sequence-dependent fashion and its fluorescent efficiency depends on the environment in which the molecule is located, on the type of the attachment and on the base pairs of DNA in the vicinity of the dye.[14]

Lifetimes of excited states of the dye depend strongly on a DNA sequence. The DNA sequence determines the nature of interactions between the dye and the DNA bases. The photoisomerization process deactivates the excited state. The efficiency of the isomerization decreases and the lifetime of the singlet excited state increases[15] when the interaction between the dye and the DNA oligomer avoids the rotation around the double bonds of the dye.

1.4.2 Techniques Used in Biotechnology

1.4.2.1 Fluorescent Resonance Energy Transfer

FRET is a non-radiative energy transfer from the excited donor fluorophore to a nearby acceptor fluorophore, which is mediated by dipole-dipole interactions.[16] FRET requires overlap between the donor emission spectra and the acceptor excitation spectra.

FRET can be used for investigation of the structures and dynamics of DNA as well as it enables a measurement of distances between the two fluorescent dyes.[17] FRET can be applied for the analysis of global conformation and folding transitions in nucleic acids.[18]

According to the FRET theory all fluorescence properties of the probes are independent on the environment.[18] However, this approximation is inaccurate because the photophysical properties of most probes depend on solution.[18] Some environmental contributions are critical.[18] Solvent properties such as temperature or viscosity cannot be neglected since they influence the rotational freedom of the linkage between the fluorophore and the DNA oligomer.[18]

The FRET theory provides us the long-range (up to 80 Å.[19]) information about a structure in

solution.[19] The distance is determined on the base of efficiency of the transfer of excitation energy from a donor to an acceptor fluorophore. This transfer is a result of the coupling between the transition dipoles of the two fluorophores.[19] The efficiency of the transfer is given by

$$E_{\mathrm{FRET}} = \frac{1}{1 + \left(\frac{R}{R_0} \right)^6}$$

where R is a distance between the fluorophores, R_0 is the Förster distance for the donor-acceptor combination at which the efficiency of energy transfer is 50 %.[19,20] R_0 depends on the relative orientation of the transition dipole vectors as well as on the spectroscopic properties of the fluorophores and the medium and it is given by [20]

$$R_0 = \frac{0.529 \cdot \kappa^2 \Phi_{\mathrm{D}} \cdot J(\lambda)}{N \cdot n^4} \quad ,$$

where κ^2 describes the relative orientation of the fluorophores, Φ_{D} is the quantum yield of the donor, N is Avogadro number, n is refraction index of the medium and $J(\lambda)$ is spectral overlap integral. The spectral overlap integral depends on the spectral shape of donor emission and the spectral shape of acceptor excitation.

The efficiency of FRET between Cy3 and Cy5 terminally attached to the 5' ends of DNA duplex strongly depends on the relative orientation of two fluorophores[20] which is dependent on the length of the duplex, the helical periodicity of DNA duplex and on the local DNA conformation. [20]

The transition moments for the $\pi \rightarrow \pi^*$ transitions lie in the plane of the indole rings and are directed in planes that are parallel to each other.[20] Orientation factor κ^2 is a very significant parameter which describes relative orientation of dipoles of the two fluorophores and its value is between 0 and 1. The border values correspond to perpendicular and parallel orientations of the dipoles of the two fluorophores. If the isotropic reorientation of

fluorophores is shorter than the excited state lifetime of the donor then it is impossible to determine experimentally the exact value of κ^2. In such a case the value $\kappa^2 = 2/3$ is usually considered which roughly corresponds to the orientation factor of the dynamically averaged dipole orientation of rotationally free chromophores.[18]

1.4.2.2 Cy3-DNA Stacking Interaction

Cy3 dye is one of the most popular fluorophores used in DNA chips. This dye has significant stability against photobleaching and is compatible with green lasers. Cyanine dye Cy5 is often used in combination with Cy3 due to its compatibility with red lasers.[18]

Despite of the high similarity of their structures, these cyanine dyes have different properties. Cy3 has a shorter polymethine chain with respect to Cy5 and that is why Cy3 is more sensitive to environmental factors such as temperature or viscosity compared to Cy5. Most frequently Cy3 is used as the donor in FRET because the efficiency of the energy transfer is calculated from donor signal. The efficiency of the energy transfer depends on the location of the dye. [18]

Norman at al. described the interaction between Cy3 and the DNA duplex by a combination of two experimental methods, NMR and fluorescence. They found that the fluorophore is attached to the end of duplex DNA by stacking interactions as an additional base pair. [18]

Cy5 interacted similarly also being stacked onto the end of the helix. These results were confirmed by another NMR study which revealed that Cy5 stacks on the top of the first base pair with similar stacking interactions that occur between Cy3 and free nucleotides in solution.[19]

The cyanine dye could be attached to the duplex of DNA by tethers of various lengths. It was described in many studies that the increasing length of the tethers changes the mutual orientation of the dye and DNA but it does not lead to a higher rotational freedom.[21,22] On the other hand, Cy5 may bind to DNA in two different conformations and their preference is

influenced by the tethers. Conformational changes influence dye-to-dye distances in FRET experiments.[18]

Stacking interactions impede the photoisomerization of the dye to a nonfluorescent *cis* isomer. The all *trans* conformation corresponds to the ground state. When the absorption takes place, the singlet-excited state is deactivated to the ground state by many different processes which include fluorescent emission, internal conversion or photoisomerization: a rotation around the C-C bond of the polymethine chain.[19]

The fluorescent lifetime and quantum yield of Cy3, the parameters for FRET efficiency calculations, depend on temperature and steric interactions that are able to reduce a speed of isomerization. The isomerization competes with the fluorescent emission.[19] From a comparison of activation energies for isomerization one can deduce that the affinity of Cy3 towards to a T-A base pair is significantly smaller then for the C-G base pairs.[19]

1.4.2.3 Photophysics of Cyanine Dyes Linked to DNA

Fluorescence techniques are frequently used for the investigation of the DNA structure and dynamics. Fluorophores are often attached to the 5´ terminus of nucleic acids because the longest lifetime of fluorescence was observed for the fluorophore – 5' ssDNA complexes being more than 10 times longer than the fluorescence lifetime of the free dye in solution. This is interpreted as a result of the *cis-trans* isomerization reaction from the first excited state. [20] The activation energy for photoisomerization depends strongly on the rigidity of the microenvironment in which the dye is located.[20]

Environmental effects also influence the fluorescence properties. For instance, the dependence of fluorescence anisotropy on temperature and viscosity can be used to study rotational dynamics of the molecule to which the fluorophore is attached. Fluorescence anisotropy characterizes the rotational freedom of the linkage between the fluorophore and DNA.[20]

1.4.2.4 Time-Resolved Single Molecule Fluorescence Spectroscopy of Cy5

Time-resolved single molecule fluorescence spectroscopy is one of the useful tools for better understanding of biological processes – time dependent dynamic studies and single molecule polarization techniques are other used methods. The long observation time requires the immobilized molecules. The effect of immobilization methods on the biological systems and their photophysical properties was analyzed in many studies.[23]

Cy5 dye is one of the most utilized probes for single molecule fluorescence spectroscopy. This carbocyanine derivative is used for investigating the conformational dynamics of large biomolecules and forms a suitable FRET pair with other single-molecule probes like tetramethyl rhodamine and Cy3.[23]

Photophysics of Cy5 dye was analyzed in many studies. The deactivation mechanism, triplet state formation and *cis-trans* isomerism revealed unexpected results, such as reverse intersystem crossing, delayed fluorescence and millisecond blinking. It is important to reduce Cy5 blinking and photobleaching, if the Cy5 dye is used as a probe for single molecule fluorescence spectroscopy.[23]

1.4.2.4.1 Delayed Fluorescence of the Cy5 Dye

The fluorescence and intersystem crossing properties compete with the possible isomerization by carbon-carbon bond rotation without any involvement of the triplet states in the formation of the isomers. The delayed fluorescence is generated from the S1 state of *trans*-Cy5 through an intersystem crossing from the *cis*-triplet state T1 to the trans-singlet state S1 via thermal activation. The back-isomerization of cyanine dyes from the *cis*-T1 state to *trans*-S1 state that leads to the unusual delayed fluorescence. [24]

1.4.2.4.2 Characterization of Photoinduced Isomerization and Intersystem Crossing of the Cyanine Dye Cy3

Single-molecule fluorescence experiments reveal important photophysical properties of cyanine dyes related to the *cis-trans* isomerization and triplet-state formation. [25]

For instance, it is unexpected that the conversion of Cy5 from the dark state to fluorescence state is dramatically enhanced in the presence of Cy3. The conventional Förster resonance energy transfer theory is not able to explain this observation. On the other hand, the dark states of Cy5 are able to quench the fluorescence of the donor Cy3. Reaction between an excited Cy3 and blinked Cy5 is necessary to bring back Cy5 from blinking. Cy3 and Cy5 are often used as a donor-acceptor pair for FRET measurement. A longer polymethine chain of Cy5 may lead to different competition behaviors between isomerization and intersystem crossing (ISC). [25]

The photophysical properties of Cy3 with respect to triplet state and isomerization revealed that the triplet-state absorption (T1-T*n*) of *trans*-Cy3 and the ground state absorption of the photoisomer (*cis*-Cy3) of Cy3 largely overlap each other around 580 nm with broad absorption region. [25]

1.4.2.4.3 Molecular Photonic Switches

Molecular photonic switches and optical data storage elements also depend on the fluorescent dyes. The well-known quantum processes such as quenching, energy transfer via intersystem crossing can change the fluorescent intensity. If energy fluctuations can be controlled then the molecular photonic switching can be achieved. The switching occurs between nonfluorescent and fluorescent states. For instance we are able to control the emission from the excited singlet state of the chromophore via photoinduced electron transfer.[20]

1.5 DNA Microarrays

DNA microarrays have become a handy tool in molecular biology. They consist of hundreds up to tens of thousands microscopic spots of DNA single stranded oligonucleotides attached to a solid surface such as glass. Each spot contains multiple copies of a unique DNA sequence (probes) which corresponds to a single gene sequence. These oligonucleotides then interact with solution containing fragments of nucleic acids known as targets. Target sequences than hybridize with probe sequences in case these sequences are complementary. The extent of binding is detected mainly fluorometrically.[26,27]

Scientists are using DNA microarrays to investigate a wide range of problems from cancer to pest control. One common use of DNA chips is the determination of active and suppressed genes in comparisons of two populations of cells. Identification of gene activity changes during a gene expression is important for the understanding of toxicological processes. For example from the differences in gene activity between healthy and cancer cells it is possible to determine genes in cancer cells which are working abnormally.[26,27]

1.5.1 Manufacturing of DNA Microarrays

Two main approaches exist in the manufacturing of DNA microarrays – *in situ* synthesis of oligonucleotides on a solid substrate or a physical deposition of presynthesized DNA onto a solid substrate (spotted arrays).[26]

The first approach involves a direct synthesis of the nucleotides on the solid surface. The oligonucleotide arrays are produced by the photolithographic method pioneered by Fodor et al.[28]. This method allows manufacturing of high-density DNA microarrays (> 250 000 oligonucleotide spots per cm^2) which are especially suitable for gene expression analysis. However, photolithografic masks are too expensive for the fabrication of custom DNA arrays.[26]

The second approach, the physical deposition of presynthesized DNA on a solid substrate, is

more flexible but requires labor-intensive preparation.[26]

Negatively charged DNA fragments can be bound to glass-surfaces often coated by a positively charged layer of (bio)molecules such as poly-L-lysine.[26]

DNA microarrays techniques are evolving. The requirements for increasing spot density, improved reproducibility and reduced production time and price gave rise to a development of new screening instruments, such as the Maskless Array Synthesizer (MAS). Here a digital light processor (DLP) enables the gene expression analysis using oligonucleotide arrays produced by a maskless photolithography.[26]

1.5.2 Microarray Design

To construct a DNA microarray, several procedures have to be combined (such as immobilization, or fluorescence).[26]

The first step is to decide what question has to be answered, what kind of genes will be investigated. For diagnostic analysis the smaller microarrays with known target sequences are used while for studies of metabolic pathways large microarrays with unknown target sequences are used. [26]

The next step is to select DNA capture probes, from which the genetic information can be obtained. Then the selected capture probes are immobilized on a solid substrate. Commonly used are glass slides because the glass surface can be easily modified by silanes. For immobilization several methods can be performed, such as microspotting, inkjetting, electronic addressing, and photolithography.[26]

Finally the target probes from biological samples are prepared and hybridization is performed. The signal from the hybridization reactions can be obtained using several methods, such as fluorescence based methods, electrochemical methods, mass spectroscopy, and radiolabeling.[26]

1.5.3 Hybridization

Before hybridization the mRNA strands are labeled by a fluorescent dye and modified nucleotides are incorporated during the cDNA synthesis. The most popular fluorescent dyes are Cy3, Cy5, fluorescein and rhodamin. They are chosen with respect to their resistance to photobleaching, excitation and emission peak, and other parameters.[26]

The hybridization conditions are influenced by size of the DNA fragments attached to a microarray and must be determined before every experiment.

The hybridization signal has a linear dependence on the RNA expression level if the target DNA is present on array in at least a ten-fold excess. However, the two to three fold difference is still detectable.[26]

Microarrays readers can be divided into CCD cameras, non-confocal laser scanners, and confocal laser scanners. [26]

2 Methods

2.1 Starting Structures Preparation

Starting geometry of canonical B-DNA (LBDNA) was generated by the NAB program. The NAB program is one of the programs from the AmberTools package that builds DNA chains in a given conformation. The sequence of the LBDNA was identical to that sent by dr. Norman – CCACTAGTGG (see below).

Then Cy3 or Cy5 dyes were attached to LBDNA. The dye was attached to the LBDNA in exactly the same conformation as in the 'experimental' structure which had been sent to us by dr. David G. Norman from The University of Dundee after request. This structure corresponds to an average structure from 19 structures which were calculated using restrained molecular mechanics and dynamics. The distance restraints were applied in these simulations to force the system to be in agreement with experimental data obtained by NMR and FRET methods.[29]

The first C---G (DG_20-DC_1) base pair with the attached fluorescent dye (Figures 18 and 22) were taken from the experimental structure. Then this structure was uploaded into the Sirius program together with the structure of the LBDNA decamer. The first C---G base pair of LBDNA decamer was aligned with the G---C pair attached to the fluorescent dye and then the first C---G base pair of the DNA decamer was deleted. Finally, we saved coordinates into a new *pdb* file containing fluorescent dye attached to the DNA decamer.

2.2 Atom Types and Charge Derivation

For molecular simulation we need to know atom types and charges of every atom in the simulated system. The Amber program is able to distinguish the atoms of DNA chains and determine their atom types and charges. On the other hand, the fluorescent dyes are non-standard residues and the Amber program needs the library files which contain their atom

names, types and charges.

This problem can be solved by the Anterchamber program that is suited to determine the atom types and charges for organic molecules. In our case the Anterchamber program was used only for determination of the atom types and the charges were calculated using the method called the Restrained Electrostatic Potential (RESP) method. The detailed procedure is described in Appendix 1.

2.2.1 Dihedral Angles of the Methidine Chains

The molecules of Cy3 and Cy5 are symmetrical with respect to the central atom of the methidine chain. In the GAFF force field, the linear chain conjugated systems are defined as a sequence of alternating single and double bonds. However for the methidine chains of Cy3 and Cy5 dyes with the even number of conjugated bonds it is not possible to ensure a symmetrical arrangement of the double and single bonds with respect to the central atom. Therefore the dihedral angles of the methidine chains were re-parametrized and new atom types were introduced as depicted in Figure 8.

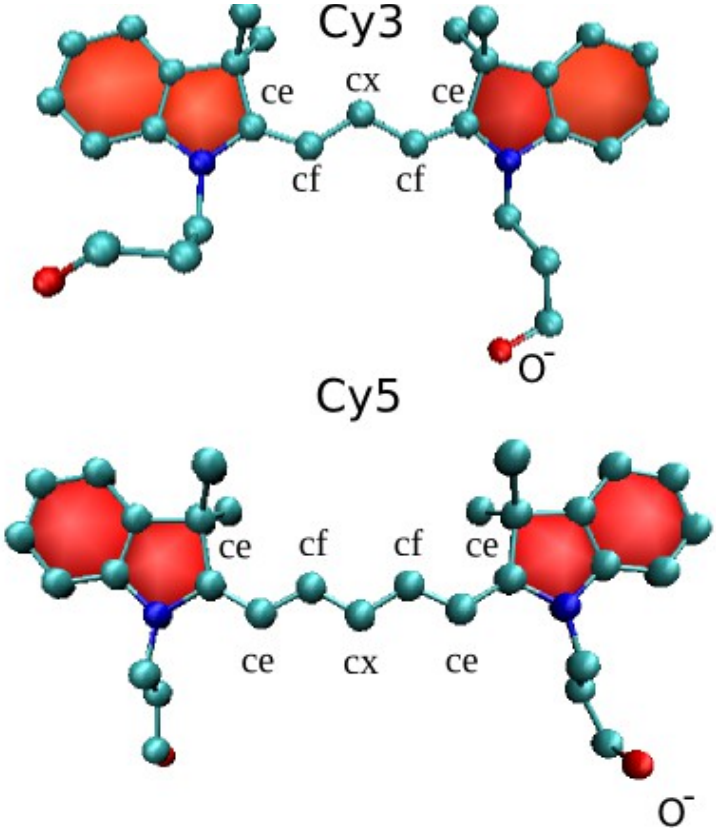


Fig. 8 *The dihedral angles of the methidine chains and new atom types cx.*

Dihedral flexibility of the methidine chains was investigated by the MP4(SD)/6-31G*//B3LYP/6-31+G** method which included restrained optimizations by the B3LYP/6-31+G** method with subsequent MP4(SD)/6-31G* single point energy calculations. All these calculations were performed with the Gaussian 09 program package[30]. The torsional profiles of the dihedral angles were calculated performing the optimizations without any restrictions except the investigated dihedral angles which were sampled in the whole range of angles <-180°; +180°> with the step size of 20°.

The torsional MP4(SD)/6-31G* profiles were then fitted by a linear combination of the cosine functions to reproduce the heights of the rotational barriers.

In the Amber parameter file one has to define the torsional parameters IDIVF, PK, PN a PHASE which are defined by the formula:

$$\mathbf{E_{tors} = (PK / IDIVF) * (1 + cos (PN * PHI - PHASE))}$$

IDIVF … the total number of torsions about a single bond that the potential applies to

PK … is equal to one-half of the barrier magnitude

PN … periodicity, the number of potential barriers as the C-C bond is rotated -180 to 180 degrees

PHASE … PHASE = 0 degrees if an energy *maximum* is at 0 degrees; PHASE = 180 degrees if an energy *minimum* is at 0 degrees.

In Amber the dihedral angle can be defined as generic (X – CT – CT – X) or explicit (HC – CT – CT – HC). For explicit definition the IDIVF value is equal to 1 and the total potential is divided by the number of torsions involved.

2.3 Starting Structures for Molecular Simulations

The molecular dynamics simulations were performed by the Sander program. Sander is a simulation program that needs information about the Cartesian coordinates of each atom in the structure, their velocities and topologies and also force field parameters and periodic box dimensions (see Chapter 1.2.2).

This information is included in the *prmtop* and *inpcrd* input files. (or a *restrt* file from a previous run). The *prmtop* file contains a description of the molecular topology and the force field parameters. The *inpcrd* file (or *restrt*) contains a description of the atom coordinates and optionally velocities and current periodic box dimensions.

For DNA chains, the *prmtop* and *inpcrd* files can be built directly using the xleap or tleap AmberTools programs. For non-standard residues such as fluorescent dyes it is necessary to create *frcmod* and *lib* files which contain custom parameters.

The *frcmod* file is required to define the mass and vdW parameters for new atom types of non-standard residues. It also provides all the bond, angle and dihedral parameters that are not defined in the standard ff99bsc0 force field. The *frcmod* files were generated using the

Antechamber program. We used the GAFF force field for all non-standard residues. Parameters for dihedral angles of the central methine linker were derived to reproduce rotational barriers calculated by the MP4/6-31G*//B3LYP/6-31+G* calculations.

The library file is required to define atom types and their charges. The Antechamber program was also used to create library files but the default atomic charges were manually replaced by the atomic charges calculated by the RESP method.

After that the structures were uploaded into xleap implemented in the AMBER 10 program package, the complex was solvated with explicit TIP3P water molecules (4047 for Cy3 dye, 3504 for Cy5 dye). To ensure periodic boundary conditions the structures were immersed in water solvent truncated octahedron box (8 angstrom buffer of TIP3P water).

In the next step nineteen Na^+ ions and then one Cl^- ion were added to the system (using Coulombic potential on a 1.0 angstrom grid at the places with the extreme electrostatic potential) to compensate charged groups in the Cy3-(Cy5-) DNA system. Then the *prmtop* and *inpcrd* files were created.

2.4 Minimization, Equilibration and Production Run

Molecular simulations consist of three basic steps: minimization, equilibration and production run. All calculations were performed in parallel on eight processors to save the time using periodic boundary conditions.

2.4.1 Minimization

The starting system does not correspond to a minimum on the potential energy surface within the force field used. Before starting molecular dynamics calculations, the structure has to be minimized. If the structure is not minimized, it could lead to crashes during the MD run.

The minimization was performed in three steps:

In the first step the DNA decamer, fluorescent dye and counter ions were kept fixed and only water molecules were free to move. We performed 500 steps with the steepest descent algorithm followed by equal number of steps of conjugate gradient minimization. A value for non-bonded cut off was set to 10 angstrom.

In the second step we released the counter ions still holding the DNA decamer with fluorescein dye fixed. The minimization algorithm was identical as that in the previous stage.

In the third step the entire system was minimized. We ran 2, 500 steps of minimization without the restraints, for the first 1,000 steps the steepest descent algorithm was used.

2.4.2 Equilibration

During the equilibration stage our system was heated up from 0 K to 300 K. This stage lasted 1 ns.

Temperature regulation was performed using a weak-coupling ensemble that roughly interpolates between the microcanonical and canonical ensembles. This thermostat is also called as the Berendsen thermostat. The bonds involving hydrogen atoms were constrained by the SHAKE algorithm that removed their bond stretching freedom and enabled prolongation of the time step up to 2 fs. The Particle Mesh Ewald method was used to treat long-range electrostatic interactions, while for non-bonded interactions the cut off of 9 angstroms was utilized. A relaxation time of 5 ps was used.

The equilibration was divided into four stages, each 250 ps long.

In the first stage we hold the DNA decamer, fluorescent dye and counter ions fixed. The system was heated up to 300 K. In the second stage the counter ions were released and temperature was kept constant at 300 K in the beginning and then the system was cooled

down to 0 K. Finally, all restrictions were released and the system was heated up to 300 K.

2.4.3 Production Run

While equilibration protocol used NVT ensemble the production run was performed using an NPT ensemble at 298 K and a pressure of 1 atm. All other settings of the calculations were the same as in the equilibration run.

Timescale of MD simulations is currently in the range of tens up to hundreds of nanoseconds. A correct determination of a timescale of MD simulations is very important because the structure may undergo many conformational changes in time. In our case the total time of the production run was 150 ns.

2.5 Analysis of the Results

There was a number of properties that we have monitored:

1. RMSd of the whole system and then RMSd of each base with respect to the initial structure
2. Base-pair step parameters – *Shift*, *Slide*, *Rise*, *Roll*, *Tilt*, and *Twist*
3. Base-pair parameters – *Shear*, *Stretch*, *Stagger*, *Buckle*, *Propeller*, and *Opening*
4. Puckering of the ribose rings including the six nucleotide torsion angles: α, β, γ, δ, ε, ζ, then the glycosidic torsion angle χ and the five internal sugar torsions ν_0 - ν_4
5. Dihedral angles of the linker between the DNA chain and the fluorescent dye

Before that we had to re-image our trajectory files. Because the simulations were performed under the periodic boundary conditions, the system of interest may move partly to a neighbor box and its coordinates are arbitrarily shifted. Therefore we used the *ptraj* program to re-image our trajectory so that both strings of DNA chain are centered in the primary box and

coordinates in different times can be compared with each other. The water molecules as well as the counter ions were usually not considered in our analysis.

For analysis we used the average structures computed over time intervals of 100 ps, i.e. 1500 structures in total. All these structures were analyzed by the X3DNA program. However, the fluorescent dyes are not recognized by X3DNA program being focused only on nucleotide bases. To analyze relative position of the fluorescent dye with respect to DNA we converted it to a A---A base pair by the algorithm defined in the next chapter.

2.5.1 Conversion of the Cy3/Cy5 Dye to the Adenine-Adenine Base Pair

X3DNA is a comprehensive software package for the analysis of nucleic acid structures. Moreover it allows the exact reconstruction of the structure from a set of parameters such as shift or twist. However the use of the program is restricted to nucleic acids only. It is not able to analyze geometry of other molecules such as the organic dyes bound to the DNA chain.

This problem can be solved by suitable geometric conversion of the organic dye to a nucleic acid base-pair since we are interested in the relative position of the planar fluorescent dye with respect to the last base pair and this can be evaluated by the same structural parameters as it is done for two adjacent base pairs.

The Cy3/Cy5 dyes contain two indole rings (also called proximal and distal rings) which are joined by a polymethine chain. For both cyanine dyes we used the same methods of conversions.

We decided to convert the two indole rings of the dyes to an artificial A-A base-pair which can be analyzed by X3DNA program (see Figure 9).

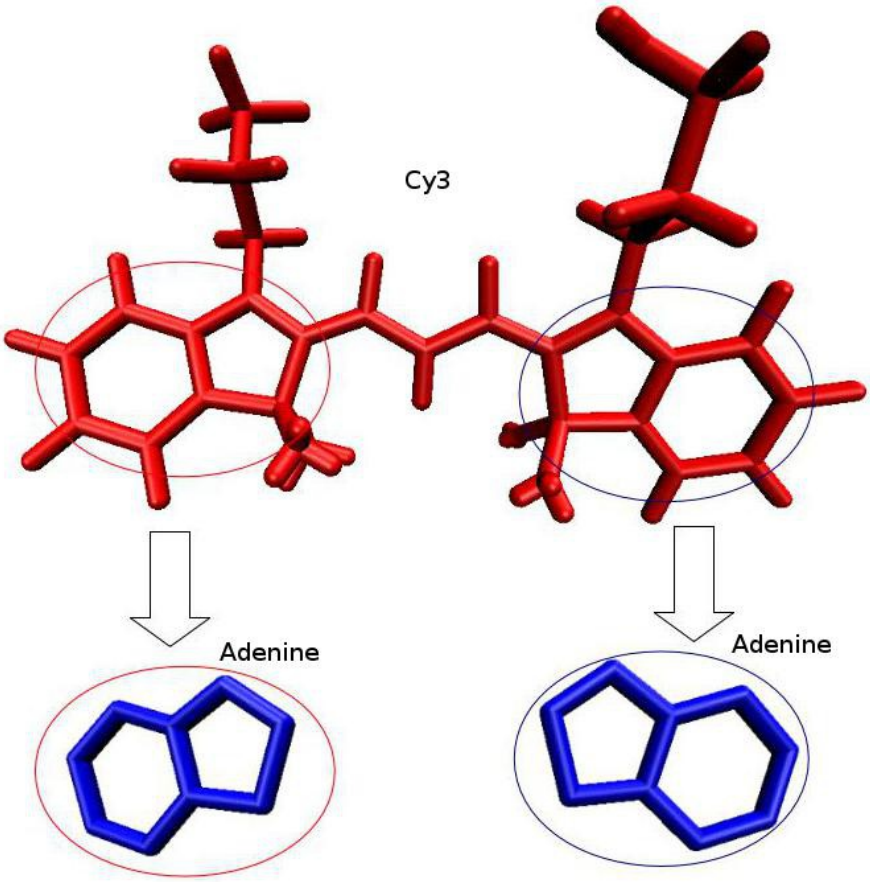


Fig. 9 *The figure shows how the two indole rings of Cy3 dye are converted to the aromatic rings of the A-A pair. This method keeps the original angle and base-pair step parameters of the indole rings because all the atoms of the two adenine rings are in the same positions as the the atoms in the indole rings. The ring atoms of the two indole rings (red and blue circles) were simply renamed to ring atoms of the two adenines. See Appendix 2 for the Fortran 95 code.*

Finally, the results for Cy3 and Cy5 dyes can be only partly compared to each other. The main reason is that the structures differ in the length of polymethine linkers and the proximal indole ring and its three-carbon linker are displaced outward from the center of DNA decamer.

3 Results

3.1 Atomic Charges of the Cy3 and Cy5 Residues

The charges on the phosphate linker groups (atoms P70, O71, O72, and O73 of Cy3; P28, O29, O38, and O40 of Cy5, see Figures 10 and 11) were kept the same as for the phosphate group in the ff99bsc0 force field. The charges of Cy3 and Cy5 dyes were calculated by a standard method which included B3LYP/6-31G* optimization of the structures with the subsequent RESP fitting of the charges which reproduce electrostatic potential of the molecule calculated at the HF/6-31* level.[31]

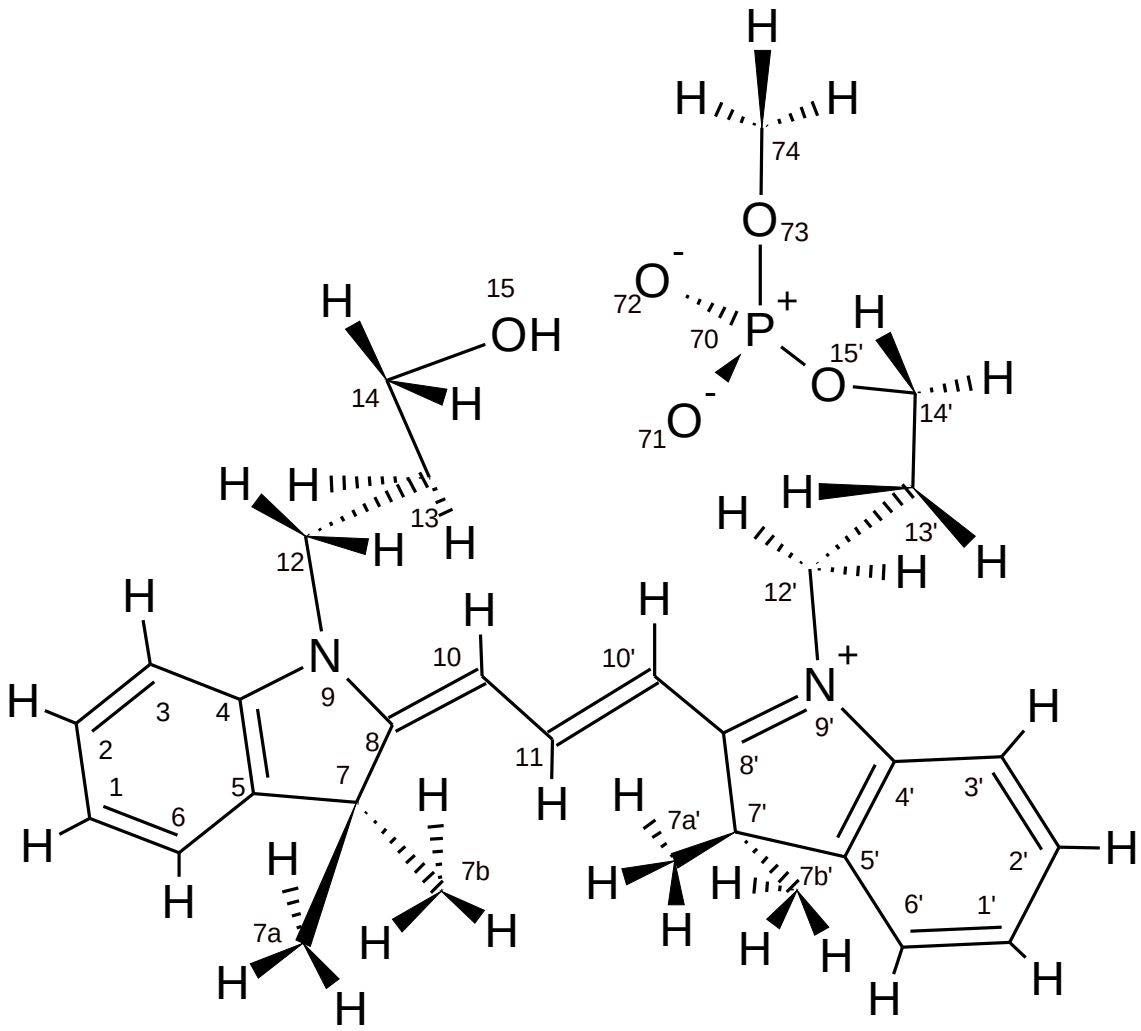


Fig. 10 *The atom numbering in the Cy3 dye.*

Comparing charges of Cy3 and Cy5 dyes the charges of most equivalent atoms on the two indole rings are similar but some differences can be found in Cy5 due to different type of a linker attached to the two indole rings (Tables 1 and 2). For example charges of N9 and N9' atoms of indole rings in Cy5 differ substantially, the N9 atom being more positive, while in Cy3 these charges are similar. The similar situation occurs for the C8, C8' atoms of Cy5. The charge of C8 atom is more negative (by 0.2e) than the charge of C8' atom.

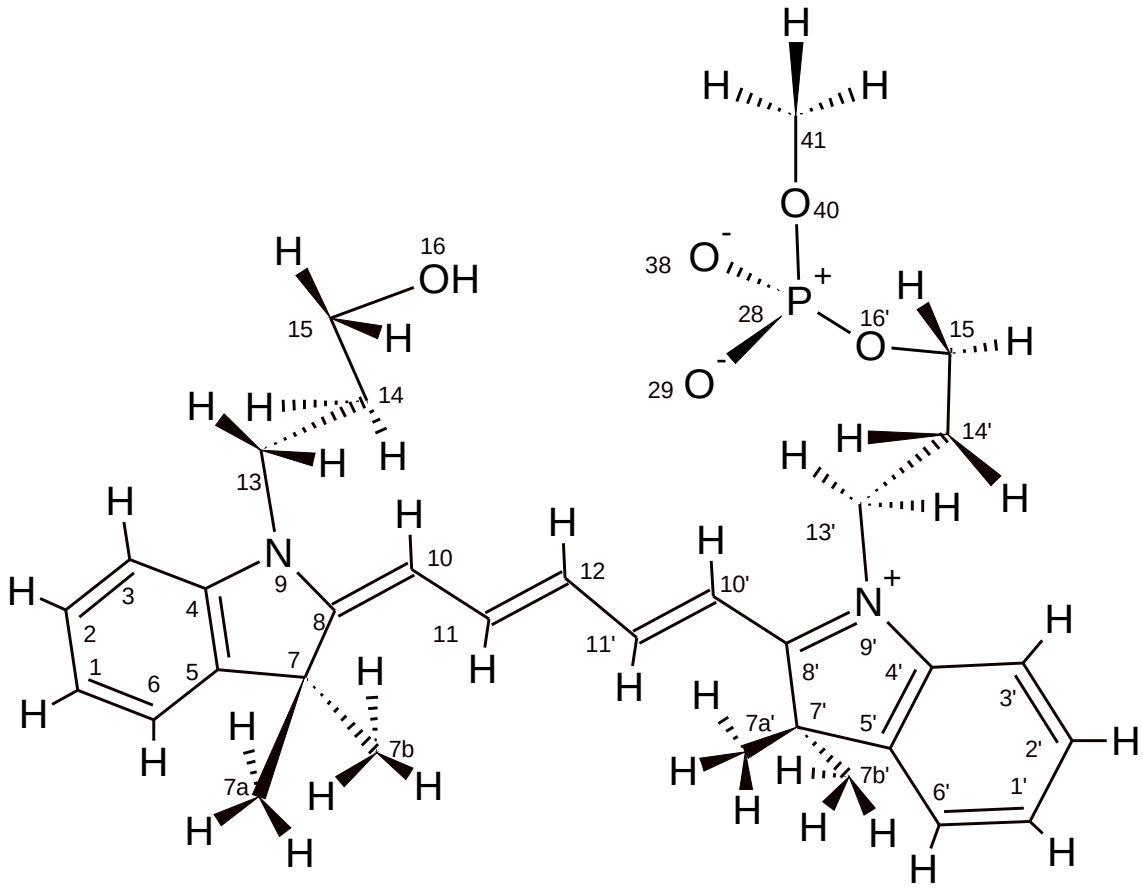


Fig. 11 *The atom numbering in the Cy5 dye.*

Nb.	Element	Charge	Nb.	Element	Charge
1	C	-0.164028	1'	C	-0.162437
2	C	-0.107608	2'	C	-0.111121
3	C	-0.251992	3'	C	-0.258169
4	C	0.195055	4'	C	0.182945
5	C	-0.056256	5'	C	-0.063240
6	C	-0.186790	6'	C	-0.184905
7	C	0.288054	7'	C	0.315676
7a	C	-0.371112	7a'	C	-0.323167
7b	C	-0.371112	7b'	C	-0.323167
8	C	0.354928	8'	C	0.314423
9	N	-0.205290	9'	N	-0.227430
10	C	-0.541782	10'	C	-0.570276
11	C	0.170547			
12	C	-0.079256	12'	C	0.082016
13	C	0.036298	13'	C	-0.221697
14	C	0.148386	14'	C	0.215445
15	O	-0.731253			

Table 1 *The atomic charges of Cy3 dye.*

Nb.	Element	Charge	Nb.	Element	Charge
1	C	-0.156243	1'	C	-0.136225
2	C	-0.105812	2'	C	-0.164432
3	C	-0.256228	3'	C	-0.048610
4	C	0.154646	4'	C	0.105396
5	C	-0.070742	5'	C	-0.064495
6	C	-0.178389	6'	C	-0.212922
7	C	0.343498	7'	C	0.386898
7a	C	-0.393179	7a'	C	-0.455631
7b	C	-0.388929	7b'	C	-0.455631
8	C	0.212182	8'	C	0.391679
9	N	-0.089779	9'	N	-0.262369
10	C	-0.460748	10'	C	-0.554502
11	C	0.077426	11'	C	0.131027
12	C	-0.354824			
13	C	-0.248034	13'	C	0.109060
14	C	0.042052	14'	C	-0.248439
15	C	0.313286	15'	C	0.365299
16	O	-0.707402	16'	O	-0.432105

Table 2 *The atomic charges of Cy5 dye.*

3.2 Atom Types

Atom types of Cy3 and Cy5 dyes were determined by the Antechamber program (see Table 4 and 5). The analogous atoms have the same atom type. DNA decamer is attached to a fluorescent dye by a linkage which consists of the three carbon atoms and the phoshodiester group. These three carbon atoms are sp3 hybrids and the atom types of phoshodiester linkage were defined in accordance with ff99bsc0 force field. Pseudoether group contains the same sp3 carbons as hydrocarbon linker attached to the 5'-phosphate terminus of the DNA. The GAFF force field was used for all atoms of Cy3 and Cy5 dyes with exception of the polymethine chain whose dihedral angle parameters were calculated since the default GAFF parametrization of the linker would lead to the carbon chain with alternating ce/cf atom types. However this parametrization would not enable the analogous atoms to have the same atom type. Therefore the central carbon in the polymethine linker was defined by a new atom type 'cx'. The bond cx-cf stretching parameter was set as an average value of the ce-ce and ce-cf stretching modes while cf-cx-cf and cx-cf-ce angle parameters were set to be equal to the ce-ce-cf (or ce-cf-cf) value. The dihedral parameters c3-ce-cf-cx, ce-cf-cx-cf, c3-ce-ce-cf, and ce-ce-cf-cx were derived to reproduce rotational barriers calculated by quantum chemical calculations (see Figures 12-17) using the formula:

$$E_{tors} = \frac{PK1 - PK1_{corr}}{IDIVF1} * (1 + \cos(PN1 * PHI - PHASE1)) + \\ + \frac{PK2}{IDIVF2} * (1 + \cos(PN2 * PHI - PHASE2)) \quad ,$$

where all torsional parameters are summarized in Table 3.

Dye	Bond	PK1	IDIVF1	PN1	PHASE	PK2	IDIVF2	PN2	PHASE2
Cy3	c3-ce-cf-cx	5.90	1	2	180.00	0.91	1	1	180.00
Cy3	ce-cf-cx-cf	7.85	1	2	180.00	3.50	1	1	0.00
Cy5	c3-ce-ce-cf	4.71	1	2	180.00	1.28	1	1	180.00
Cy5	ce-ce-cf-cx	7.55	1	2	180.00	3.25	1	1	0.00
Cy5	ce-cf-cx-cf	8.20	1	2	180.00	1.72	1	1	0.00

Table 3 *The torsional parameters IDIVF, PK, PN a PHASE.*

Nb.	Element	Atom type	Nb.	Element	Atom type
1	C	CA	1'	C	CA
2	C	CA	2'	C	CA
3	C	CA	3'	C	CA
4	C	CA	4'	C	CA
5	C	CA	5'	C	CA
6	C	CA	6'	C	CA
7	C	C3	7'	C	C3
7a	C	C3	7a'	C	C3
7b	C	C3	7b'	C	C3
8	C	CE	8'	C	CE
9	N	NH	9'	N	NH
10	C	CF	10'	C	CF
			11'	C	CX
12	C	C3	12'	C	C3
13	C	C3	13'	C	C3
14	C	C3	14'	C	C3
15	O	OH			

Table 4 *The atom types of the fluorescent dye Cy3*

Nb.	Element	A. types	Nb.	Element	Charges
1	C	CA	1'	C	CA
2	C	CA	2'	C	CA
3	C	CA	3'	C	CA
4	C	CA	4'	C	CA
5	C	CA	5'	C	CA
6	C	CA	6'	C	CA
7	C	C3	7'	C	C3
7a	C	C3	7a'	C	C3
7b	C	C3	7b'	C	C3
8	C	CE	8'	C	CE
9	N	NH	9'	N	NH
10	C	CE	10'	C	CE
11	C	CF	11'	C	CF
12	C	CX			
13	C	C3	13'	C	C3
14	C	C3	14'	C	C3
15	C	C3	15'	C	C3
16	O	OH	16'	O	O

Table 5 *The atom types for the fluorescent dye Cy5*

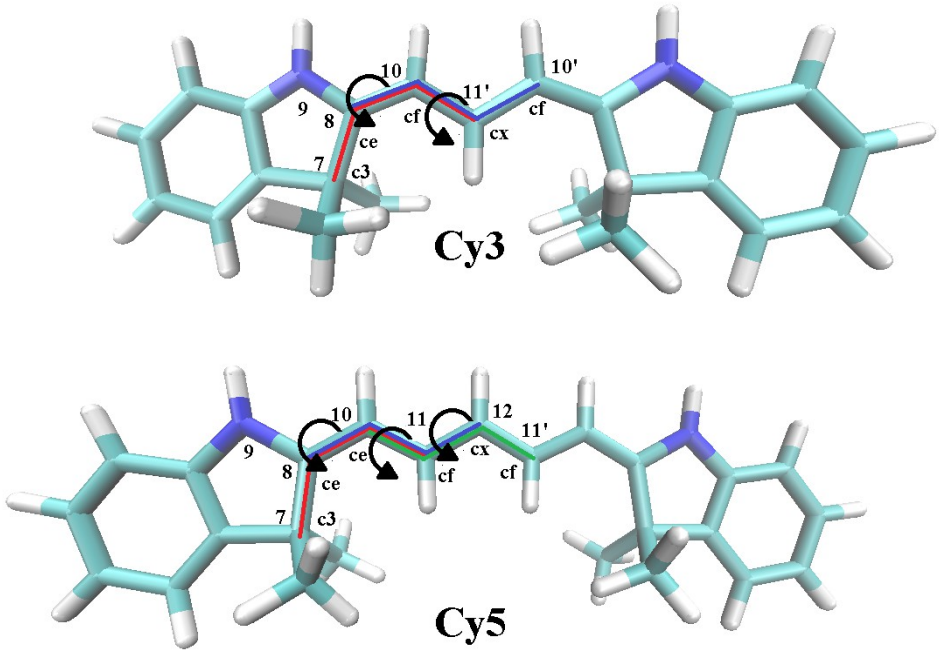


Fig. 12 *Rotational barriers were determined by rotation of dihedral angles in Cy3 (c3-ce-cf-cx, ce-cf-cx-cf) and Cy5 dye (c3-ce-ce-cf, ce-ce-cf-cx, ce-cf-cx-cf).*

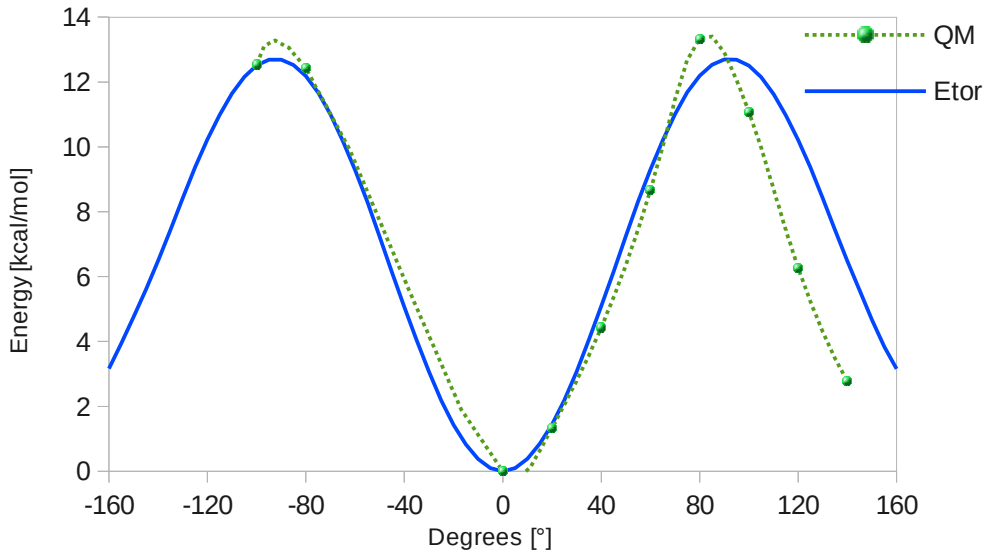


Fig. 13 *Calculated rotational barrier for the c3-ce-cf-cx dihedral angle of the Cy3 dye obtained from quantum chemical calculations (green points) and the fitted function* E_{tors} *whose parameters were used for molecular dynamics simulations.*

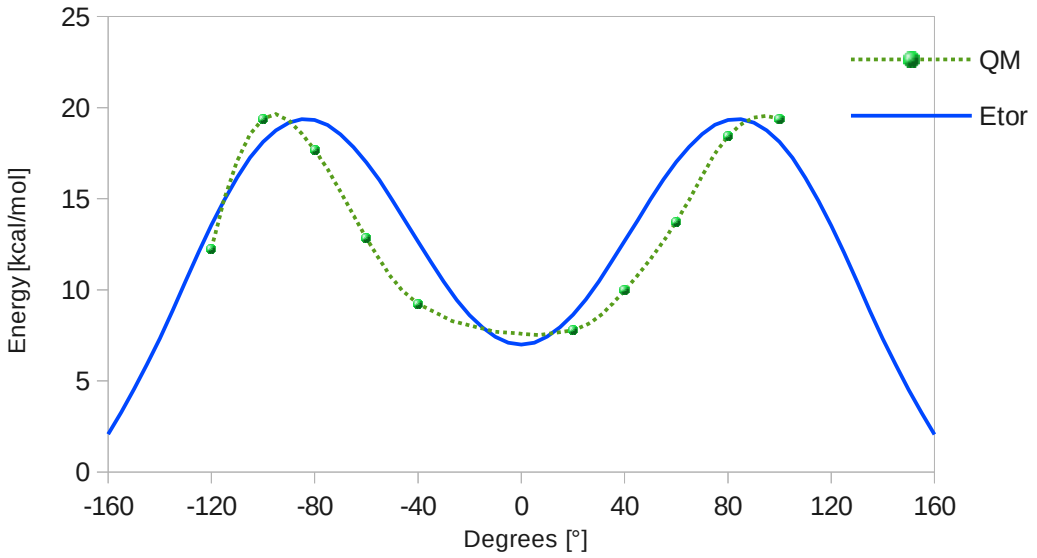


Fig. 14 *Calculated rotational barrier for the ce-cf-cx-cf dihedral angle of the Cy3 dye obtained from quantum chemical calculations (green points) and the fitted function* E_{tors} *whose parameters were used for molecular dynamics simulations.*

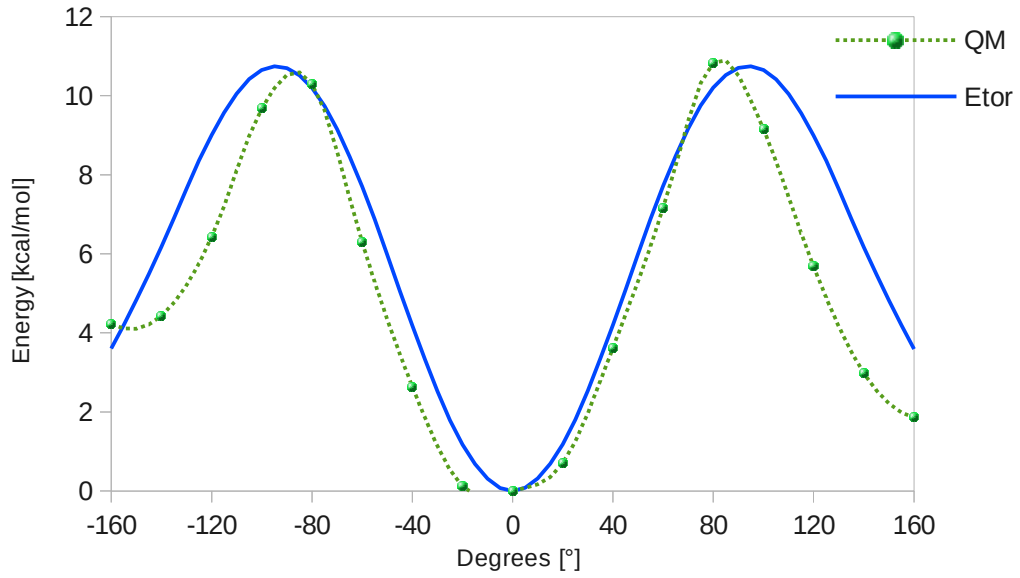


Fig. 15 *Calculated rotational barrier for the c3-ce-ce-cf dihedral angle of the Cy5 dye obtained from quantum chemical calculations (green points) and the fitted function* E_{tors} *whose parameters were used for molecular dynamics simulations.*

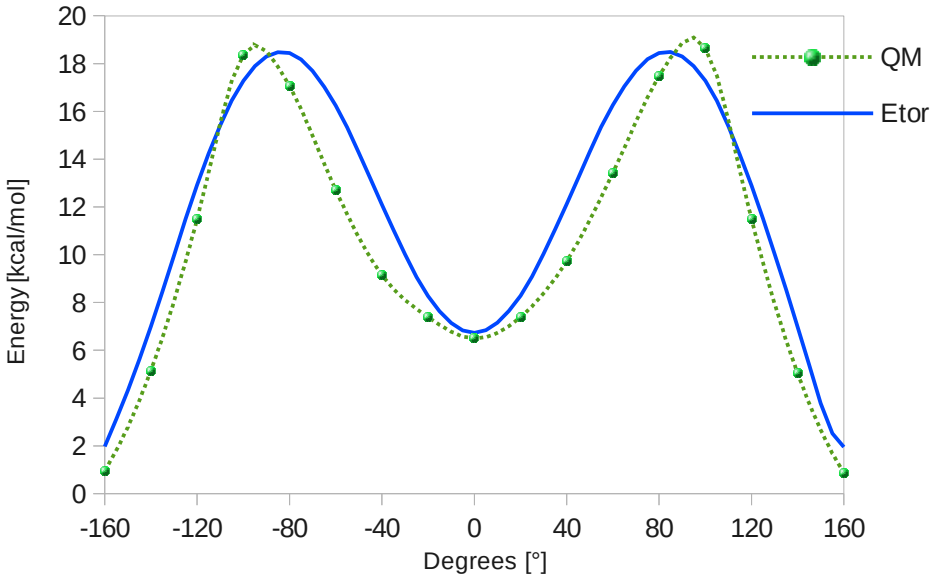


Fig. 16 *Calculated rotational barrier for the ce-ce-cf-cx dihedral angle of the Cy5 dye obtained from quantum chemical calculations (green points) and the fitted function* E_{tors} *whose parameters were used for molecular dynamics simulations.*

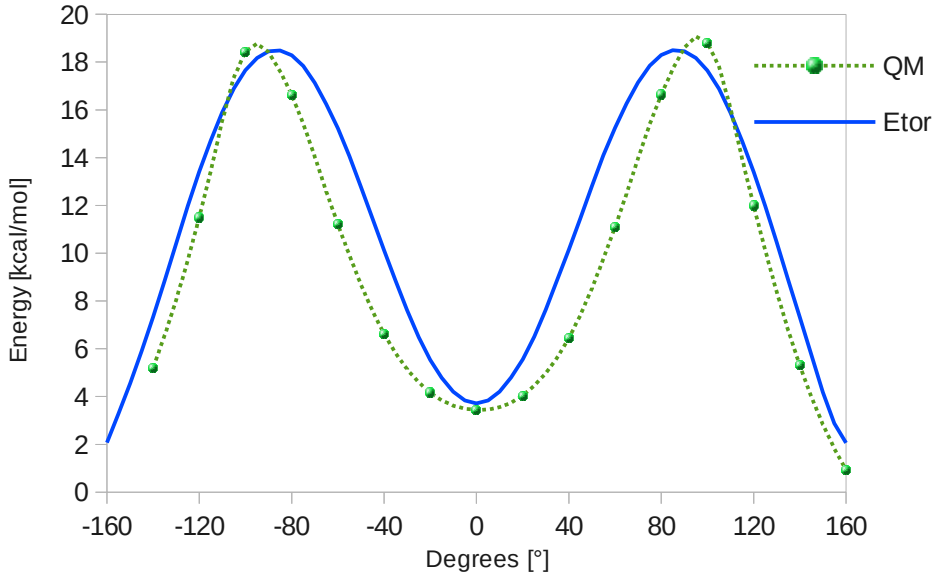


Fig. 17 *Calculated rotational barrier for the ce-cf-cx-cf dihedral angle of the Cy5 dye obtained from quantum chemical calculations (green points) and the fitted function* E_{tors} *whose parameters were used for molecular dynamics simulations.*

3.3 Rigid Body Parameters

3.3.1 Rigid Body Parameters for the Cy3 Dye

It is known that the key DNA parameters of *Twist* and *Rise* are strongly linked to the stacking interactions between nucleic acid base-pairs.[32] But could be the same consideration applied to distinguish all the conformations of the Cy3-DNA complex (see Figure 18)? To answer this question we calculated all base-pair and base-pair step parameters between the DC_1 and DG_20 base pair and the Cy3 dye. The analysis were performed by the X3DNA program and the Cy3 dye was transformed to the two adenine bases as described above.

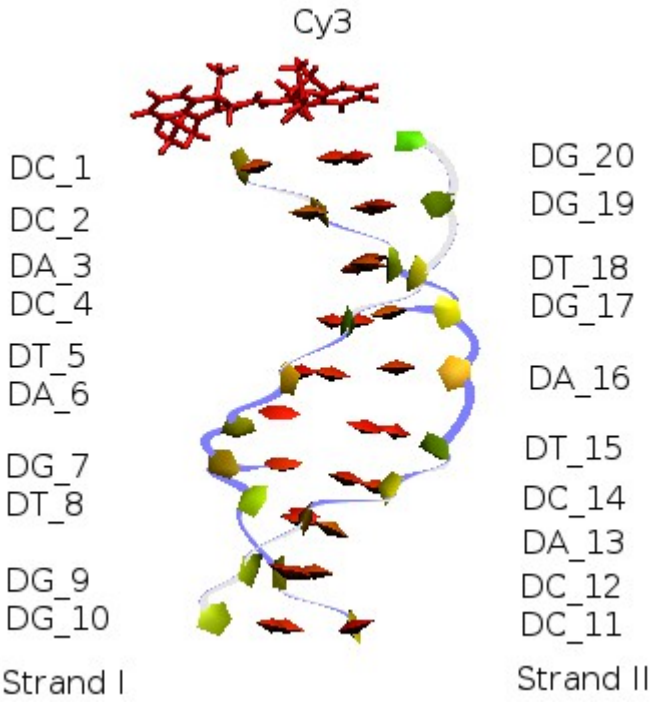


Fig. 18 *The Cy3-DNA complex*

Additionally, the correlation between the calculated parameters and RMSd was tested. The mean values of the parameters for all the conformations are shown in Tables 6 and 7. The representative structures corresponding to these values are shown in Figure 19. The changes of these parameters during the production run are presented in Figures 20 and 21.

The data in the Table 6 show that *Twist* and *Shift* parameters are really the key parameters which change during the conformational switches. The calculated RMSd values reasonably well correlate with *Shift* and *Twist* parameters (Table 6 and Figure 20). On the other hand, there are some conformations which have the same values of these two parameters: A, H an G. For this reason, we consider two additional parameters: *Tilt* (to distinguish between A and H conformations) and *Slide* (between A and G conformations). No significant differences between the conformers were found for the *Rise* and *Roll* parameters.

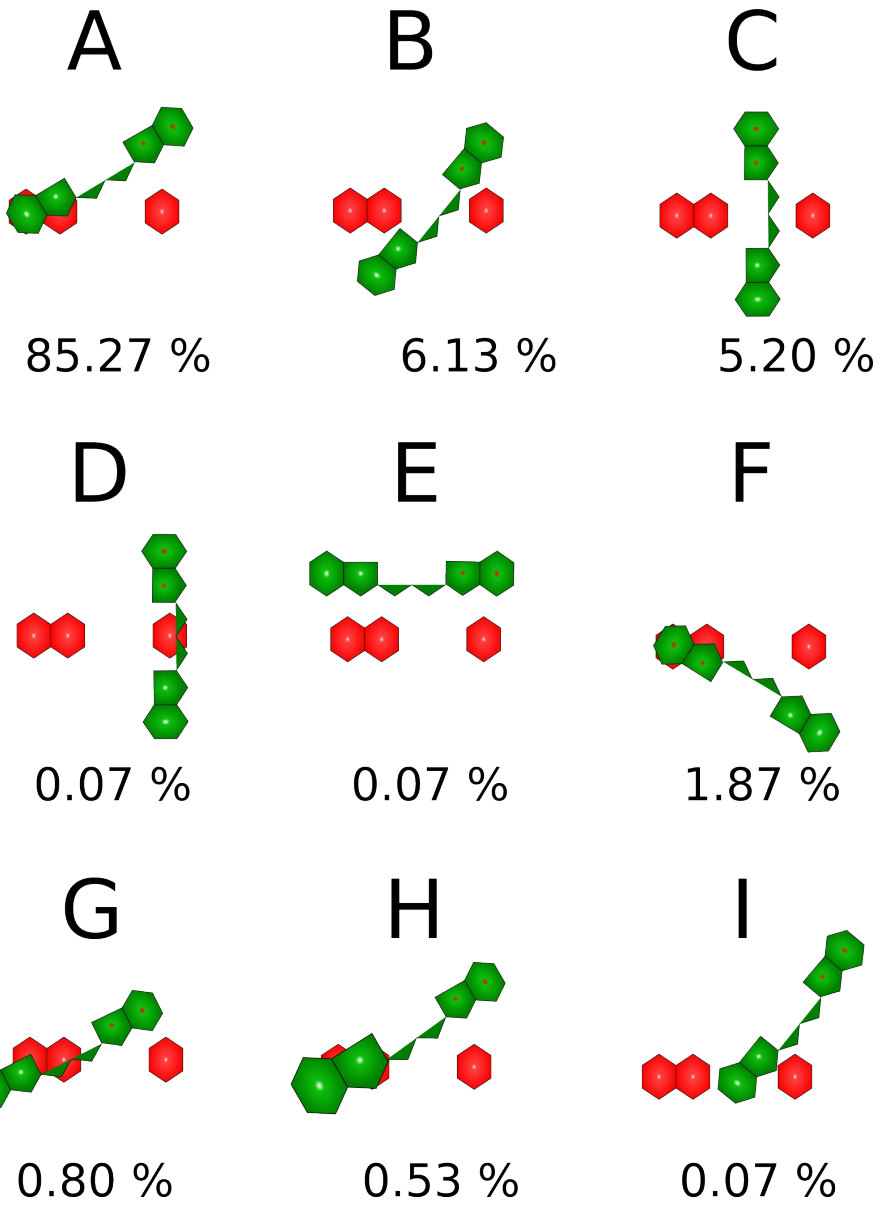


Fig. 19 *The schematic drawings of all observed conformations of Cy3-DNA complex. The DC_1 - DG_20 base pair and the Cy3 dye are shown in red and green, respectively. The percentage abundance of the conformations is also shown.*

On the other hand the base-pair parameters were almost constant during the production run (Figure 21) and independent on the conformer (Table 7). Thus these parameters cannot be used to distinguish different Cy3-DNA conformations. It is not surprising since the structures differ in the mutual orientation of the Cy3 dye and the DC_1 and DG_20 base pair. The conformation changes have only a minor effect on the geometries of the Cy3 dye and the DC_1 and DG_20 base pair. The exceptions are D, E, G and I conformers which differ in the *Buckle* and *Propeller* parameters but all these conformers belongs to the least probable ones.

Interestingly the Cy3 dye is not planar but the two indole rings are turned slightly with respect to each other by about 15 degrees with respect to each other (see *Propeller* values in Table 7).

Conf.	Rise	Roll	Shift	Slide	Tilt	Twist
A	4.12 ± 0.25	-19.40 ± 4.10	-4.11 ± 0.53	-1.67 ± 0.74	-4.82 ± 2.85	47.42 ± 4.97
B	4.66 ± 0.25	-12.44 ± 4.73	-1.47 ± 0.82	-0.90 ± 0.82	-14.22 ± 4.82	67.07 ± 6.82
C	4.75 ± 0.27	-10.39 ± 5.54	-0.28 ± 0.93	-0.01 ± 1.22	-16.09 ± 6.32	76.99 ± 8.49
D	3.86	-5.68	-1.61	1.95	-5.86	18.98
E	-c	-c	-c	-c	-c	-c
F	4.34 ± 0.62	-9.13 ± 7.57	2.37 ± 0.63	2.75 ± 0.75	-3.30 ± 14.79	103.45 ± 7.61
G	4.16 ± 0.22	-11.59 ± 2.62	-3.88 ± 0.46	0.36 ± 1.08	-7.10 ± 2.44	52.44 ± 7.36
H	4.36 ± 0.58	2.21 ± 8.69	-3.52 ± 2.39	-1.21 ± 1.63	-13.22 ± 15.82	46.37 ± 20.38
I	-c	-c	-c	-c	-c	-c

Table 6 *The average values of the base-pair-step parameters of Cy3-GC for nine conformations that were recognized during the 150 ns MD simulation of Cy3 DNA.*

Conf.	Buckle	Opening	Propeller	Shear	Stagger	Stretch
A	-3.79 ± 3.73	68.20 ± 1.05	-10.96 ± 9.58	0.01 ± 0.03	0.32 ± 0.39	-13.09 ± 0.1
B	2.83 ± 4.52	68.80 ± 1.13	-16.64 ± 8.06	-0.00 ± 0.03	0.65 ± 0.34	-13.02 ± 0.11
C	4.98 ± 4.16	69.08 ± 1.28	-14.15 ± 9.25	0.01 ± 0.04	0.61 ± 0.39	-13.01 ± 0.12
D	-10.41	66.72	-5.24	-0.02	0.28	-13.18
E	-3.46	68.47	-27.46	-0.13	1.13	-12.97
F	-1.87 ± 6.17	67.99 ± 1.35	-12.93 ± 20.62	-0.04 ± 0.06	0.48 ± 0.81	-13.07 ± 0.15
G	-2.98 ± 3.73	68.87 ± 0.87	12.52 ± 5.17	0.00 ± 0.02	-0.64 ± 0.31	-13.04 ± 0.10
H	-0.88 ± 6.34	68.49 ± 1.96	-10.36 ± 11.09	0.02 ± 0.07	0.38 ± 0.53	-12.92 ± 0.18
I	-0.2	66.53	-37.98	-0.05	1.47	-13.15

Table 7 *The average values of base-pair-parameters of the Cy3 residue for nine conformations A-I .*

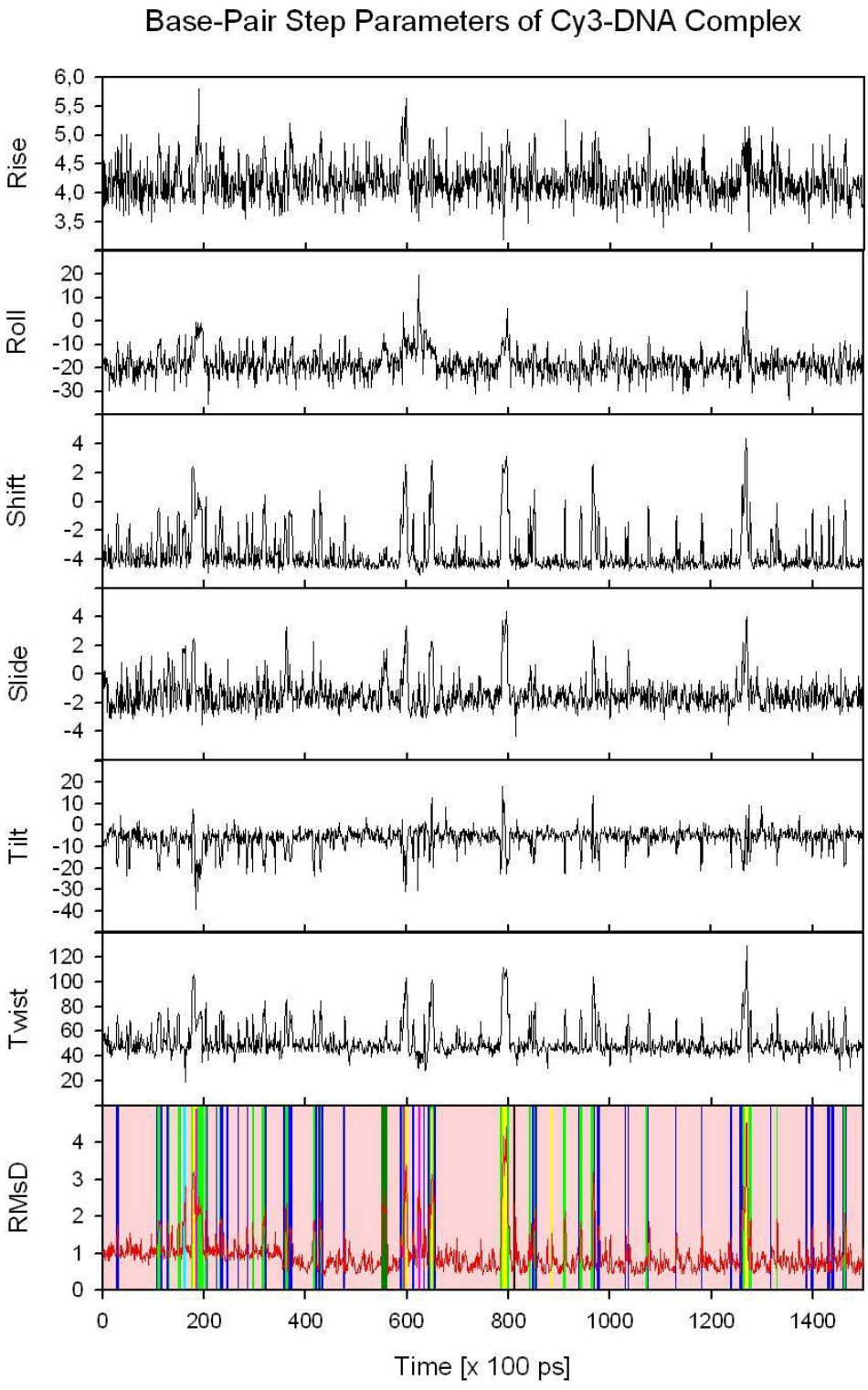


Fig. 20 *Changes of the base-pair step parameters and the RMSd values of the Cy3-DNA complex during the 150 ns MD simulation. In the RMSd graph the occurence of the conformers is distinguuished by different colors of the background (red – A, blue – B, green – C, light blue – D, orange – E, yellow – F, dark green – G, pink – H, black – I).*

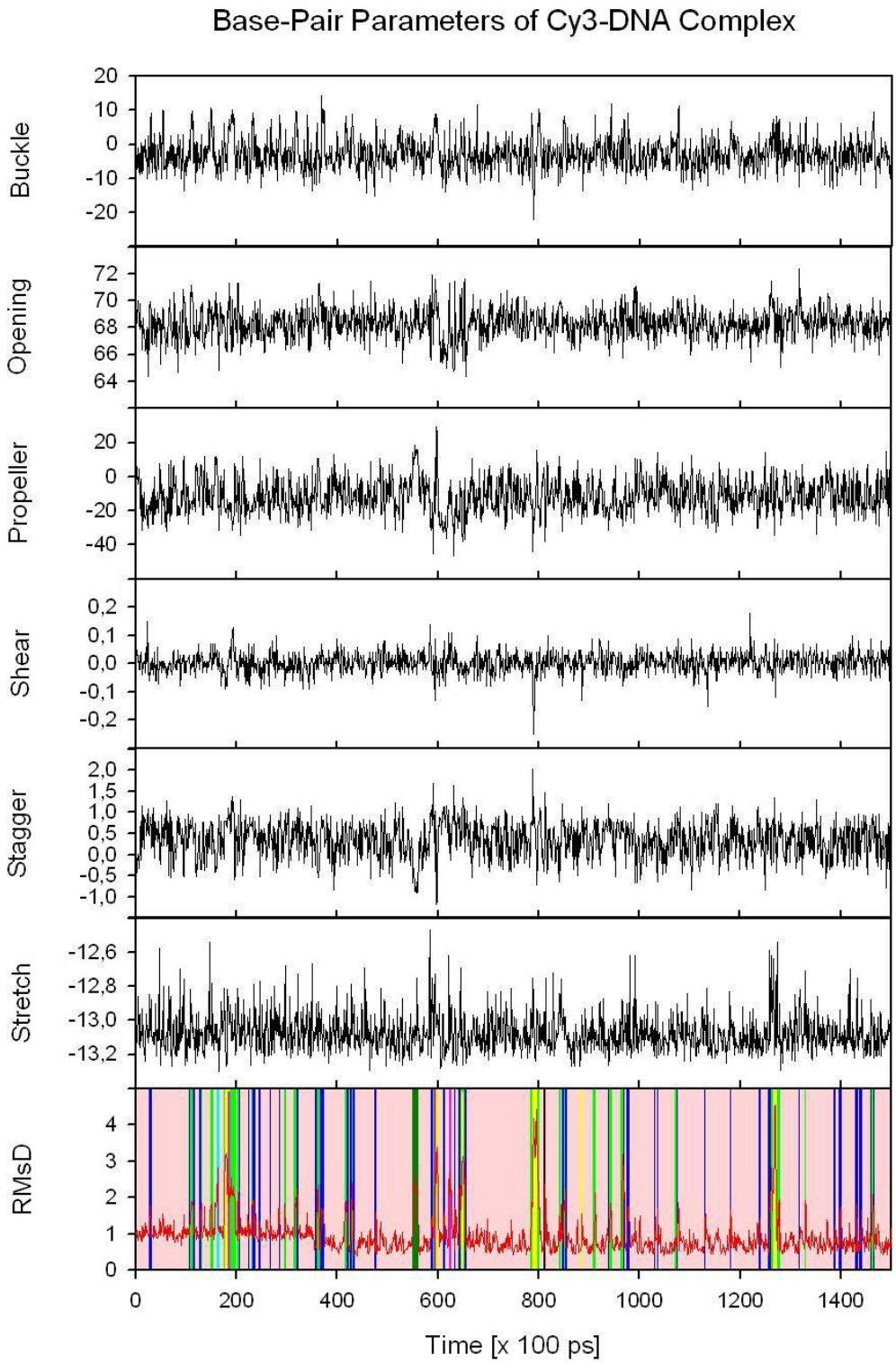


Fig. 21 *Changes of the base-pair parameters and the RMSd values of the Cy3-DNA complex during the 150 ns MD simulation. In the RMSd graph the occurence of the conformers is distinguuished by different colors of the background (red – A, blue – B, green – C, light blue – D, orange – E, yellow – F, dark green – G, pink – H, black – I).*

3.3.2 Rigid Body Parameters for the Cy5 Dye

If we summarize the main results from the previous chapter, all conformations can be reliably distinguished on the basis of base-pair step parameters (*Twist* and *Shift*, for some cases *Slide* and *Tilt*). The representative structures are presented in Figure 23 and their designation was chosen with respect to the Cy3 conformers: i.e. conformers of Cy3 and Cy5 dyes designated by the same letter roughly correspond to each other.

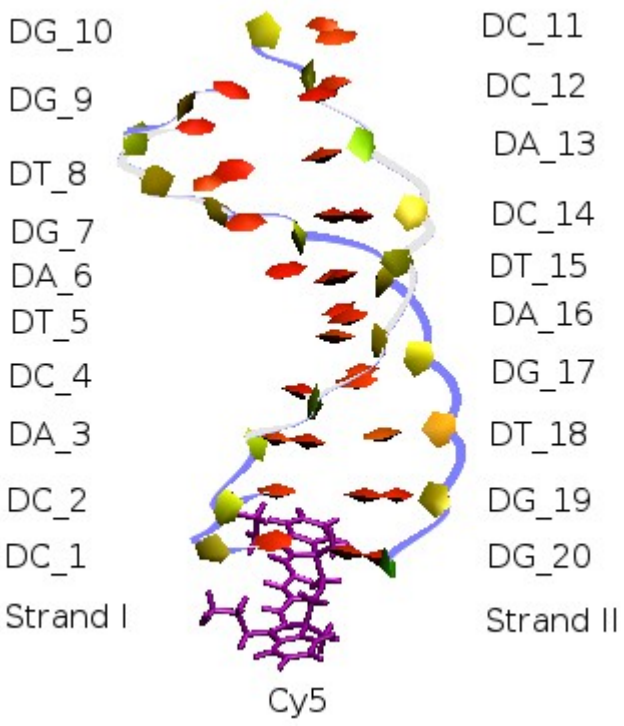


Fig. 22 *The Cy5-DNA complex.*

For Cy5-DNA complex (see Figure 22) some parameters are not fully comparable with the Cy3 results. The indole rings of Cy3 dye are connected by three-carbon polymethine chain while Cy5 dye has the two additional carbon atoms in the polymethine chain. Thus for instance the *Shift* values are more positive for Cy5-DNA than for Cy3-DNA when the same conformations of the two complexes are compared.

The data in the Table 8 show that mainly *Twist, Slide* and *Shift* parameters can be considered as important parameters for a recognition of the conformations.

On the basis of *Shift* parameters only two conformations can be distinguished – A and J. D and I conformations can be reliably distinguished by *Twist* and *Slide* parameters (see Figure 24).

The base-pair parameters cannot be used to distinguish different Cy5-DNA conformations see Table 9, Figure 25). It is not surprising since the structures differ in the mutual orientation of the Cy5 dye and the DC_1 and DG_20 base pair. The conformation changes between Cy5 and the GC pair have only a little effect on the geometries of the Cy5 and the GC residues themselves.

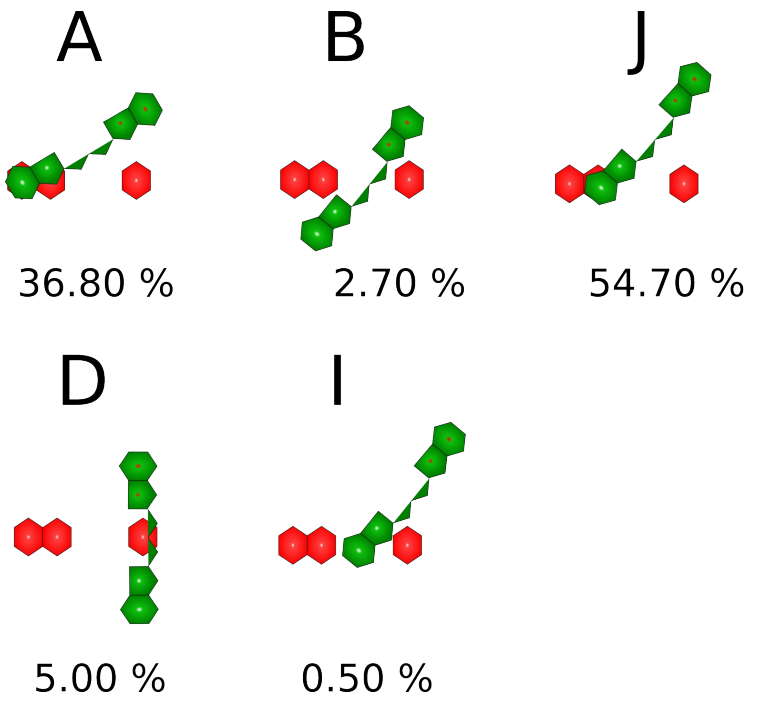


Fig. 23 *The schematic drawings of all observed conformations of Cy5-DNA complex. The DC_1 - DG_20 base pair and the Cy5 dye are shown in red and green, respectively. The percentage abundance of the conformations is also shown.*

C.	Rise	Roll	Shift	Slide	Twist	Tilt
A	4.0 ± 0.6	-20.5 ± 9.2	4.2 ± 0.6	-2.8 ± 0.8	34.8 ± 14.8	4.7 ± 16.8
B	3.7 ± 1.0	-1.0 ± 9.7	0.5 ± 0.8	-4.2 ± 0.9	56.3 ± 2.3	15.3 ± 9.8
D	3.4 ± 2.1	-16.4 ± 30.1	-0.7 ± 2.5	-2.9 ± 1.4	61.1 ± 42.9	-20.1 ± 50.5
I	3.5 ± 3.2	-28.4 ± 25.1	1.3 ± 1.1	-6.2 ± 2.2	5.7 ± 14.8	-32.7 ± 47.9
J	4.3 ± 0.9	-17.2 ± 13.7	2.8 ± 0.8	-4.0 ± 1.5	33.6 ± 15.1	-0.1 ± 15.4

Table 8 *The average values of the base-pair-step parameters of Cy5-GC for five conformations that were recognized during the 150 ns MD simulation of Cy5-DNA.*

C.	Buckle	Opening	Propeller	Shear	Stagger	Stretch
A	3.7 ± 6.5	62.7 ± 1.6	-8.9 ± 15.5	0.0 ± 0.1	0.2 ± 0.6	-15.7 ± 0.1
B	0.5 ± 21.3	9.3 ± 22.8	-2.7 ± 7.5	-0.1 ± 0.1	-0.3 ± 0.5	-16.7 ± 0.5
D	-5.2 ± 17.8	18.6 ± 36.7	-5.9 ± 15.9	-0.2 ± 1.8	-0.7 ± 1.0	-16.2 ± 2.3
I	-19.4 ± 19.9	1.5 ± 6.3	-13.0 ± 37.8	-0.3 ± 0.6	-0.7 ± 0.3	-16.9 ± 0.2
J	-0.4 ± 17.0	40.8 ± 33.4	-3.7 ± 22.6	-0.2 ± 1.2	0.1 ± 0.9	-16.0 ± 1.6

Table 9 *The average values of the base-pair-parameters of the Cy5 dye for five conformations A, B, D, I, J.*

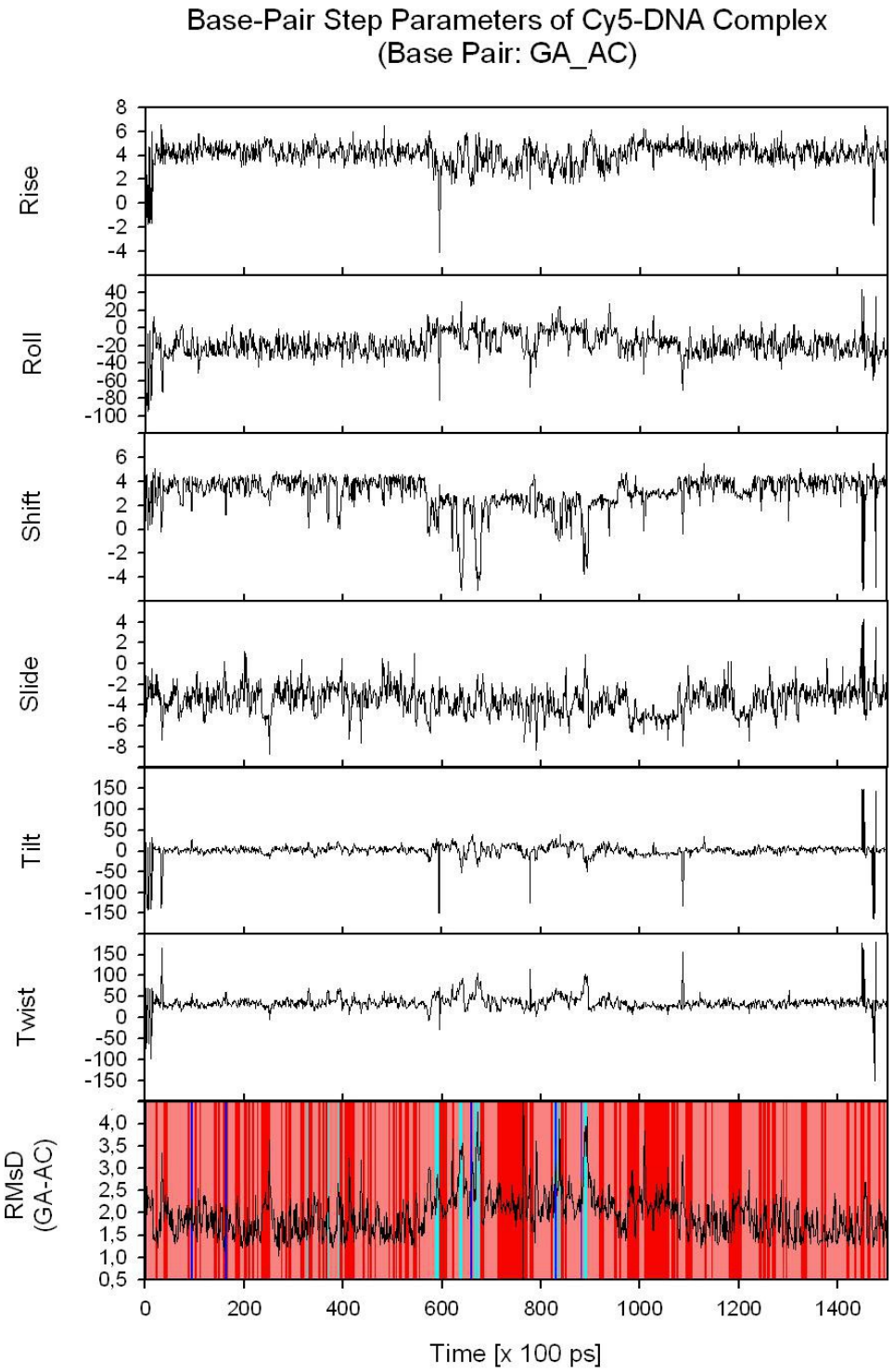


Fig. 24 *Changes of the base-pair step parameters and the RMSd values of the Cy5-DNA complex during the 150 ns MD simulation. In the RMSd graph the occurence of the conformers is distinguuished by different colors of the background (light red – A, blue – B, light blue – D, red – J, black – I, yellow – N).*

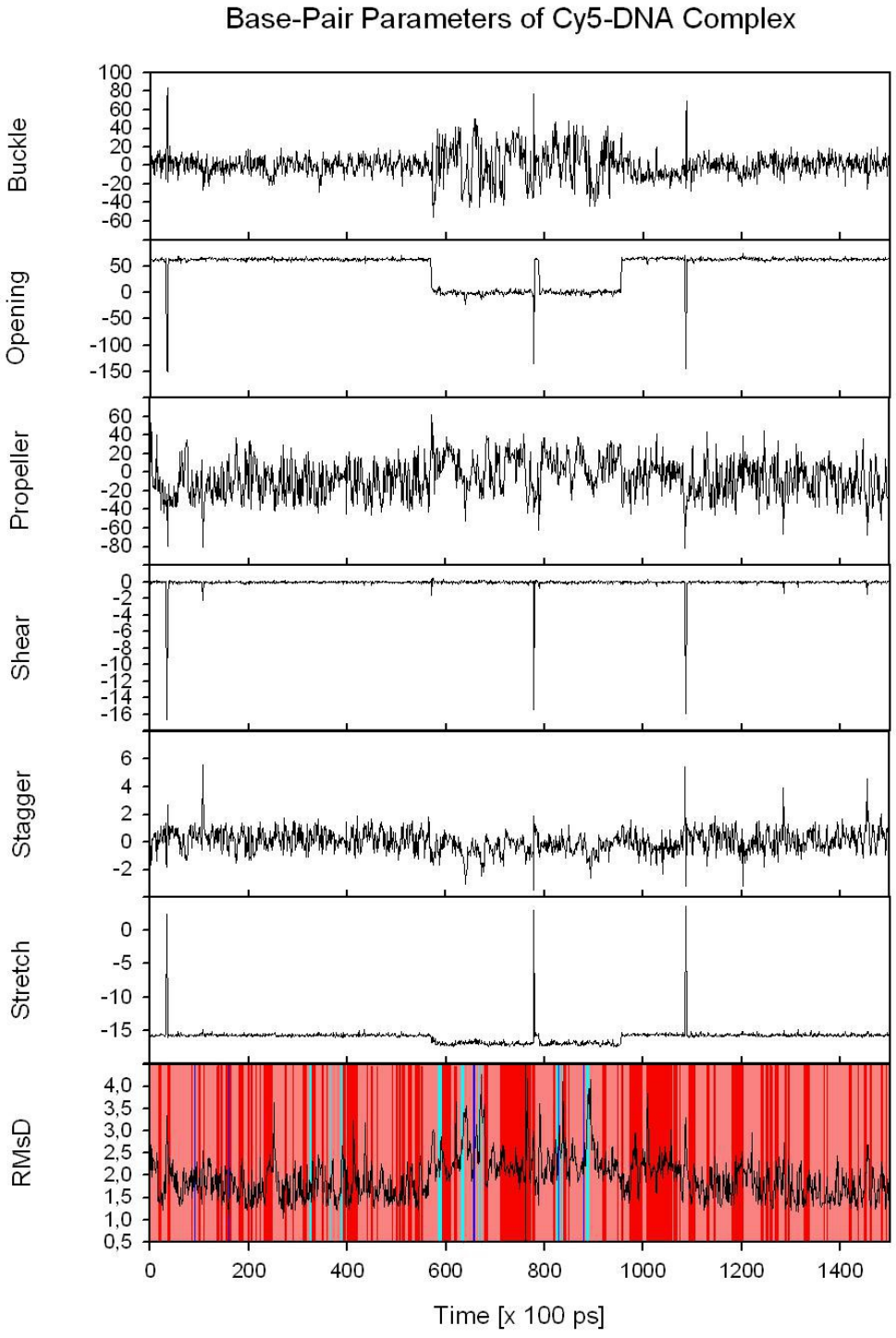


Fig. 25 *Changes of the base-pair parameters for the A-A base pair (converted from the two indole rings of Cy5) and the RMSd values of the Cy5-DNA complex during the 150 ns MD simulation. In the RMSd graph the occurrence of the conformers is distinguuished by different colors of the background (light red – A, blue – B, light blue – D, red – J, black – I, yellow – N).*

3.4 Backbone Torsion Angles of Cyanine Dyes

3.4.1 Backbone Torsion Angles in the Cy3-DNA Complex

The X3DNA program calculates a set of nucleic acid backbone parameters, including the six main chain torsion angles (α, β, γ, δ, ε, and ζ) around the covalent bonds, χ about the glycosidic bond, and the sugar pucker. We focused our attention on some of them, on the main chain torsion angles and then on rotation angle about the glycosidic bond.

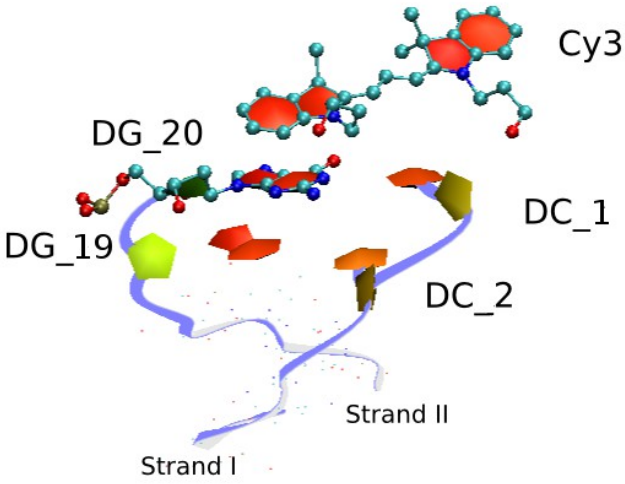


Fig. 26 *The Cy3-DNA complex with highlighted sugar attached to DG_20 base.*

At first let me explain some assumptions about the dihedral angles of Cy3-DNA and Cy5-DNA complexes.

Double-stranded DNA consists of two antiparallel strands I and II (see Figure 26). Therefore the order of six main chain torsion angles for the *strand I* has also antiparallel orientation with respect to the *strand II* (see Figure 27).

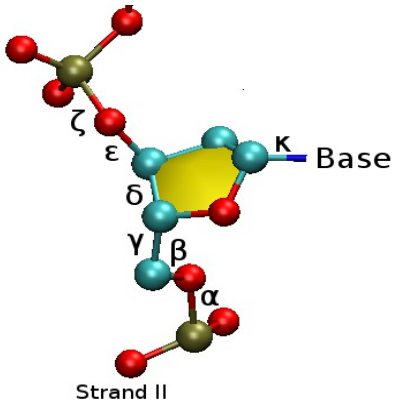


Fig. 27 *The dihedral angles of DG_19 base for strand II of doubled-stranded DNA*

Cy3 dye is attached to DC_1 of the *strand I*, so all torsion angles of DC_1 can be analyzed. However, its α dihedral angle is not exactly the sugar-phosphate angle because the oxygen atom on the O3' position is the hydroxyl oxygen atom from the Cy3 dye. To distinguish this case the corresponding α dihedral angle is designated with an asterisk (α*).

The DG_20 base is described only by four main chain torsion angles (α, β, γ, and δ), while ε and ζ dihedral angles cannot be determined since DG_20 is the 3' end of the *strand II*.

The initial structure of our DNA model have canonical geometry of B-DNA (LBDNA: the structure generated by NAB program). After equilibration some significant differences were observed in its dihedral angles (Figure 29). Firstly the conformation of the linker between Cy3 dye and DC_1 changed by a rotation around α* dihedral angle (Figure 28 and 31). Secondly the sugar ring of DG_20 on the *strand II* and hydroxypropyl side chain of the Cy3 distal indole ring changed their orientations (Figure 29). The positions of other atoms did not change significantly with respect to the initial structure as it can be seen from low RMSd values of about 1 Å calculated for Cy3, DC_1 and DG_20 residues (Figure 20).

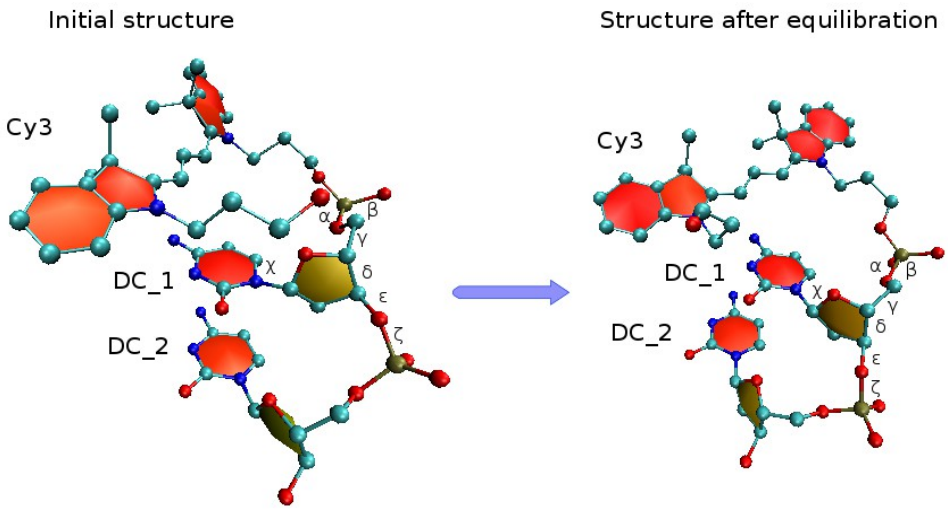


Fig. 28 *The comparison of sugar-phoshate backbone and glycosidic torsion angle between initial and final structure of Cy3-DNA complex of the equilibration.*

During molecular dynamic calculations the most significant change on the sugar-phosphate backbone occurred at time of 35 ns. The sugar of DG_20 rotated about its glycosidic bond back to the position it had in the initial structure (Figure 30). This change can be described by the changes of χ dihedral angle which changed from the starting value of -165.78 ± 3.70 degrees to the value of -115.40 ± 15.02 degrees (Table 11, 12 and Figure 32). This motion also influences the dihedral angles of DG_19 and RMSd values (Appendix 3) being accompanied by the decreased *Shift* between the DG_20-DC_1 and DC_2-DG_19 base

pairs). The relative position of the Cy3 indole rings with respect to the DG_20-DC_1 base pair was not affected by this change and the RMSD value for Cy3, DC_1 and DG_20 residues decreased to ca. 0.7 Å.

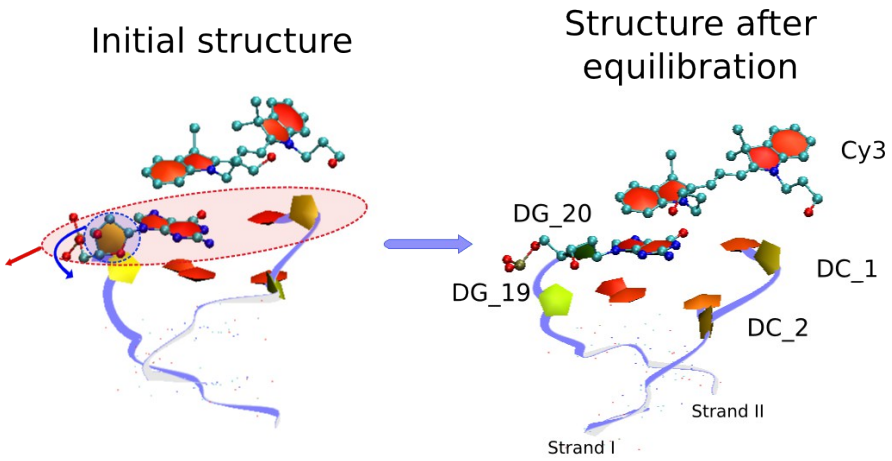


Fig. 29 *Cy3-DNA: the comparison of the initial structure and the structure after the equilibration.*

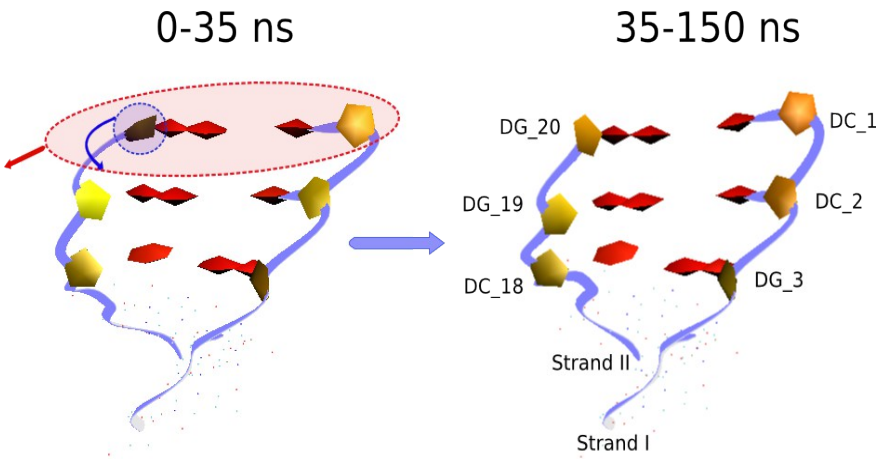


Fig. 30 *The reorientation of the sugar ring of the DG_20 nucleotide at the time of 35 ns (cf. Fig.27)*

DC_1	α	β	γ	δ	ε	ζ	χ
A	-77.5 ± 29.3	160.8 ± 45.9	55.1 ± 18.5	136.1 ± 4.3	-168.1 ± 25.5	-84.7 ± 36.8	-115.3 ± 8.2
B	-75.0 ± 4.8	162.2 ± 4.6	52.6 ± 2.5	135.2 ± 3.0	-158.7 ± 62.5	-88.9 ± 4.2	-117.7 ± 5.4
C	-44.6 ± 56.7	93.0 ± 130.6	32.8 ± 86.3	134.0 ± 5.4	-149.6 ± 86.4	-90.1 ± 3.9	-112.7 ± 8.8
D	-54.8	145.3	59.6	132.4	-68.5	108.3	-60.4
E	-c	166.8	58.9	137.1	-172.9	-94.2	-115.8
F	-28.8 ± 65.6	90.0 ± 107.5	19.3 ± 91.5	129.8 ± 11.5	-134.4 ± 110.2	-95.8 ± 11.4	-114.4 ± 10.5
G	-153.4 ± 14.8	-139.6 ± 100.0	63.5 ± 2.0	96.4 ± 6.9	-151.7 ± 10.0	-67.0 ± 4.0	-91.6 ± 9.7
H	-120.3 ± 80.5	-129.3 ± 112.3	76.2 ± 40.7	111.7 ± 17.9	-149.8 ± 19.2	-71.5 ± 7.6	-101.1 ± 12.8
I	-c	176.7	-64.2	133	-168.7	-86.9	-126.4

Table 10 *The average values of dihedral angles of DC_1 nucleotide for all conformations which were found during the MD production run.*

DG_20	α	β	γ	δ	χ
A	-77.6 ± 20.1	141.4 ± 52.9	51.2 ± 82.8	130.8 ± 9.8	-126.8 ± 22.3
B	-78.2 ± 13.7	143.2 ± 45.2	64.2 ± 81.8	126.2 ± 11.8	-137.1 ± 19.1
C	-80.7 ± 10.4	126.3 ± 50.4	64.9 ± 105.2	129.7 ± 12.6	-142.2 ± 22.9
D	-88.5	63.5	176.1	148.2	-162.6
E	-70.2	174.7	55.8	118.2	-125.1
F	-79.6 ± 9.2	147.4 ± 38.4	49.1 ± 75.7	129.7 ± 10.9	-124.9 ± 20.4
G	92.0 ± 10.9	-166.4 ± 5.3	144.1 ± 102.6	124.1 ± 2.6	-29.7 ± 5.2
H	-115.5 ± 33.2	63.3 ± 151.0	68.2 ± 45.2	121.0 ± 9.1	-56.6 ± 65.1
I	-71	172.3	55.2	107.1	-127.9

Table 11 *The average values of dihedral angles of DC_20 nucleotide for all conformations which were found during the MD production run.*

DC_1	α*	β	γ	δ	ε	ζ	χ
Initial	81.9	-172.5	55.0	136.0	178.8	-83.5	-100.5
After eq.	-74.0	164.6	52.7	139.2	-174.2	-90.3	-107.9
DG_20	**α**	**β**	**γ**	**δ**	**ε**	**ζ**	**χ**
Initial	-67.3	134.3	37.6	144.6	-	-	-98.7
After eq.	-89.0	69.3	-178.4	140.1	-	-	-166.6
DG_19	**α**	**β**	**γ**	**δ**	**ε**	**ζ**	**χ**
Initial	-62.8	176.7	40.9	141.2	-87.1	136.2	-85.2
After eq.	-81.0	172.3	47.2	128.8	-162.0	-84.3	-97.6

Table 12 *The comparison of initial structure and the average structure of the first 100 ps of production run with respect to all dihedral angles of DC_1, DG_20, and DG_19 bases.*

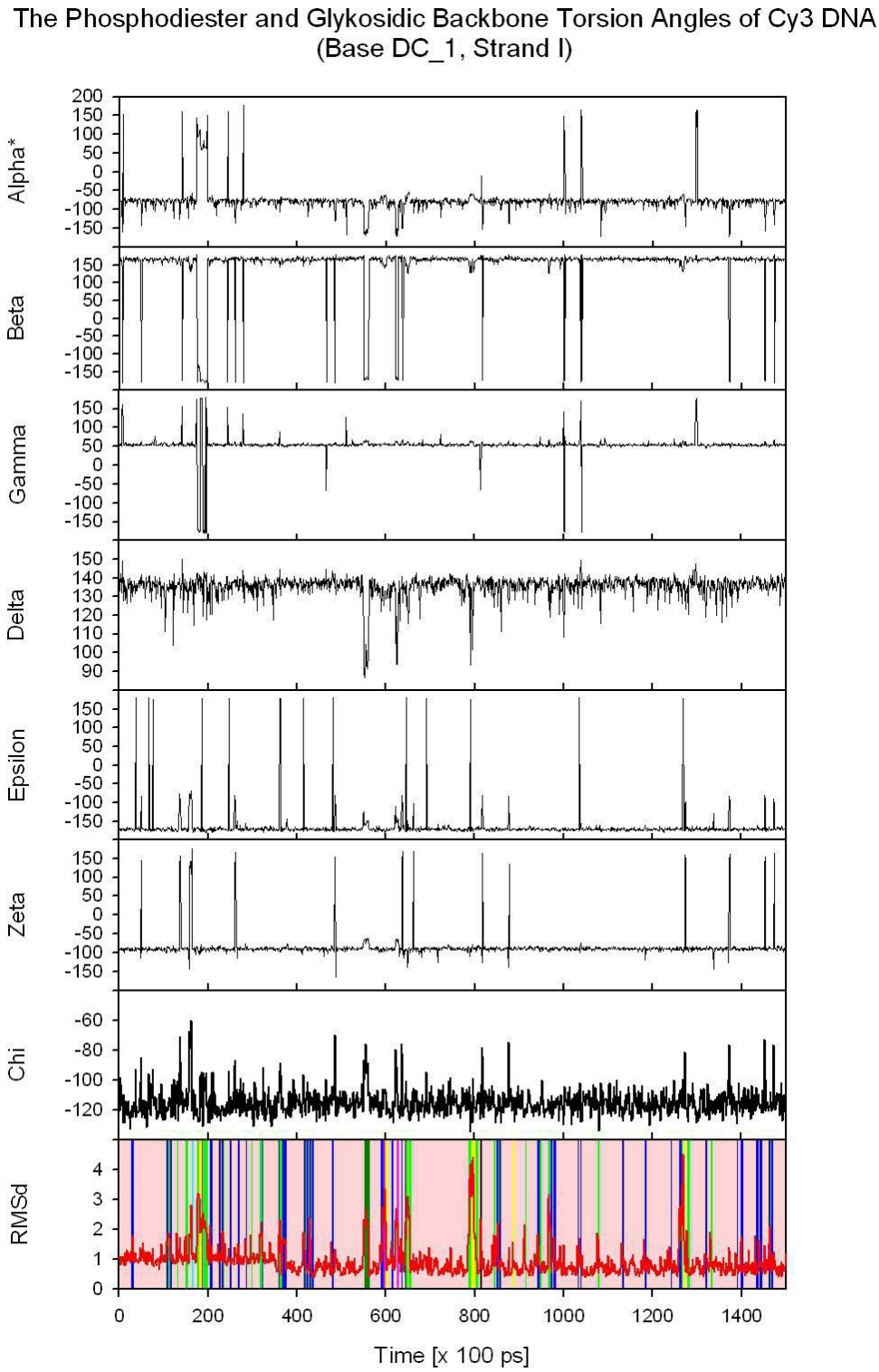


Fig. 31 *The phosphodiester and glycosidic backbone torsion angles of Cy3-DNA complex for DC_1 base of strand I.*

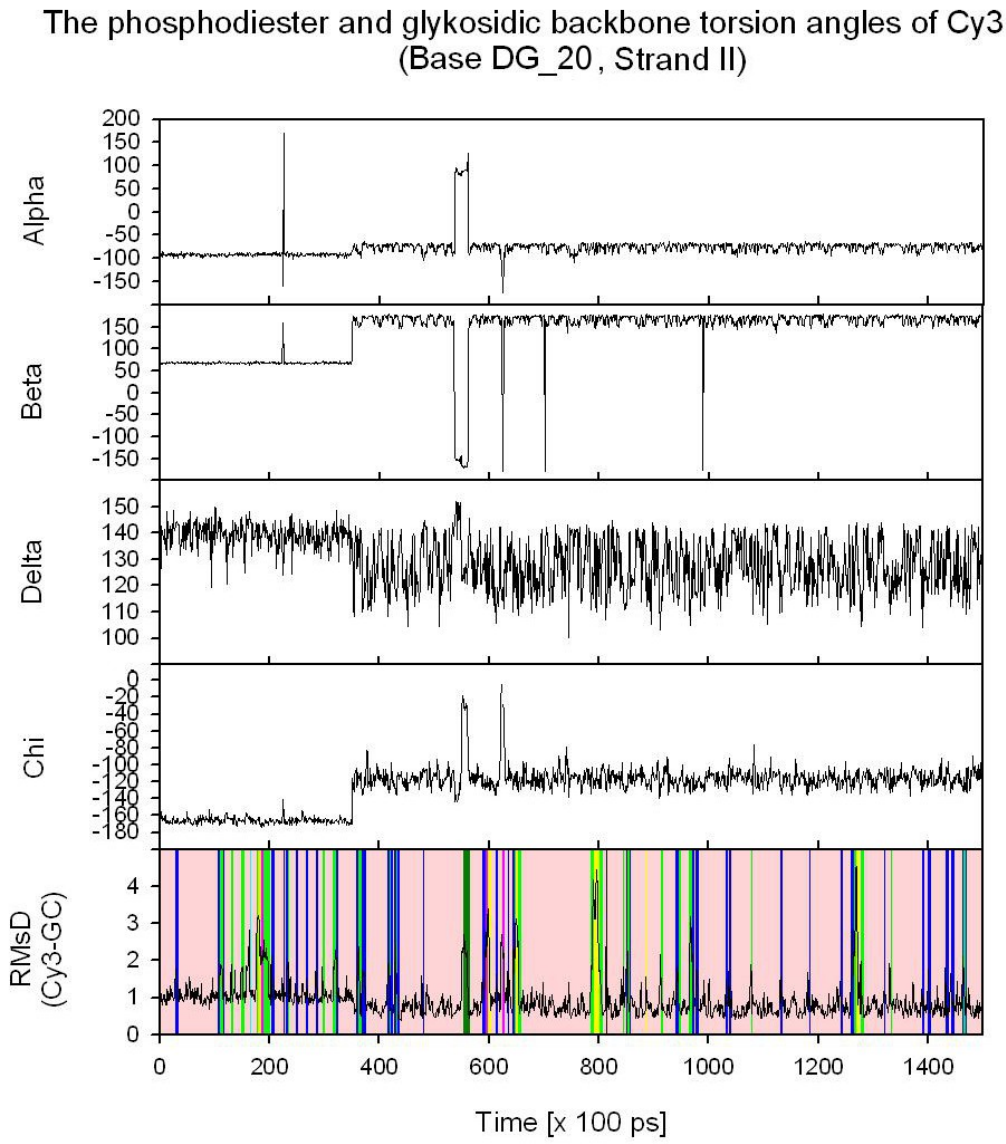


Fig. 32 *The phosphodiester and glycosidic backbone torsion angles of Cy3-DNA complex for DG_20 base of strand II.*

3.4.2 Backbone Torsion Angles in the Cy5-DNA Complex

The DNA backbone is more stable during the equilibration run than in the Cy3-DNA case: the dihedral angles for the initial structure and average structure of the first 100 ps of the production run for Cy5-DNA complex differ only in α* torsion angle which rapidly decreased by about 60 degrees (Table 15). For more information see Tables 13, 14 and Figures 33, 34.

DC_1	α	β	γ	δ	ε	ζ	χ
A	-104.4 ± 37.5	42.6 ± 168.3	55.5 ± 17.0	134.7 ± 6.9	-165.3 ± 6.1	-84.9 ± 5.9	-130.9 ± 8.6
B	-122.6 ± 56.4	-43.4 ± 170.2	46.0 ± 53.2	132.5 ± 9.4	-167.9 ± 4.7	-85.7 ± 5.4	-125.9 ± 8.5
D	-94.3 ± 36.2	104.8 ± 129.6	49.8 ± 30.1	131.3 ± 9.8	-168.0 ± 4.9	-86.1 ± 4.9	-125.3 ± 9.5
I	-42.6 ± 30.4	-163.2 ± 24.4	28.4 ± 55.4	131.4 ± 7.7	-169.7 ± 3.1	-91.0 ± 4.2	-128.1 ± 8.9
J	-83.0 ± 53.5	21.1 ± 167.8	48.0 ± 31.7	133.8 ± 7.4	-154.5 ± 56.9	-77.3 ± 49.5	-122.8 ± 16.9

Table 13 *The average values of dihedral angles of DC_1 base for all conformations occurring during MD production run.*

DG_20	α	β	γ	δ	χ
A	-73.8 ± 7.3	169.1 ± 17.0	52.3 ± 3.6	124.6 ± 10.2	-119.3 ± 10.7
B	-72.0 ± 6.3	169.7 ± 7.9	53.6 ± 3.1	119.4 ± 9.2	-127.4 ± 6.2
D	-73.2 ± 12.4	171.7 ± 5.4	55.3 ± 2.4	119.0 ± 8.4	-124.7 ± 9.8
I	-69.1 ± 2.9	171.7 ± 2.4	55.3 ± 1.8	115.3 ± 4.8	-128.8 ± 6.7
J	-74.6 ± 7.7	166.2 ± 15.1	52.8 ± 3.8	125.4 ± 10.7	-118.8 ± 10.2

Table 14 *The average values of dihedral angles of DG_20 base for all conformations occurring during MD production run.*

DC_1	α*	β	γ	δ	ε	ζ	χ
Initial	-49.6	-178.2	-53.4	152.3	-175.5	-116.1	-98.4
After eq.	-110.7	175.9	51.9	139.9	-169.7	-89.1	-117.9
DG_20	**α**	**β**	**γ**	**δ**	**ε**	**ζ**	**χ**
Initial	-79.0	-147.7	57.6	130.0	-	-	-93.4
After eq.	-68.8	173.0	55.0	122.8	-	-	-123.1

Table 15 *The comparison of initial structure and average structure of the first 100 ps of the production run with respect to all dihedral angles of DC_1 ang DG_20 bases.*

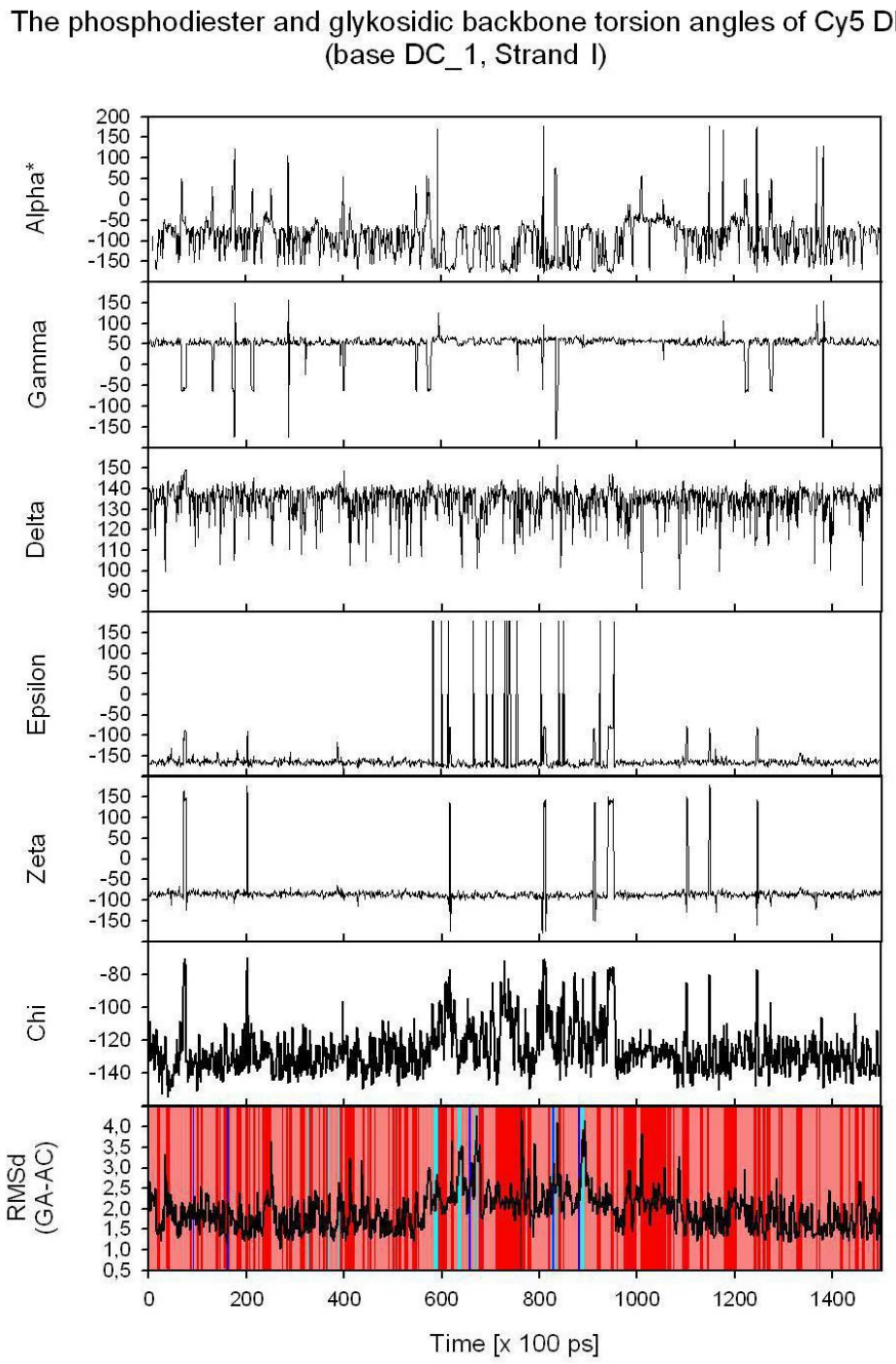


Fig. 33 *The phosphodiester and glycosidic bakbone torsion angles of Cy5-DNA complex for DC_1 base of strand I.*

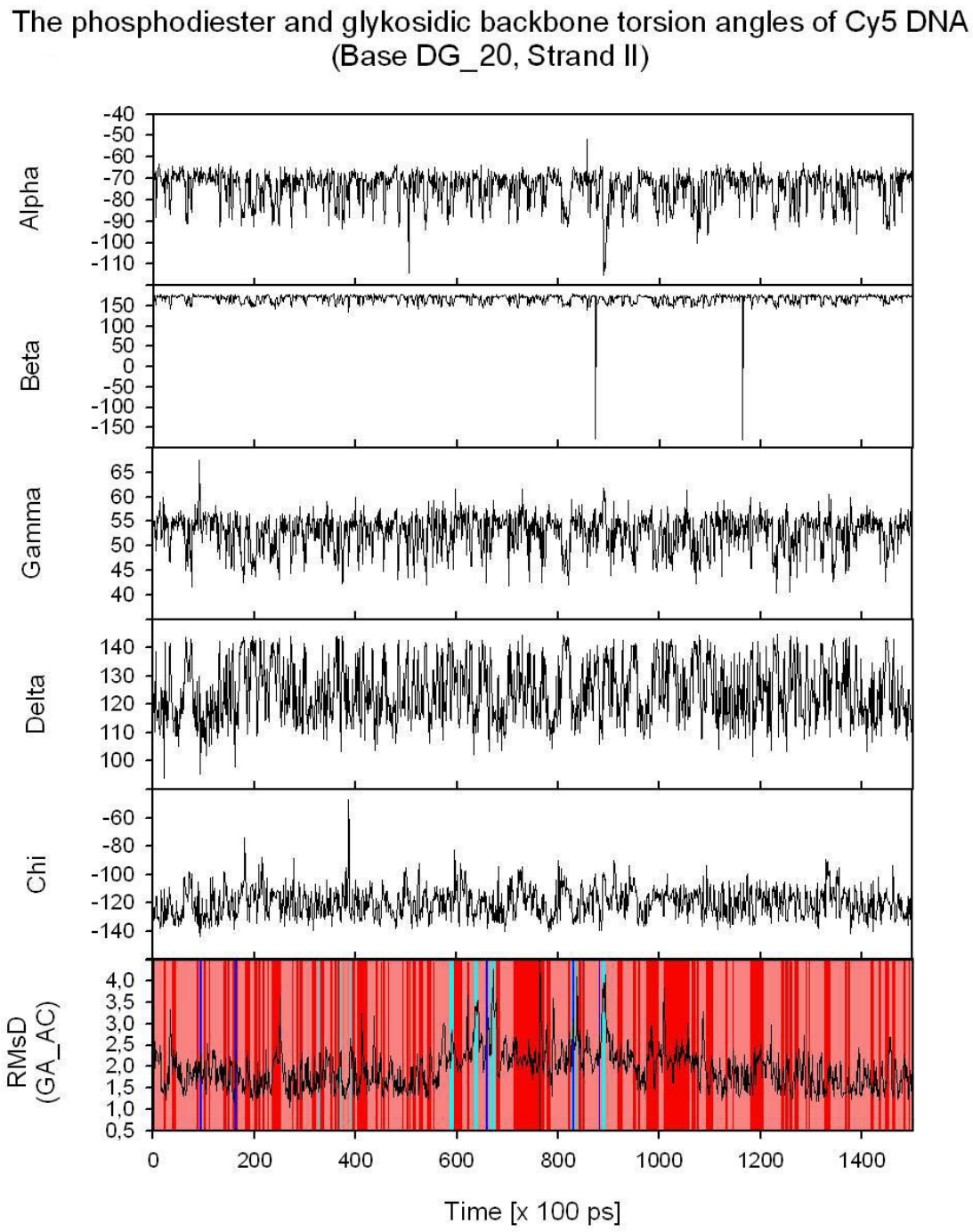


Fig. 34 *The phosphodiester and glycosidic bakbone torsion angles of Cy5-DNA complex for DG_20 base of strand II.*

3.5 RMSd of Cyanine Dyes

3.5.1 RMSd of the Cy3 Dye

All averaged structures were compared with the initial structure of molecular dynamics to calculate the values of the Root Mean Standard Deviation (RMSD). According to our results during the 150 ns production run all states A-I were populated some of them being the transition structures in which the dye was completely unstacked for a time interval of a few nanoseconds.

For more information see Figure 35, 36 and Table 16, 17.

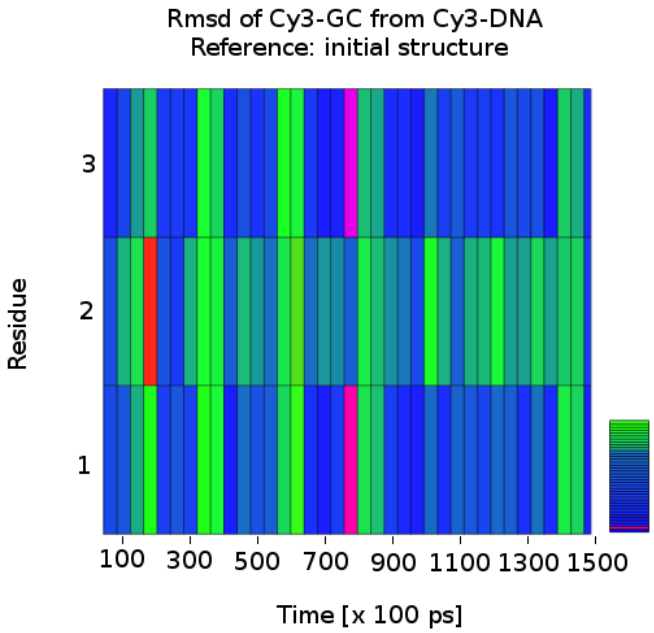


Fig. 35 *3D-RMSd graph for Cy3 residue (denoted as residue 1 on the y-axis) , residue C1 (2) and residue G20 (3). The 2-D colorplot shows the changes of the RMSD values of the residues in the average structures during the production run. The RMSD values are differentiated by the colors (see a small color scale on the right side).*

State	RMSd value [Å]	Abundance [%]	State	RMSd value [Å]	Abundance [%]
A	0.81 ± 0.24	85.27	F	3.43 ± 0.55	1.87
B	1.48 ± 0.25	6.13	G	2.26 ± 0.18	0.80
C	2.04 ± 0.34	5.20	H	2.37 ± 0.25	0.53
D	2.81 ± 0.00	0.07	I	2.37 ± 0.00	0.07
E	1.57 ± 0.00	0.07			

Table 16 *The abundance of individual states of Cy3 dye and their average values of RMSd.*

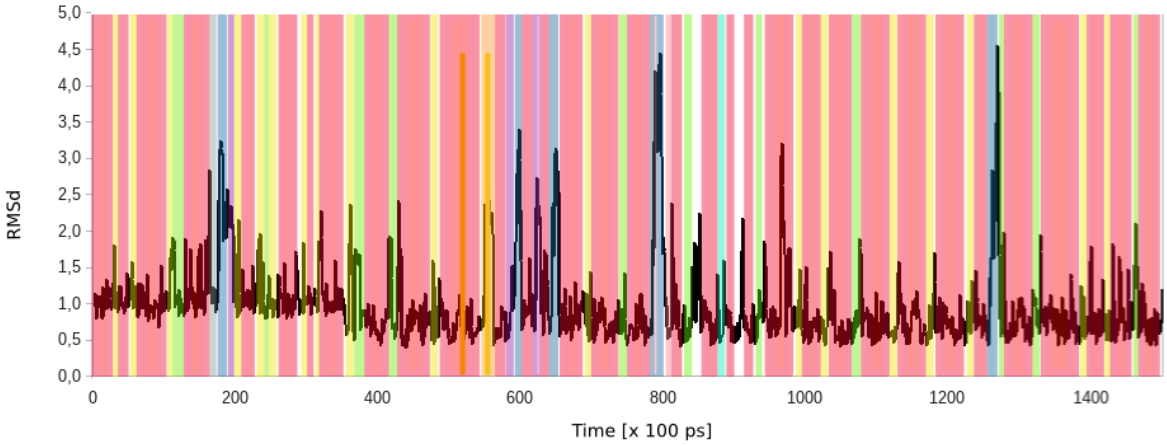


Fig. 36 *Changes of RMSD of the Cy3+GC (the first base pair) for the production run with respect to the initial structure . The occurrence of the conformers is distinguuished by different colors of the background. (A – red; B – yellow; C – green; D – grey; E – light blue; F – blue; G – orange; H – violet; I – pink).*

The RMSd values for the residues of the Cy3-GC structure were calculated using the vmdICE plugin[33] implemented in the VMD program. The average values for the residues are summarized in Table 17. The Cy3 dye is attached to the cytidine phosphate via a three carbon linker and it is the most flexible part of the system. The changes are shown in RMSd graph (see Figure 36).

The average values of RMSd for each residue of Cy3-GC	
Residue	***The average value of RMSd***
Cy3 dye	1.7 ± 0.9
DC_1	1.2 ± 0.7
DG_20	1.4 ± 0.6

Table 17 *The average values of RMSd for each residue of Cy3-GC.*

3.5.2 RMSd of the Cy5 Dye

The RMSd values were calculated for the average structures with respect to the initial structure (see Figure 37, Tables 18, 19).

State	RMSd value [Å]	Abundance [%]	State	RMSd value [Å]	Abundance [%]
A	1.6 ± 0.2	36.8	I	3.43 ± 0.55	0.5
B	2.5 ± 0.3	2.7	J	2.26 ± 0.18	54.7
D	2.9 ± 0.6	5.0	Not Class.	2.37 ± 0.00	0.3

Table 18 *The abundance of individual states of Cy5 dye and their average values of RMSd.*

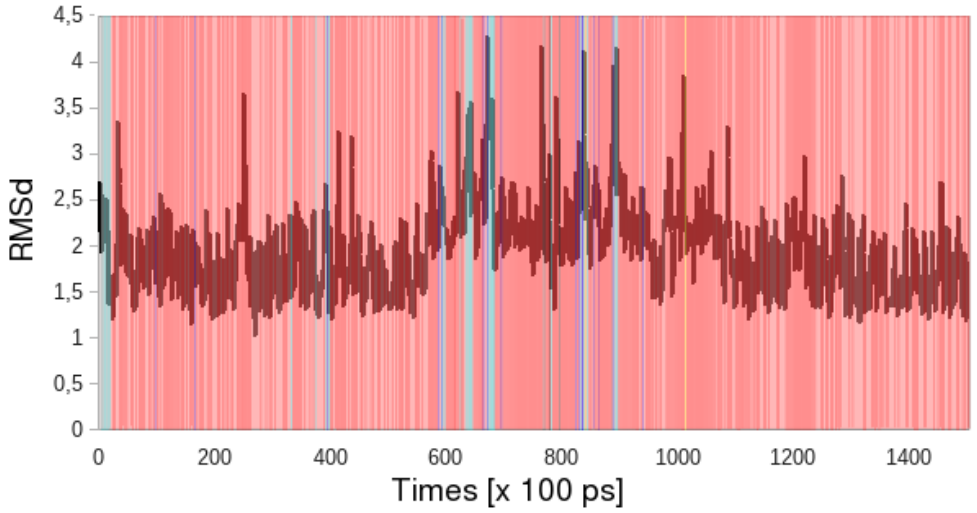


Fig. 37 *Changes of RMSD of the Cy5+GC (the first base pair) for the production run with respect to the initial structure . The occurrence of the conformers is distinguuished by different colors of the background. (A – red; B – blue; D – light blue; I – black; J – red; Not Classified – yellow).*

The average values of RMSd for each residue of Cy5-GC	
Residue	***The average value of RMSd***
Cy5 dye	2.1 ± 0.9
DC_1	1.5 ± 0.8
DG_20	1.0 ± 0.4

Table 19 *The average values of RMSd for each residue of Cy5-GC.*

3.5.3 Influence of the Attached Cy3 and Cy5 Dyes on the DNA Structure

We investigated the influence of Cy3 and Cy5 dyes on conformational changes of the DNA chain. The ten base pairs were analyzed by vmdICE plugin.

The influence of the probe on the DNA structure was explored by the comparison of the values of RMSd for single residues for the bare LBDNA and for DNA with attached Cy3 and Cy5 dyes. The same reference structure of DNA was used for the three simulations: the canonical B-DNA structure which was generated by the program NAB and used as a starting structure for LBDNA simulation and attached to the Cy3 and Cy5 dyes in the Cy3-DNA and Cy5-DNA simulations (see above).

With the cyanine dye attached the DNA chain shows only small changes with respect to the initial structure. The highest values of RMSd occur at the terminal base pairs DG_10, DC_11 and surprisingly for Cy3-DNA also on DA_6 and DA_13 residues (see Table 20).

It can be clearly seen that in presence of Cy3 and Cy5 dyes the DNA duplex is closer to the ideal B-DNA conformation. Stabilization is clear on the ends of the duplex where the dye prevented fraying.

Time	Cy3 DNA		Cy5 DNA		LBDNA	
	RMSd	σ_{RMSd}	RMSd	σ_{RMSd}	RMSd	σ_{RMSd}
DC_1	1.5	0.8	1.7	0.9	5.4	1.2
DC_2	1.0	0.5	1.3	0.6	3.8	0.9
DA_3	0.9	0.5	1.3	0.5	2.7	0.8
DC_4	0.7	0.3	1.2	0.5	3.1	0.8
DT_5	0.9	0.3	0.9	0.4	2.2	0.7
DA_6	1.5	0.6	1.6	0.6	2.2	0.6
DG_7	0.5	0.2	0.6	0.3	1.7	0.4
DT_8	0.7	0.3	0.9	0.4	1.2	0.4
DG_9	1.1	0.5	1.3	0.4	2.2	0.7
DG_10	1.3	0.5	3.2	0.7	3.1	0.9
DC_11	1.7	0.8	1.6	0.9	5.6	1.3
DC_12	1.1	0.5	1.4	0.6	3.8	0.9
DA_13	1.4	0.6	1.2	0.5	3.2	0.7
DC_14	0.9	0.4	0.8	0.4	3.0	0.7
DT_15	0.8	0.3	0.9	0.4	2.8	0.7
DA_16	0.9	0.4	0.9	0.4	1.8	0.6
DG_17	0.6	0.3	0.7	0.3	1.6	0.5
DT_18	0.6	0.3	0.7	0.3	1.4	0.5
DG_19	0.7	0.3	0.9	0.5	1.9	0.6
DG_20	1.8	0.7	1.1	0.6	3.3	1.0

Table 20 *- The average RMSd values for all the nucleotides of the DNA decamer calculated for the simulations of the Cy3-DNA, Cy5-DNA and LBDNA systems.*

3.6 Conformational Transitions

3.6.1 Conformational Transitions of the Cy3-DNA Complex.

As described in the previous section the state A represents the most stable conformation of the Cy3 dye with respect to DNA. Starting from A the dye may perform (1) a rotational movement roughly around the normal of the Cy3 complex plane (A → E transition), (2) a swinging movement with a partial unstacking (A → H transition), (3) a rotational-translational motion in which Cy3 unstacks from the guanine being moved roughly above the center of the GC base pair (A → I transition), and most importantly (4) an anti-clockwise rotational-translational motion which includes the sequence of transitions (A ↔ B ↔ C ↔ F).

The conformational transitions are very frequent. The first one can be found around the time

of 2.8 ns. The Cy3 molecule is very flexible with respect to DNA (Figure 38 and Table 21). We recognized nine types of possible conformational changes during 150 ns of the production run. The A ↔ B change is the most frequent transition in the production run (53.84 % from all conformational changes). The Cy3 dye slides along the short axis of the GC base pair and in the same time rotates around the DNA axis. A and B are the most probable conformations during the whole production run (probabilities 85.3 % and 6.1 %, respectively).

B and C conformations can be regarded as the intermediate structures between A and F conformations. The B → C transition is less frequent than B → A and it occurred firstly at the time of 11 ns. It means that B conformation will change back to A conformation with 60.7 % probability and with 39.3 % probability it will unstack to form the C conformer. B ↔ C transition represents 25.18 % of all transitions.

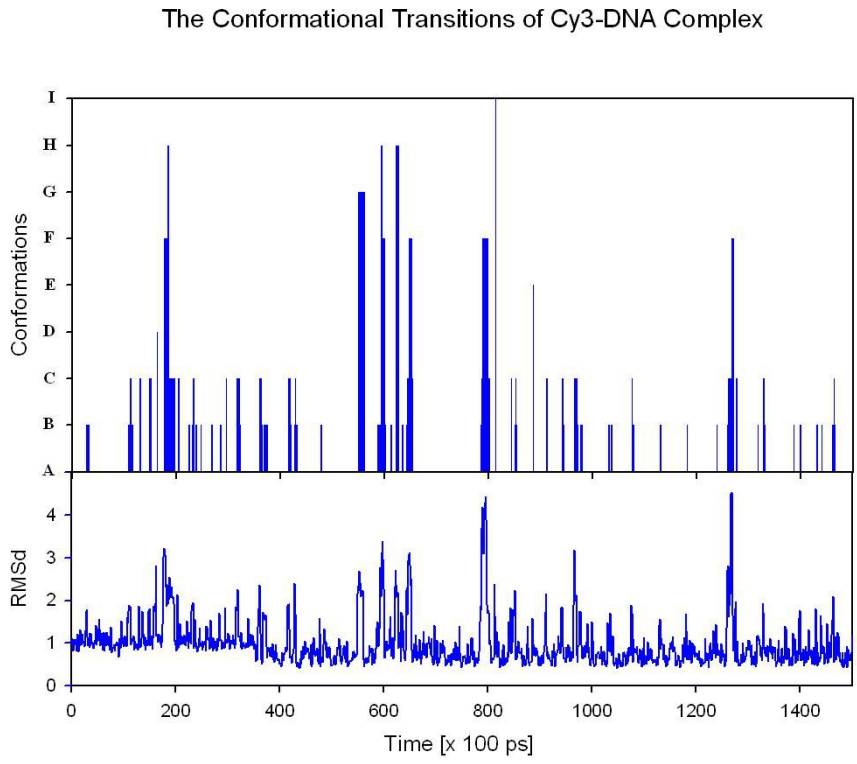


Fig. 38 *The bar chart which shows the occurrence of the conformationsin the Cy3-DNA system (upper part) and the RMSd values (lower part).*

Finally the conformer C will change back to B with the highest probability (78.5 %). The

conformational change to F represents the termination of the rotational-translational movement in the anti-clockwise direction (transitions (A ↔ B ↔ C ↔ F)) and it occurred four times (i.e. with ca. 14 % probability). Unlike the other conformers the F conformer is fully stacked to DG_20 by its proximal indole ring. Its low occurrence is probably caused by an increased strain in the pseudoether linker compared to others conformers. With the least probability (2 cases ca. 7 %) the structure can make the rotational-translational movement in the 'A direction' but unlike the A conformer it remains mostly unstacked (too large tilt) and forms the H conformer (see Figure 39).

Transition	Percentage [%]	Transition	Percentage [%]
A → B	27.97	D → A	0.70
B → A	25.87	A → G	0.70
B → C	16.78	G → A	0.70
C → B	15.38	A → H	0.70
C → F	2.80	H → A	0.70
F → C	2.80	A → I	0.70
C → H	1.40	I → A	0.70
H → C	1.40	A → E	0.70
A → D	0.70	E → A	0.70

Table 21 *The probability of the conformational transitions in the Cy3-DNA system.*

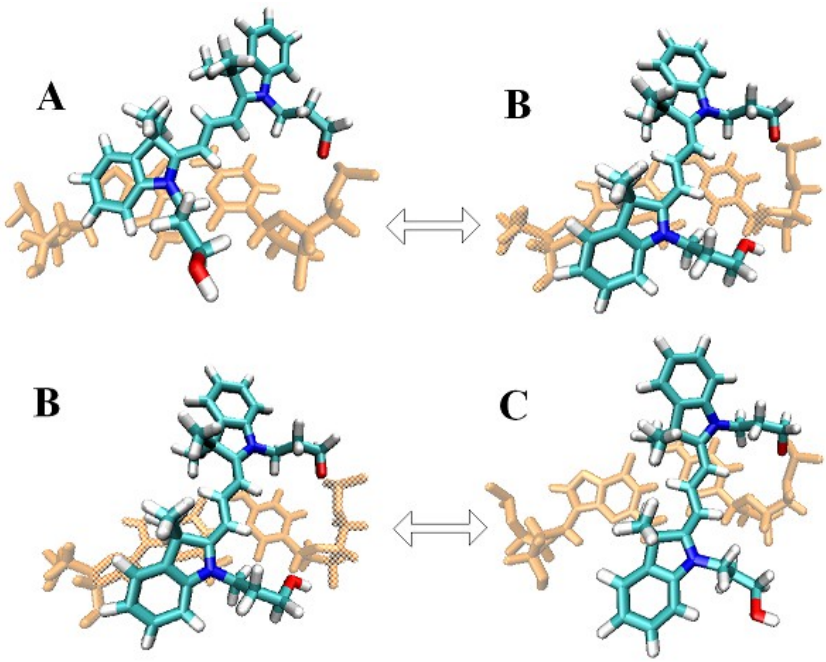


Fig. 39 *The most frequent conformational transitions of Cy3-DNA complex.*

3.6.2 Conformational Transitions of the Cy5-DNA Complex

The most stable conformations of Cy5 dye are represented by the states A and J. The differences between these two states are small and A and J differ in the values of slide and shift which change the extent of the stacking between guanine and indole rings: in the A state the both rings of the non-bonded indole are stacked just above guanine while in case of the J state only six-membered indole ring is stacked with guanine. The A ↔ J transition is the most frequent one in the production run (76.4 % of all transitions, see Figure 40). A and J conformations occur with the similar probabilities of 36.8 % and 54.7 %, respectively.

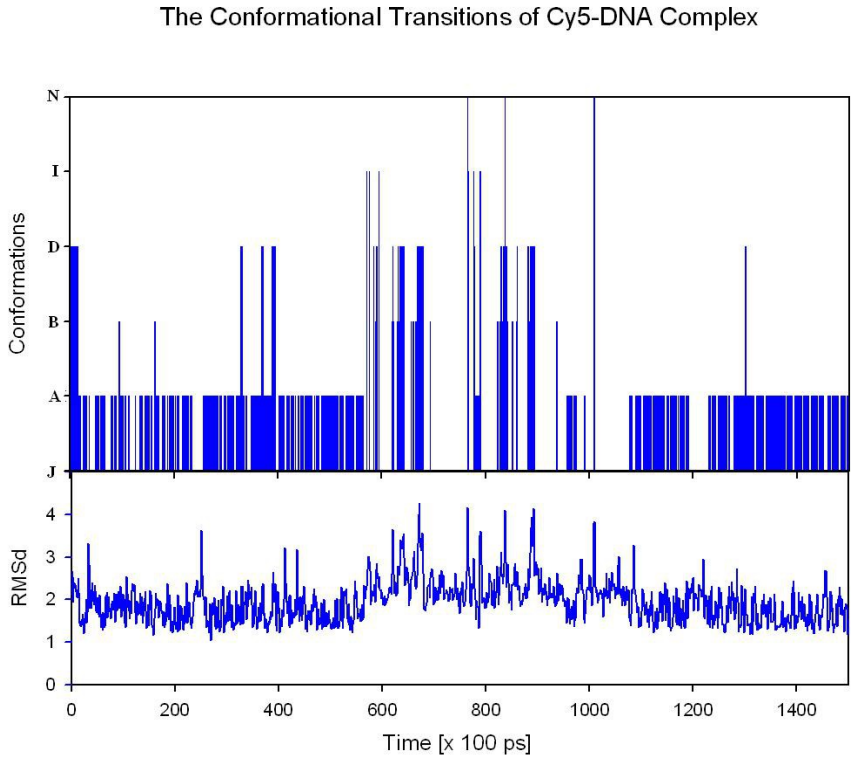


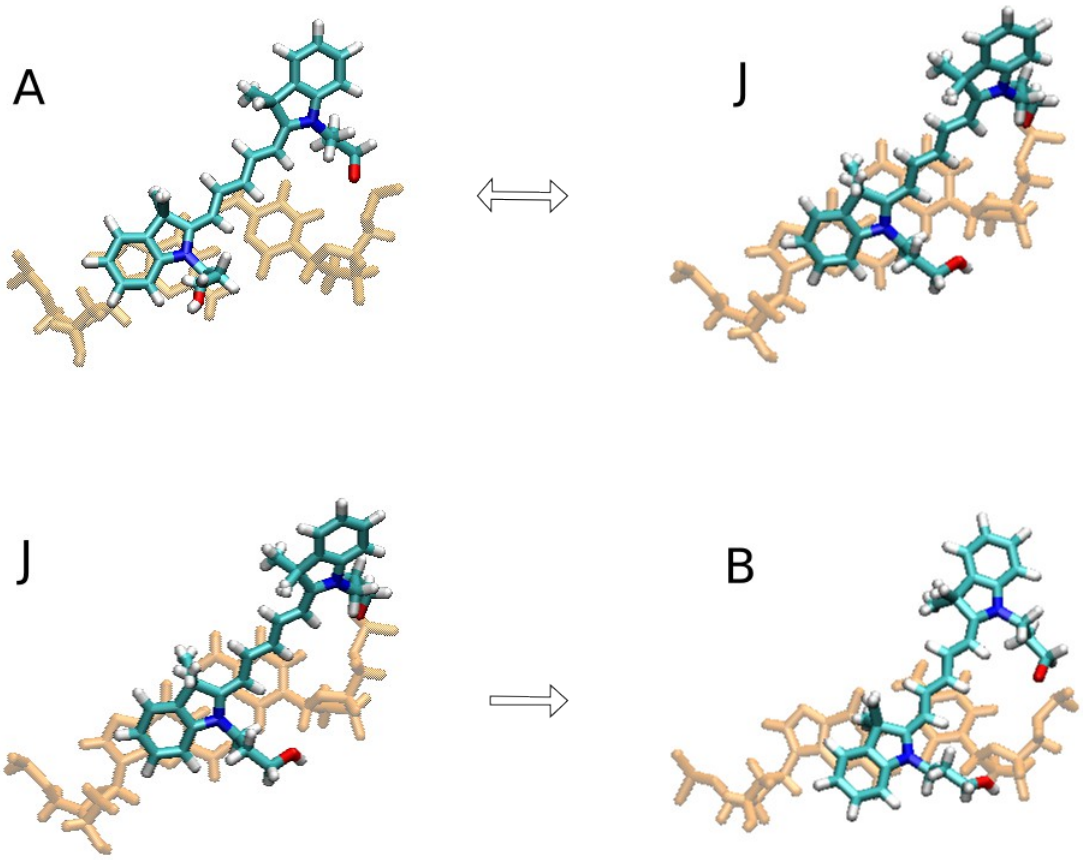
Fig. 40 *The bar chart which shows the occurrence of the conformationsin the Cy5-DNA system (upper part) and the RMSd values (lower part)*

Starting form the conformation J the Cy5 dye may perform similar movements as the Cy3 dye: (1) a rotational-translational motion in which Cy5 unstacked from the guanine being moved roughly above the center of the GC base pair (J → I transition, see Figure 41), (2) an anti-clockwise rotational-translational motion which includes the sequence of transitions (J ↔ B ↔ D). The rotational movement roughly around the normal of its plane (the "swinging" movement) with partial unstacking was not observed.

J conformation can be changed into A, B, D, and I conformations. The transitions occurred with the following probabilities: J → A (84.1 %, translational motion), J → B (9.9 %, anti-clockwise rotational motion), J → D (2.1 %, anti-clockwise rotational motion), J → I (0.6 %, translational motion).

Transition	Percentage [%]	Transition	Percentage [%]
A → J	38.20	D → A	0.30
J → A	38.20	A → B	0.30
J → B	4.50	A → D	0.30
D → J	3.30	D → I	0.30
B → D	3.30	I → D	0.30
B → J	2.40	A → I	0.30
J → D	2.10		
D → B	2.10		
I → J	1.20		
B → A	0.60		
J → I	0.60		

Table 22 *The probability of conformational transitions in the Cy3-DNA system.*



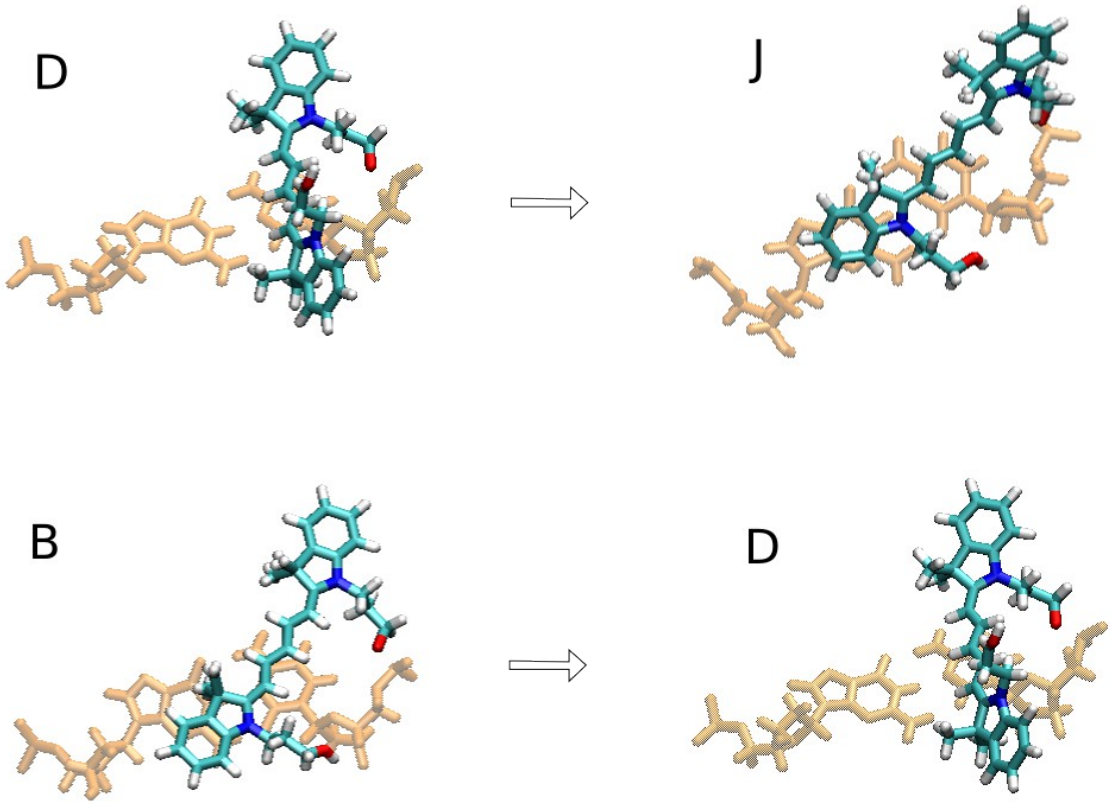


Fig. 41 *The most frequent conformational transitions in the Cy5-DNA complex.*

B conformation can be regarded as the intermediate structure between J and D conformations. The B → J transition is less frequent than B → D and it occurs firstly at the time of 62,2 ns. It means that B conformation will change to D conformation with 52.4 % probability and only with 38.1 % probability will change to the most probable J conformer. However these numbers can be influenced by the relatively low number of transitions observed from the structure B (totally 21 transitions). B ↔ D and B ↔ J transitions represents 3.3 % and 6.9 % of all transitions, respectively (see Table 22).

4 Discussion

The mutual position of the Cy3 and Cy5 dyes bound to DNA is a critical information for determination of the parameters for fluorescence resonance energy transfer (FRET)- mainly of the orientation factor κ^2.

However, the fluorophores are highly flexible and their mutual orientation changes very quickly. Our results will enable calculation of the mean value of the orientation factor κ^2.In DNA κ^2 is a function of distance (or the number of base pairs between the dyes) and the local DNA conformation.

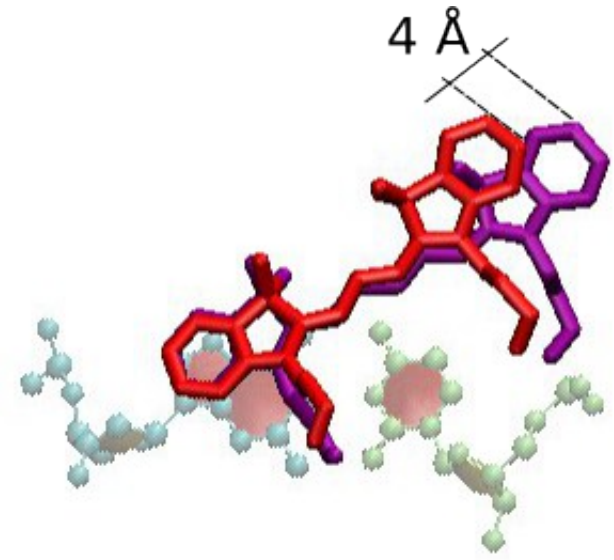


Fig. 42 *Comparison of the positions of the dye in the A conformations of the Cy3-DNA and Cy5-DNA complexes with respect to the terminal DC_1-DG_20 base pair. (Cy3 dye – red; Cy5 dye – purple). While the distal rings of the two dyes are in almost the same positions the proximal indole ring of Cy5 is shifted away by about 4 Å compared to Cy3.*

The differences in the behaviour between the Cy3-DNA and Cy5-DNA complexes can be attributed to the length of the polymethine linker between the distal and proximal indole rings. Two indole rings of the Cy5 molecule are connected by the pentamethine linker, while the indole rings of the Cy3 dye are connected by the trimethine linker.

For the most cases the distal indole ring is stacked on the base of the terminal DG_20 residue. However, the longer linker of Cy5 dye causes that the position of proximal indole rings differ. While the proximal indole ring of Cy3 dye can be partially stacked on the terminal DC_1 residue or even on the DG_20 residue. The proximal indole ring of Cy5 is shifted too much away from the terminal base pair (Figure 42).

The Cy5 dye attached to DNA complex is less flexible than the Cy3 dye. For the Cy5-DNA

complex five different conformations were recognized, while for the Cy3-DNA nine conformations were determined. Cy3-DNA conformers included also structures with stacked proximal indole ring. These conformers were not observed for the Cy5-DNA complex.

For the Cy5-DNA complex only transitions which include translational displacement of the dye are possible. The purely rotational movements such as the rotational movement roughly around the normal of its plane and a swinging movement with a partial unstacking were not observed.

Finally we verified the influence of the cyanine dyes on the structure of the DNA decamer and we compared the structures of the DNA molecule in the three systems: Cy3-DNA, Cy5-DNA and bare LBDNA. The DNA base sequence was the same for all the systems.

For each step of the MD trajectories we computed base-pair parameters, step parameters and RMSd values for every base. From the comparison of average values of RMSd for the base pairs we can see a stabilizing effect of the Cy3 and Cy5 dyes on the B-DNA conformation of the duplex structure. The RMSd values for the terminal bases were decreased by up to 4.0 Å in presence of the dye (Table 20). For internal bases the stabilization effect of the dye was smaller but still significant (RMSD values being by about 1 Å lower).

5 Conclusions

In this work we investigated the structure and dynamics of molecular complexes which involve the fluorescence probes Cy3 and Cy5 dyes bound to DNA decamer. We used the explicit water solvent molecules and Cl^- and Na^+ ions for charge neutralization of the system. To study dynamical and structural properties of the complexes we used the molecular dynamics calculations. *Ab initio* quantum chemical calculations were used for determination of rotational barriers of the bonds of the polymethine linkers which were used for the parametrization of the torsions. The cyanine dyes were connected covalently to the 5'-phosphate terminus of DNA via N9' attached three-carbon linker. The flexibility of Cy3 and Cy5 dyes enabled the systems to adopt optimal stacking geometries.

Our MD data have shown that in agreement with experimental evidence[34] the cyanine Cy3 and Cy5 dyes that are attached to 5'-C end of DNA becomes stacked via their distal indole ring over the terminal G base. However a few unstacking events were observed for both dyes during the 150 ns production run. The Cy3 dye was completely unstacked for the time interval of about 7.9 ns (5,3 % of the time), while Cy5 dye was unstacked for about 7.5 ns (5.0 % of the time).

Thus, the position of cyanine dyes with respect to the terminal base pair is highly flexible: for Cy3 and Cy5 dyes we recognized nine and five different conformations, respectively. The Cy3-DNA is more flexible since the proximal indole ring can be also involved in stacking interactions. All conformations were characterized by the base pair step parameters (*Twist, Shift, Slide, Tilt, Roll* and *Rise*).

Both probes have stabilization effect on the B-DNA conformation of the helix under our experimental conditions. If the cyanine dye is attached to the terminal cytidine, the average RMSd values for all the bases are rapidly decreased compared to bare LBDNA simulation (see Table 20). The stabilization effect of the Cy3 (Cy5) dye was largest for the first base pair for which the RMSd values decreased substantially compared to LBDNA. To see stabilization effects of the dyes in a more detail the analysis of the base pair and base pair step parameters is necessary.

References

1. Li W., Ying X., *Applied Bioinformatics*, **2006**, 5, pp. 181-186
2. Wernersson R.,Nielsen H. B., *Nucleic Acids Research,* **2005**, 33, pp. W611-W615
3. The Amber 11 authors: D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B.P. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P.A. Kollman.
4. Cornell W. D., Cieplak P., Bayly Ch. I., Gould I. R., Merz K. M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J.W., Kollman P.A., *Journal of the American Chemical Society,* **1995,** 117 (19), pp. 5179–5197
5. Wang et al., *Journal of the Computational Chemistry,* **2004,** 25 (9), pp. 1157-1174
6. Xiang-Jun Lu, Olson W. K., *Nature Protocols*, **2008**, 3 (7), pp. 1213-27
7. Xin Y., Olson W. K., *Nucleic Acids Research*, **2009**, 37, pp. D83-D88
8. Gu J., Bourne P. E., *Structural bioinformatics*, second edition, Wiley-Blackwell **2009**
9. Šponer J., Lankaš F., Leszczynski J., *Computational Studies of RNA and DNA*, Springer **2006**
10. Neidle S., *Principles of Nucleic Acid Structure.* London: 1st edn. Elsevier; **2008**
11. Kuhn, H., *Journal of Chemical Physics,* **1949**, 17, 1198–1212.
12. Zhang X. H., Wang L.-Y., Zhai G.-H., Wen Z.-Y., Zhang Z.-X., *Bulletin of the Korean Chemical Society,* **2007**, 28, pp. 2382-2388
13. Shindy H.A., El-Maghraby M.A., Eissa F. M., *Dyes and Pigments,* **2006**, 68, pp. 11-18
14. Harvey B., Levitus M., *Biophysical Journal*, **2008**, 96, pp. 401a-401a
15. Harvey B. J., Perez C., Levitus M., *Photochemical and Photobiological Sciences,* **2009**, ; 8(8), pp. 1105-10
16. Šimková E., Staněk D., *International Journal of the Molecular Sciences,* **2012**, 13, pp. 14929-14945
17. Rietdorf J., Gadella T. W. J., *Microscopy techniques,* Springer, **2005**
18. Dolghih E., Roitberg A. E., Krause J. L., *Journal of Physical Chemistry A,* **2007**,190 (2–3), pp. 321–327

19. Spiriti J., Binder J. K., Levitus M. , van der Vaart A., *Biophysical Journal*, **2011**, 100(4), pp. 1049-57

20. Heilemann M., Margeat E., Kasper R., Sauer M., Tinnefeld P., *Journal of the American Chemical Society*, **2005**, 127 (11), pp. 3801–3806

21. Urnavicius et al., *Biophysical Journal*, **2011**, 102, pp. 561–568

22. Ouellet et al., *Biophysical Journal*, **2012**, 101, pp. 1148–1154

23. Singh M.K., *Physical Chemistry Chemical Physics*, **2009**, 11(33) pp. 7225-30.

24. Huang Z., Ji D., Xia A., Koberling F., Patting M., Erdmann R., *Journal of the American Chemical Society*, **2005**, 127(22), pp. 8064-6

25.Jia K., Wan Y., Xia A., Li S., Gong F., Yang G., *Journal of Physical Chemistry A*, **2007**, *111* (9), pp. 1593–1597

26. van Hal N.L.W. et al., *Journal of Biotechnology*, **2000**, 78, pp. 271–280

27. Nuwaysir E. F., Huang W., Albert T. J., et al., *Genome Research*, **2002,** 12, pp. 1749-1755

28. Fodor S., Read J. L., Pirrung M. C., Stryer L., Tsai L. A., and Solas D., *Science*, **1991**, 251, pp. 767–773

29. Iqbal A., Wang L., Thompson K. C., Lilley D. M. J., Norman D. G., *Biochemistry*, **2008**, 47, pp. 7857-7862

30. Gaussian 09, Revision **A.1**, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, **2009**.

31. Cieplak, P., Cornell, W.D., Bayly, C., Kollman, P.A., *Journal of the Computational*

Chemistry, **1995**, 16(11), pp. 1357-1377

32. Cooper V. R., Thonhauser T. at col., *Journal of the American Chemical Society*, **2008**, 130, pp. 1304-1308
33. Knapp B., Lederer N., Omasits U. at al., *Journal of the Computational Chemistry*, **2010**, 31(16), pp. 2868-2873
34. Norman D.G., Grainger R. J., Uhrín D., Lilley D. M. J., *Biochemistry*, **2000**, 39, pp. 6317-6324

Appendices

Appendix 1 – Charge Derivation, Starting Structure

Starting structure was based on avaraged minimized coordinates of the hybrid Cy3-DNA-Cy5 complex which were sent by dr. David G. Norman from The University of Dundee. Starting structure was used for a preparation of Sander input files *prmtop* and *inpcrd*. Because the fluorescein dyes are non-standard residues, the atomic charges cannot be taken from the Amber or GAFF force field but were calculated to reproduce electrostatic potential (ESP) of the molecule by the Restrained Electrostatic Potential (RESP) method.
Firstly the geometry was optimized by the B3LYP/6-31G* method and then electrostatic potential (ESP) was evaluated using HF/6-31G* level of theory with the following line in the Gaussian 09 input file:

#P HF/6-31G* SCF=Tight Pop=MK IOp(6/33=2,6/41=10,6/42=17)

The Gaussian ESP data were converted into the RESP format using the following *bash script* (for Cygwin - Linux-like environment in Windows):

esp.sh
#!/bin/csh
g77 readit.f
grep "Atomic Center " $1 > a
grep "ESP Fit" $1 > b
grep "Fit " $1 > c
./a.exe
rm -f a b c a.exe readit.o

Table A1 *Bash script for the conversion of the Gaussian ESP data into the RESP format.*

The *readit.f* program can be downloaded from the www address: http://ambermd.org/tutorials/advanced/tutorial1/files/readit.f. Furthermore the conversion script needed the numbers of ESP centers and ESP Fit centers as the input. After running the

script we obtained the **.dat* files.

The charge derivation by the RESP program included two basic steps. At first the charges of all atoms were allowed to vary and then the charges of all degenerate hydrogen atoms were constrained to have the same value. The second step was important for realistic evaluation of the charges of methyl and methylene groups of Cy3 and Cy5 dyes.

cy5-resp.in

Page 1	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7
cy5-linker-resp run	6 0	6 0	1 0	1 0	1 0	1 0
#1	6 0	6 0	1 0	1 0	1 0	1 0
&cntrl	7 0	6 0	1 0	1 0	1 0	8 -0.69210
nmol=1,	6 0	6 0	1 0	1 0	1 0	1 74 1 75 1 76 1 77
ihfree=1,	6 0	6 0	1 0	1 0	1 0	1 78 1 79 1 80 1 81
qwt=0.0005,	6 0	6 0	1 0	1 0	1 0	
iqopt=2,	6 0	6 0	1 0	1 0	1 0	
/	6 0	6 0	1 0	1 0	1 0	
	6 0	6 0	1 0	1 0	8 0	
1.0	6 0	6 0	1 0	1 0	15 -1	
cy5-linker	6 0	6 0	1 0	1 0	8 -1	
0 81	7 0	6 0	1 0	1 0	8 -1	
6 0	6 0	6 0	1 0	1 0	6 0	
6 0	6 0	6 0	1 0	1 0	8 -1	
6 0	6 0	8 0	1 0	1 0	1 0	
6 0						

Table A2 *The input files for RESP.*

Also we created a qin file:

cy5-resp.qin

0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 1.165900 -0.776100 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 -0.495400 0.000000 -0.776100
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000

Table A3 *Atom's charge were read from the qin file and constrained to this value.*

cy5-resp2.in						
str. 1	str. 2	str. 3	str. 4	str. 5	str. 6	str. 7
Linker Trial 5	6 -1	6 -1	1 -1	1 0	1 61	1 79
Stage2	7 -1	6 -1	1 -1	1 51	1 0	8 -0.69210
&cntrl	6 -1	6 -1	1 -1	1 0	1 67	1 74 1 75 1 76
ihfree=1,	6 -1	6 0	1 -1	1 53	1 0	1 77 1 78 1 79
qwt=0.001,	6 -1	6 0	1 -1	1 0	1 69	1 80 1 81
iqopt=2	6 -1	6 0	1 -1	1 55	1 0	
/	6 -1	6 0	1 -1	1 55	1 71	
	6 -1	6 27	1 -1	1 55	8 -1	
1.0	6 -1	6 0	1 -1	1 55	15 -1	
cy5-linker	6 -1	6 29	1 -1	1 55	8 -1	
0 81	7 -1	6 0	1 -1	1 0	8 -1	
6 -1	6 -1	6 0	1 -1	1 61	6 0	
6 -1	6 -1	6 0	1 -1	1 61	8 -1	
6 -1	6 -1	8 -1	1 0	1 61	1 0	
6 -1	6 -1	1 -1	1 49	1 61	1 79	
6 -1						

Table A4 *The input files for RESP – the second part.*

The final evaluation of the charges was performed by the command:

*$AMBERHOME/exe/resp -O -i *.in -o *.out -p *.pch -t *.chg -q *.qin -e *.dat*

Appendix 2 - The Conversion of Cy3/Cy5 Dyes to ADE- ADE base pair (output file of this Fortran script is compatible with input file for the X3DNA program)

The conversion is performed by the bash script VMDscript that calls the Fortran 90 program CYTVMD5.f90.

VMDscript:

```
#!/bin/sh

echo 1 > nazvy.txt

for i in `seq 1  1500`
do
index=$(expr "$i" + 1)
echo cyanine-$i.pdb adenine1-$i.pdb adenine2-$i.pdb>> nazvy.txt
done

for a in `seq 1  1`
do
index=$(expr "$a" + 1)
sed -e '/TER/d' Cl_average-$a.pdb > average-$a.pdb
awk 'NR>0 && NR<34' average-$a.pdb > cyanine-$a.pdb
awk 'NR>69 && NR<704' average-$a.pdb > dna_1-$a.pdb
sed 's|DG3|GUA|g;s|DC |CYT|g;s|DC5|CYT|g;s|DA |ADE|g;s|DT |THY|g;s|DG |GUA|g' dna_1-$a.pdb > dna-
$a.pdb
done

./skript/CYTVMD5.out

for a in `seq 1  1`
do
index=$(expr "$a" + 1)
cat adenine2-$a.pdb adenine1-$a.pdb dna-$a.pdb > Cy3-dna-$a.pdb
find_pair -t Cy3-dna-$a.pdb Cy3-dna-$a.inp
analyze Cy3-dna-$a.inp
done
```

CYTVMD5.f90 – The Conversion of the Proximal Indole Ring to Adenine

```
                SELECT CASE(typprvku)
                      CASE('N')
                        typprvku='N9'
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek
                      CASE('C')
                        typprvku='N1'
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek
                      CASE('C1')
                        typprvku='C2'
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek
                      CASE('C2')
                        typprvku='N3'
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek
                      CASE('C3')
                        typprvku='C4'
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek
                      CASE('C4')
                        typprvku='C5'
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek
                      CASE('C6')
                        typprvku='N7'
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek
                      CASE('C7')
                        typprvku='C8'
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek
                      CASE('C5')
                        typprvku='C6'
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek
                      CASE('O1')
                        typprvku="O3'"
                        WRITE(5,'(A4,A5,I2,A2,A4,A2,A3,A2,I2,A4,3F8.3,2F6.2)')
atom,m_5,poradiatomu,m_2,typprvku,m_2,nazevbarviva,m_2,&
                           cisloretezce,m_4,x,y,z,sloupec1,sloupec2!,m_11,prvek

                      END SELECT
```

Appendix 3 - Backbone Torsion Angles of DG_19 nucleotide in Cy3-DNA Complex

DG_19	α	β	γ	δ	ε	ζ	χ
A	-74.4 ± 8.4	169.2 ± 31.0	53.7 ± 5.7	127.9 ± 8.5	-148.7 ± 36.2	-48.4 ± 104.1	-105.8 ± 12.4
B	-73.2 ± 5.3	171.2 ± 3.3	55.1 ± 4.2	127.7 ± 6.4	-150.1 ± 65.8	-75.3 ± 74.5	-109.3 ± 10.9
C	-74.5 ± 4.8	171.7 ± 2.7	52.7 ± 4.0	127.7 ± 6.8	-154.6 ± 45.7	-64.0 ± 82.9	-107.2 ± 11.5
D	-77.1	172.1	49.4	126.1	-167.2	-89	-95.6
E	-73.8	171.8	53.8	120.8	-176.9	-95.6	-116.7
F	-73.4 ± 5.1	172.4 ± 3.0	53.2 ± 3.7	128.4 ± 8.2	-145.9 ± 33.8	-34.9 ± 117.2	-107.0 ± 16.1
G	-77.4 ± 3.7	172.5 ± 1.9	51.2 ± 2.2	134.7 ± 3.5	-142.2 ± 2.6	-72.7 ± 4.1	-86.2 ± 6.0
H	-76.2 ± 4.7	128.0 ± 121.7	51.3 ± 4.0	132.3 ± 3.7	-137.7 ± 17.6	-82.4 ± 28.4	-91.7 ± 12.3
I	-75.8	172.1	52.2	122.9	-177	-93.8	-111.6

Table A5 *The average values of dihedral angles of DG_19 nucleotide for all conformations which were found during the MD production run.*

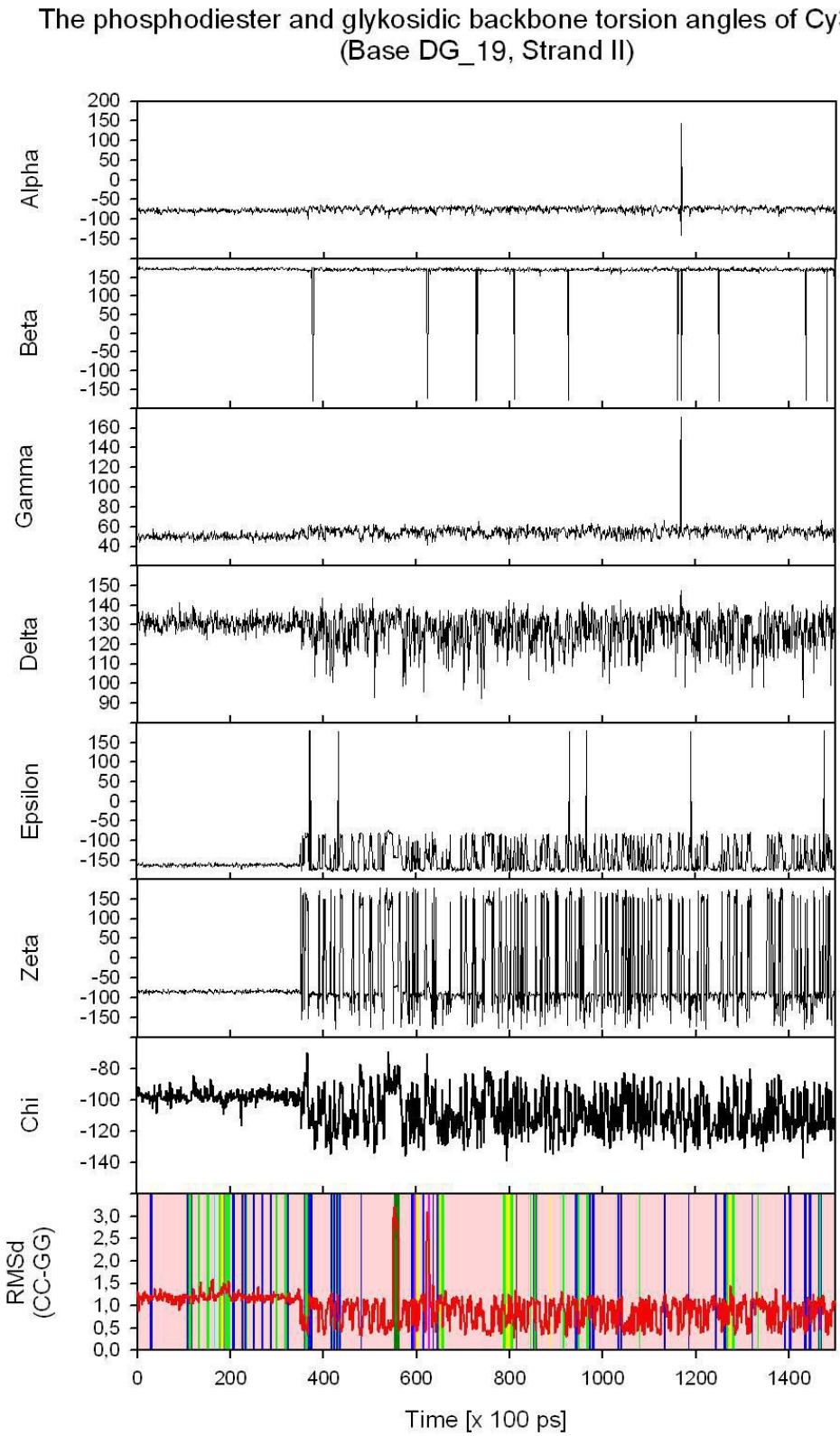


Fig. A1 *The phosphodiester and glycosidic backbone torsion angles of Cy3-DNA complex for DC_2 base of strand I.*

Appendix 4 - Influence of the Environment on the Specificity of the Mg(II) Binding to Uracil

Ingrid Romancová, Zdeněk Chval, and Milan Předota

J. Phys. Chem. A **116**(7):1786-93 (2012), PMID 22263542

Interactions of uracil with the Mg(2+) ion were studied theoretically in the gas phase and in solution. The bare Mg(2+) prefers bidentate N-C═O binding sites stabilizing rare keto-enol forms of the base. Hydration and/or phosphate binding of the Mg(2+) ion shield its positive charge, which leads to preference of monodentate binding to the oxygen keto atoms, shifting fully the equilibrium between the tautomers back toward the canonical diketo tautomer. In solution, a direct inner-sphere metal binding to uracil is not clearly advantageous compared to the outer-sphere metal binding. Similar trends were also obtained for the Ca(2+) ion. Results are supported by the natural bond orbital (NBO) and atoms in molecule (AIM) analyses and the combined extended transition-state energy decomposition analysis and natural orbitals for chemical valence (ETS-NOCV).

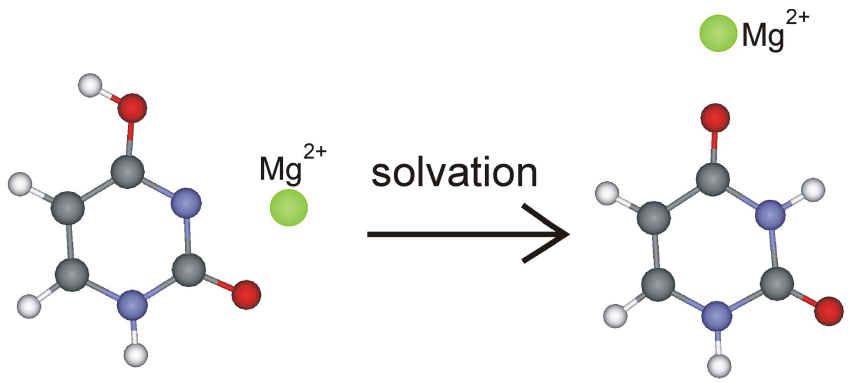


Fig. A2 *Influence of the environment on the specificity of the Mg(II) binding to uracil*