Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary material: Deprotonation mechanism of a single-stranded DNA i-motif

Jens Smiatek*,†,‡ and Andreas Heuer*,†

Institut für Physikalische Chemie, Universität Münster, D-48149 Münster, Germany, and Institut für Computerphysik, Universität Stuttgart, D-70569 Stuttgart, Germany

E-mail: smiatek@icp.uni-stuttgart.de; andheuer@uni-muenster.de

Simulation Details

We have performed all-atom Molecular Dynamics simulations of the DNA i-motif in explicit solvent at 300 K by the GROMACS software package[10] with the implementation of the AMBER03 force field ports[11,12] which include the nucleic acid parameters of the parm99 force field[13]. All parameter values are identical to recent simulational studies and have been in detail discussed in Refs.[3,4]. The single DNA strand consists of 22 nucleic acids that are given by the sequence 5′-CCC-[TAA-CCC]$_3$-T-3′ where T, A and C denote thymine, adenine and cytosine. We modeled this structure which is directly related to the sequence used in Refs.[1,2] by the PDB entry 1ELN[14] where we only changed uracil to thymine.

The cubic simulation box with periodic boundary conditions had a box length of 5.41 nm and was filled with 5495 TIP3P water molecules[15]. The negative net charge of $q = -22e$ of the DNA

*To whom correspondence should be addressed
†Institut für Physikalische Chemie, Universität Münster, D-48149 Münster, Germany
‡Institut für Computerphysik, Universität Stuttgart, D-70569 Stuttgart, Germany

molecule was compensated by 16 sodium ions and 6 protons to form the hemiprotonated cytidine base pairs. The protons were modeled by using the parameters of hydrogen atoms (Atomtype H2) in the corresponding parm99 force field[13] where we changed the partial charge to $-1e$. It has to be noted that our approach of using explicit protons is identical to earlier simulations studies[16,17] but with the neglect of unphysical covalent hydrogen-nitrogen bonds. As a prerequisite of the simulations, we deposited the maximum number of six protons between the cytosine pairs in front of the corresponding nitrogen atoms (**3**-position). After energy minimization (Minimum force on atoms: 500 KJ $\mathrm{mol}^{-1}\mathrm{nm}^{-1}$, step size 0.002 nm) by the steepest descend method, we derived the corresponding structure of the DNA i-motif which is shown in the main text.

For the following simulations, we applied a Nose-Hoover thermostat to keep the temperature constant (NVT ensemble). All bonds have been constrained by the LINCS algorithm. Electrostatics have been calculated by the PME method and the time step was 2 fs. The initial warm up phase of 1 ns has been performed by keeping the position of the DNA and the protons restrained.

The study of the unfolding motion and the associated deprotonation process has been performed by the original Metadynamics method presented in Refs.[5,6] which deposits history-dependent potential hills on the trajectory of the considered reaction coordinate. The biasing energy allows to overcome energetic barriers and to accelerate the occurrence of rare events in addition to the evaluation of the free energy landscapes and the determination of the lowest free energy pathways[7,8]. This occurrence of realistic pathways has been specifically proven for the i-motif by the comparison of high temperature and biased simulation runs[4]. To achieve more accuracy for the determination of the deprotonation mechanism and the unfolding pathway, we have performed 10 Metadynamics simulations each with 50 ns length. All biasing simulations at 300 K have been conducted by the software PLUMED[20]. The Gaussian potential energy hills were deposited each 2 ps with a height of 0.25 kJ/mol and a width of 0.25 nm. The corresponding reaction coordinates were the distance between nucleobase C1 and T22 (end-to-end distance) and the distance between the center-of-mass for the two combined nucleobases C1 and T22 which was in detail described in Ref.[3].

Electrostatic energy: Release of proton 4

As it was discussed in the main article, the initial release of proton 4 out of the energy-minimized DNA i-motif with the new protonation state as shown in 1 can be explained by the strong electrostatic interactions between the protons. To validate this assumption, we have focused on the initial distances between protons and nucleobase atoms. With regard to the specific position of proton 4 in the initial configuration, it has to be stated that protons 4 and 5 have a minimal distance of $r^{min}_{H^+(4)-H^+(5)} \approx 0.16$ nm where all other proton distances are around $r^{min}_{H^+-H^+} \approx 0.19$ nm. Furthermore, the minimal distance between proton 4 and nucleobase atoms is given by $r^{min}_{H^+-DNA} \approx 0.31$ nm which is slightly larger compared to the other protons. The function $\psi(r)^{min}_{H^+} = (1/r^{min}_{H^+-H^+}) - (1/r^{min}_{H^+-DNA})$ allows to yield a rough estimate of the electrostatic energy for each proton. The corresponding results are shown in 2. It is clearly evident that the most unfavorable electrostatic energy is given for proton 4. Thus, the release of proton 4 can be interpreted as an obvious consequence and it can be speculated if this rearrangement also occurs with regard to the more realistic presence of H_3O^+ ions.

However, the remaining five protons form a new protonation state where one proton coordinates roughly four nucleobases in an intercalating scheme as it can be seen in 1.

Correlation between individual protons and the corresponding proton-DNA atom contact pairs

The nearly combined release of protons 1 and 2 is also indicated by the Pearson correlation coefficient r for the proton nucleobase DNA contact pair number N_P for individual protons. The Pearson correlation coefficient was calculated as shown below such that we have taken all ten biased simulations into account

$$r = \frac{\sum_t^\tau (N_{P,i}(t) - < N_{P,i} >)(N_{P,j}(t) - < N_{P,j} >)}{\sqrt{\sum_t^\tau (N_{P,i}(t) - < N_{P,i} >)^2 \sum_t^\tau (N_{P,i}(t) - < N_{P,i} >)^2}} \quad (1)$$

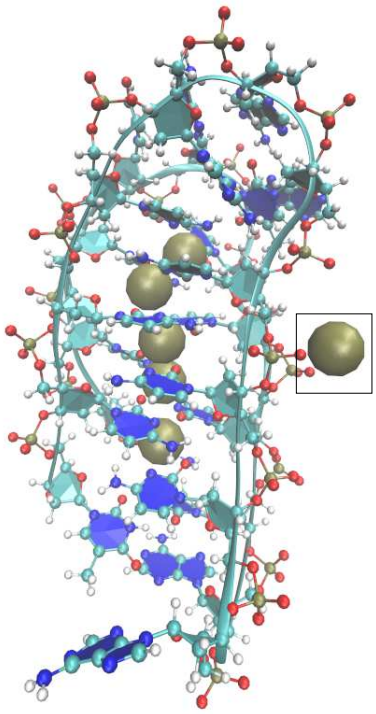

Figure 1: Proton configuration for the DNA i-motif after 100 ps. Proton 4 interacts with phosphate groups (as marked by the box) at the backbone of the DNA.

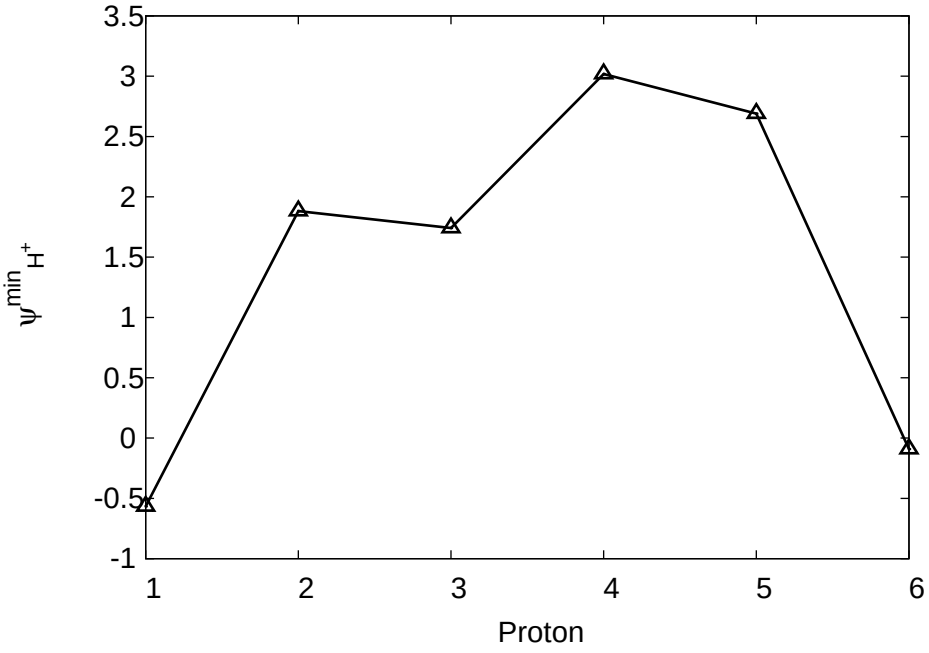

Figure 2: Estimated electrostatic potential energy $\psi(r)^{min}_{H^+}$ as defined in the text for each proton.

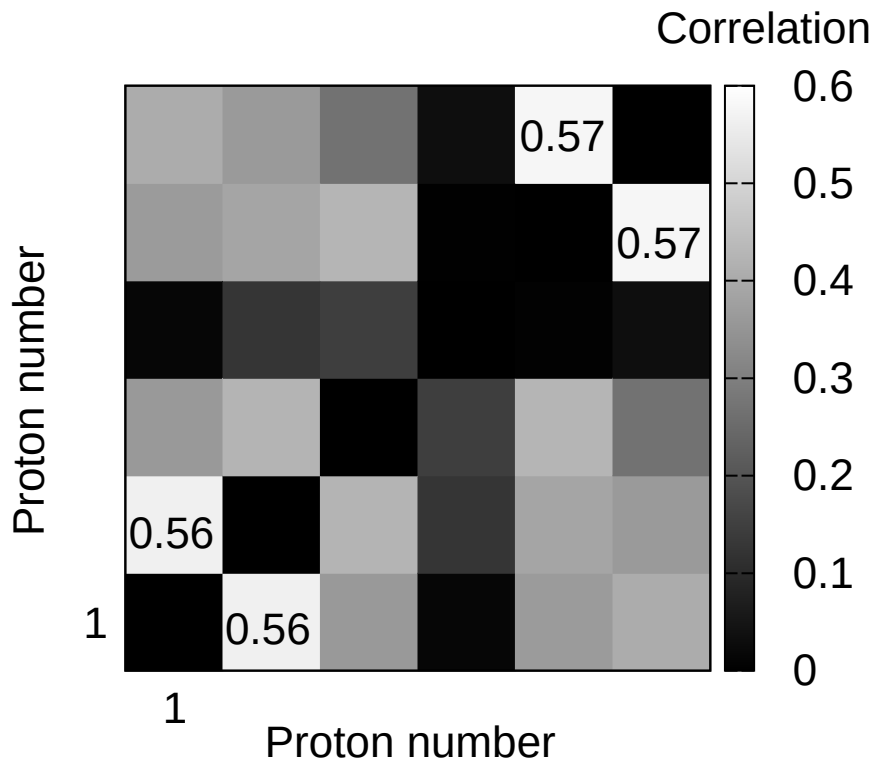

Figure 3: Pearson correlation coefficient r for the proton nucleobase contact number $N_P(t)$ according to 1 for all proton combinations.

where i and j denote different protons and $< N_{P,i/j} >$ the average number of proton-DNA atom contact pairs. The corresponding results are presented in 3. It can be clearly seen that high correlation coefficients $r > 0.5$ can be observed for the proton combinations 1,2 and 5 and 6. These values clearly indicate that a concerted deprotonation mechanism can be observed for these proton pairs. The high correlation between proton 1 and 2 can be mainly interpreted in terms of a concerted release whereas the embedded positions of protons 5 and 6 result in a nearly identical fluctuation behavior which explains the high values for the correlation coefficients. One can assume that the initial release of proton 1 destabilizes the binding properties of proton 2 due to more flexible DNA strands.

Free energy landscapes

In this section, we present the corresponding free energy landscapes for the unfolding motion of the DNA i-motif. The convergence of the free energy landscape has been achieved by combining the 50 ns simulation long constant metadynamics potential energy landscape with short simulation runs of 10 ns such that the final energy landscape can be calculated by the WHAM algorithm[18,19]. We follow with this approach the landscape convergence method presented in Ref.[9] to satisfy detailed-balance conditions. In detail, we conducted ten simulations each with 10 ns length with a constant biasing energy that start at different points of the landscape. By the overlap of the trajectories we are able to satisfy the restrictions of detailed balance which allows the calculation of the underlying free energy landscape by the histogram reweighting procedure.

For the calculation of the eigenvector[21] and the contact pair free energy landscape, we have used the projection scheme as proposed in Ref.[9]. The applied set of eigenvectors has been derived by a series of high temperature 500 K simulation runs for initially unprotonated DNA i-motifs as discussed in previous publications[3,4,9]. We have shown in a recent publication[4] that the unfolding motion along these eigenvectors resembles the low temperature unfolding pathway in good agreement. We are therefore confident that the biased unfolding pathway resembles the real unfolding pathway. The corresponding free energy landscape for the contact pair number is presented in 4. We have derived the free energy landscape by the projection scheme as proposed in Ref.[9]. The reaction coordinate is given in terms of the ratio to the initial number of all proton-DNA contact pairs. It can be clearly seen that the initial i-motif is a metastable configuration with an activation barrier of roughly 16 kcal/mol. The corresponding significantly destabilized structure after the release of two protons is a high level metastable state with a free energy difference of roughly $\Delta F^* \approx 2 - 2.5$ kcal/mol compared to the i-motif conformation. The further decrease of DNA atoms - proton contact pairs due to rearrangement and further unfolding into hairpin configurations is slightly hindered by a free energy barrier of $\Delta F^*_{23} \approx 6$ kcal/mol. The global free energy minimum at $N_P = 0.31$ is populated with hairpin structures, random coil configurations and fully unfolded conformations. A further division into its separate configurational contributions is not

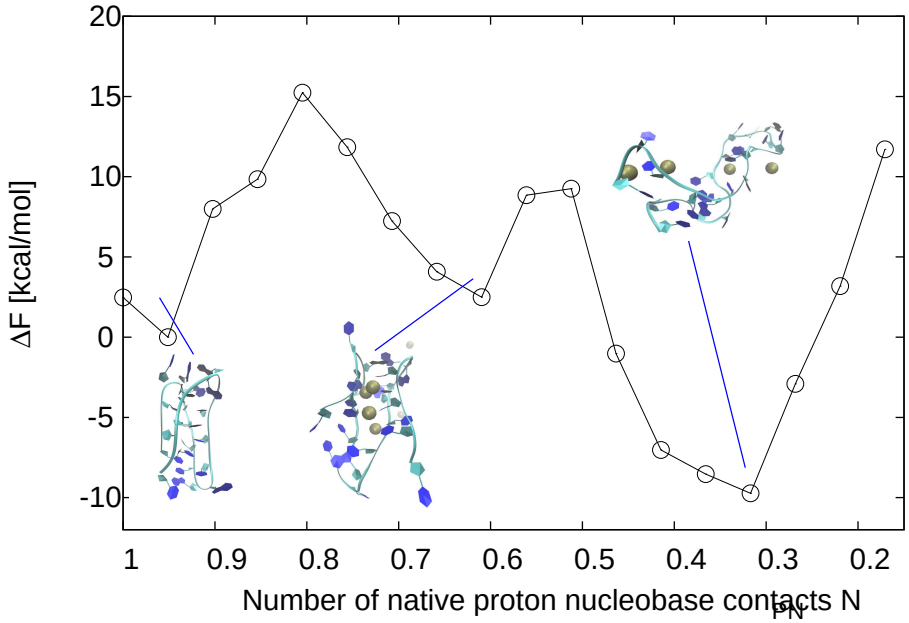

Figure 4: Free energy landscape for the number of native contact pairs between atoms of the DNA and protons. The structures visualize typical stable conformations for the reaction coordinate values. The initial structure at $N_P = 0.95$ is the rearranged DNA i-motif. The release of two protons is inhibited by a large free energy barrier at $N_P = 0.8$. The structure with 4 protons is shown at $N_H = 0.6$. The final unfolded configurations with hairpin structures and random coil configurations are located at the global free energy minimum at $N_P = 0.31$.

possible due to identical proton-DNA contact pair ratios for these conformations. The stabilization of hairpin configurations can be mainly explained by the entropic ordering of the solvent. It has been shown in Ref.[3] that DNA configurations whose nucleobases point towards the solvent are less energetically favorable than non-Watson-Crick like paired nucleobases as given for hairpins. The reason for this behavior is the loss of entropic contributions to the free energy due to additional electrostatic ordering of the water molecules in combination with excess hydrogen bonds[4] which counteracts and over-compensates favorable enthalpic contributions. The appearance of this effect has been in detail discussed in Refs.[3,24].

The separation of the several structures into different configurations can be achieved by a projection onto the eigenvector free energy landscape of the system as it has been discussed in Refs.[3,4,9]. For a better comparison to previous results[4], we have used the eigenvectors of unprotonated DNA i-motif high temperature simulations with an identical nucleobase sequence. The comparability to

an identical low temperature unfolding process has been discussed in Ref.[4]. The results are shown in 5. With regard to the discussions in Refs.[3,4,25], it has to be noted that eigenvector 1 presents

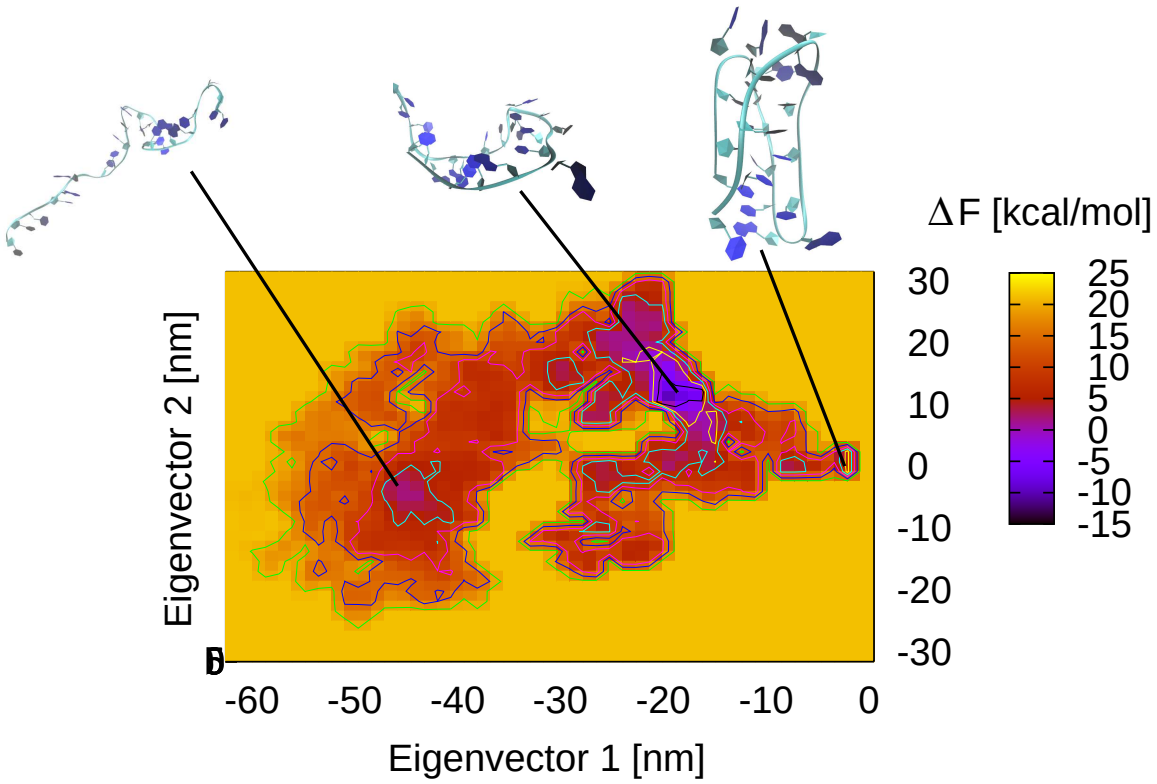

Figure 5: Free energy landscape of the first two eigenvectors for the protonated native state derived by an unprotonated high temperature unfolding simulation of an i-motif. The conformations illustrate typical conformations. The free energy landscape clearly reveals the stability of hairpin configurations.

a broadening of the end-to-end distance between nucleobases 1 and 22 while eigenvector 2 represents the relaxation behavior into planar structures[3,4]. It can be clearly seen, that the energetic minimum for the i-motif is located in a small area for eigenvector 1 and 2 values around (eigenvector 1 (EV1): 0, eigenvector 2 (EV2): 0). The presence of high free energy activation barriers that confine the accessible phase space for the native configuration can be observed in agreement to 4. The occupation of the small area for the initial structure illustrates the rigidity of the native

i-motif and therefore the strong electrostatic DNA-proton binding. The initial release of two protons occurs in the region (EV1: -10,0 $\pm$ 5) and (EV2: 0,0 $\pm$ 5). After the deprotonation process, the structure undergoes wiggling and torsional motion which is mainly located in the eigenvector region $(EV1 : -30, -10 \pm 10)$ nm and $(EV2 : -20, -10 \pm 10)$ nm. The decreased stability is in qualitative agreement to the appearance of the metastable deprotonated state at $N_P = 0.6$ as shown in 4. It has to be noted, that specifically the torsional motion is of main importance for a further rearrangement of the protons and the nucleobases[25]. The final free energy minimum is represented by hairpin configurations in agreement to 4. Random coil and fully unfolded states are energetically less stable and occur at different regions of the free energy landscape. It has to be noted that the free energy barriers as well as the free energy differences between different configurations coincide. We therefore propose that the stability of hairpin structures as it can be seen by the free energy minima can be mainly explained by the same solvent-ordering effects as it has been proposed in Ref.[3].

References

(1) Chen, C.; Li, M.; Xing, Y.; Li, Y.; Joedecke, C.-C.; Jin, J.; Yang, Z.; Liu, D. *Langmuir* **2012**, *28*, 17743-17748.

(2) Liu, D.; Balasubramian, S. *Angew. Chem.* **2003**, *42*, 5734-5736.

(3) Smiatek, J.; Chen, C.; Liu, D.; Heuer, A. *J. Phys. Chem. B* **2011**, *115*, 13788-13795.

(4) Smiatek, J.; Liu, D.; Heuer, A. *Curr. Phys. Chem* **2012**, *2*, 115-123.

(5) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562-12566.

(6) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601-126623.

(7) Ensing, B.; de Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. *Acc. Chem. Res.*, **2005**, *39*, 73-81.

(8) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. *J. Phys. Chem. B* **2006**, *109*, 6676-6687.

(9) Smiatek, J.; Heuer, A. *J. Comp. Chem.* **2011**, *32*, 2084-2096.

(10) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.

(11) Sorin, E. J. ffAMBER - AMBER force field ports for the GROMACS molecular dynamics suite. http://ffamber.cnsm.csulb.edu/ (accessed Oct 4th, 2011).

(12) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472-2493.

(13) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comp. Chem.* **2000**, *21*, 1049-1074.

(14) Phan, T. A.; Gueron, M.; Leroy, J.-L. *J. Mol. Bio.* **2000**, *299*, 123-144.

(15) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L; *J. Chem. Phys.* **1983**, *79*, 926-935

(16) Spackova, N.; Berger, I.; Egli, M.; Sponer, J. *J. Am. Chem. Soc.* **1998**, *120*, 6147-6151.

(17) Malliavin, T.E.; Gau, J.; Snoussi, K.; Leroy, J.L *Biophys. J.* **2003**, *84*, 3838-3847.

(18) Kumar, S.; Rosenbergm J. M.; Bouzida, D.; Swenden, R; Kollman, P. A.; *J. Comput. Chem.* **1992**, *16*, 1339.

(19) Roux, B.; *Comput. Phys. Comm.* **1995**, *91*, 275.

(20) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comp. Phys. Comm.* **2009**, *180*, 1961-1972.

(21) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412-425.

(22) Phan, A. T.; Gueron, M.; Leroy, J. L.; *J. Mol. Biol.* **2000**, *299*, 123-144

(23) Parkinson, G. N.; Lee, M. P. H. and Neidle, S.; *Nature* **2002**, *417*, 876-880.

(24) Finkelstein, A. V.; Ptitsyn, O. B.; *Protein Physics - A course of Lectures* **2002**, Academic Press, Amsterdam.

(25) Smiatek, J.; Janssen-Mueller, D.; Friederich, R.; Heuer, A.; *Physica A* **2014**, *394*, 136-144.