

MOLECULAR DYNAMICS OF RNA STRUCTURAL MOTIF. TEMPERATURE ENHANCED SAMPLING OF THE CONFORMATIONAL SPACE

TADEUSZ KULIŃSKI, KATARZYNA KULIŃSKA

*Institute of Bioorganic Chemistry, Polish Academy of Sciences,
Noskowskiego 12/14, 61-704 Poznań, Poland, e-mail: tadkul@ibch.poznan.pl*

Abstract. Temperature Enhanced Molecular Dynamics (TEMED) simulations were applied for the sampling of the conformational space of the RNA structural motif. During MD experiments run in explicit water and ions we observed at the atomic level the switching of the RNA tetraloop structures from the unusual conformations found in the crystal form to the conformation characteristic for the free molecule in the solution. TEMD simulations prove to be useful for the exploration of the possible conformational switches, kinetic traps in RNA folding, the detection of the barriers on folding energy surfaces as well as revealing the role of water molecules and counter ions in the stabilization of RNA structure.

1. INTRODUCTION

Molecular Dynamics (MD) simulations, based on model energy functions that describe molecular interactions, are progressively used to investigate nucleic acids structure and dynamics in addition to experimental (structural, biochemical and thermodynamic) approaches [1-3]. Allowing direct study of the forces and mechanisms that drive structure formation at atomic level, MD simulations complement structural studies in cases in which the timescales or physical properties of interest are not easily accessible to experimental methods.

RNA molecules play the key roles in all steps of gene expression, transport, catalysis and other cellular processes. Folding into compact and function-specific 3D conformations, they resemble proteins more than their double-stranded DNA cousins and are difficult objects for structural studies. The highly anionic double-stranded helical regions are tightly packed insight a globular architecture, whereas the nonhelical structural motifs are on the surface of RNA molecules. The biological function of the RNA is often directly related to the presence of the structural motifs, which form recognition sites for ligands or mediate a possible catalytic activity.

RNA structural motifs consist of ordered arrays of non-Watson-Crick base pairs forming distinctive foldings of the phosphodiester backbones of the interacting RNA strands. They mediate the specific interactions that induce the compact folding of complex RNAs [4, 5]. Some of these elements can function as isolated structures, without the context of a larger folded RNA structure. Particularly stable hairpins can form nucleation points for RNA folding. The study of the range of conformations accessible to these elements is important for understanding the mechanisms of structure formation and intermolecular interactions of RNA.

MD methods could improve structure prediction and model building of RNAs, provided that the exploration of the intermolecular interactions by force fields is sufficiently accurate and efficient. The major difficulty in determination of the RNA folding path by the simulation is related to the limited time and conformation range, which may be probed by MD in comparison to real RNA folding kinetics. Roughness of the folding energy surface, kinetic traps and stable misfolds are the factors which requires attention during *in vitro* RNA folding procedures [6], Adequately special consideration of infrequent events in structural transitions is also needed in MD experiments due to relatively low statistics reachable in current status of software and hardware available for these calculations.

In this paper we describe the application of the Temperature Enhanced Molecular Dynamics (TEMED) for the sampling of conformational space of the RNA structural motif. The method exploit MD simulations run in explicit water and ions in a range of different temperatures. In this way conformational equilibrium in temperatures close to the melting temperature of particular structural modules may be better explored indicating possible conformational switches, kinetic traps in RNA folding and barriers on folding energy surfaces as well as the role of water molecules and counter ions.

As an object for our studies we selected the structure of the hairpin of the loop D closing the extended domain IV of 5S rRNA from *Thermus Flavus* [7], This domain is of major interest because it is located near the peptidyltransferase center of the ribosome. The hairpin loop belongs to the highly conserved class of exceptionally stable GNRA tetraloops occurring in ribosomal RNAs (where N is any nucleotide, R is G or A). Surprisingly the crystal structure of the loop D shows conformational variability and is distinctly different in comparison to the other tetraloops determined by crystallography and NMR [8-10],

The aim of these MD experiments is the exploration of the switching of this tetraloop structure from the unusual conformations found in the crystal form. We expected that the molecule deprived of the intermolecular interactions imposed within the crystal network would accommodate the conformation characteristic for free molecule suspended in the solution. Previous studies have shown that the structural transformations between different conformations are difficult to be observed during standard MD simulations of limited length and required special procedures. The convergence from incorrect to the correct form of a UUCG tetraloop was observed only when a Locally Enhanced Sampling procedure was used [11] or if the loop riboses were changed to deoxyriboses [12], Several approaches have been proposed to improve conformational space sampling based on MD calculated at higher then normal temperatures [13, 14], In the method described in this paper we show the importance of explicit analysis of kinetic traps and misfolds from multiple MD trajectories parallelly run at different temperatures for the full conformational analysis and model building of RNA.

2. METHODS

Molecular dynamics (MD) simulations were performed using the CHARMM simulation program, version 27 [15], with the CHARMM27 all-atom force field for nucleic acids [16, 17].

Starting coordinates were taken from the crystal structure of a domain IV of *Thermus Flavus* 5S rRNA including two hairpins with a GCGA tetraloop (PDB code 361D, chain A and B) [7]. The 14-mer hairpins were solvated in a $42 \times 42 \times 42 \text{ \AA}^3$ water box containing TIP3P water [18]. Water molecules closer than 1.9 \AA from any RNA atom were removed. To neutralize the system, for both structures, 13 sodium counterions were placed on the bisector of the phosphate oxygen atoms, 6.0 \AA from the phosphorus atoms, which corresponds to a Na^+ concentration of 0.3 M .

The whole system was minimized, usually for 200 steps of SD with harmonic restraints on RNA (3 rounds of 50 steepest descent (SD) steps using harmonic restraints with 15, 10 and $5 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ force constant respectively on all atoms) and finally 200 steps of SD without any restraints.

MD simulations started with heating the system from 50 to the desired temperature (300 K, 330 K, 350 K) followed by short equilibration, after which trajectories were collected. MD simulations were run in the NVE ensemble, using periodic boundary conditions, a 2 fs time step and SHAKE algorithm [19] to constrain all bonds to hydrogens. The nonbonded interactions were smoothly shifted to zero at a 12.0 \AA cut-off by the atom based truncation method with a force shifting function [20]. The nonbonded pair list was generated using a 13 \AA cut off and updated when any atom had moved more than 0.5 \AA . For each tetraloop several independent simulations were performed. These repeated simulations differed in initial velocities.

Most of the simulations were run on a 16 Pentium III processor cluster *Hydra* located at the Institute of Bioorganic Chemistry PAS.

3. RESULTS AND DISCUSSION

Extensive MD simulations of two RNA structures: the hairpins with a GCGA tetraloop have been performed at 300 K and elevated 330 K and 350 K temperatures (*d300a*, *d300b*, *d330a*, *d330b*, *d350a*, *d350b*) using as starting coordinates two molecules (the sequence G82-C94 from chain A and B) included in the crystal unit (PDB code 361D). Figure 1 presents one of two initial structures of the simulated RNA hairpin with GCGA tetraloop including the sequence G82-C94 taken from the crystal structure of a domain IV of *Thermus Flavus* 5S rRNA (PDB code 361D, chain B).

Several trajectories showed the convergence of the initial RNA structures toward similar conformations with a pattern of hydrogen bonds and stacking interactions characteristic for GNRA loop motif (Fig. 2). The GNRA tetraloops occur commonly in ribosomal and others RNAs and show increased stability. They are often involved in tertiary contacts in larger RNA structures, serve as a recognition sites for protein binding and likely initiate folding events [20, 21]. The solution structures of GNRA tetraloops characterized by NMR show that they have well defined interactions which include hydrogen bonds between the G and A bases, between the G base and the phosphate oxygen of the A nucleotide, and stacking between the third and fourth loop bases [9, 20].

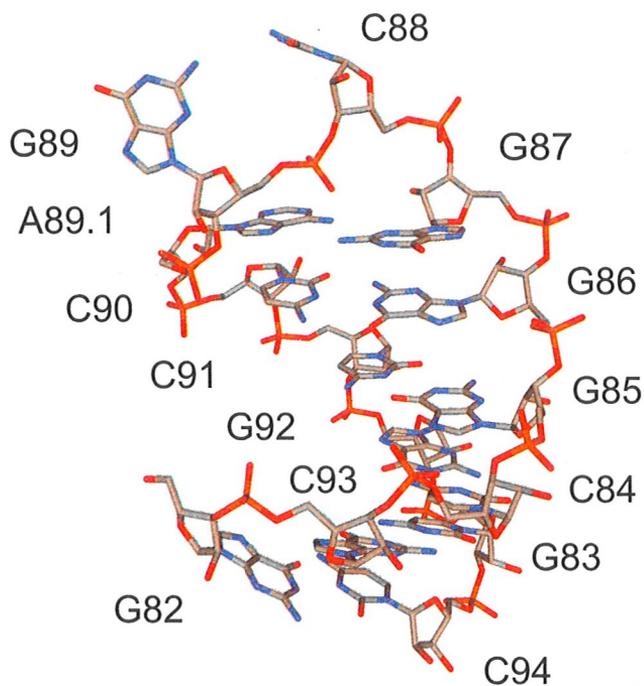


Fig. 1. Initial structure of the simulated RNA hairpin with GCGA tetraloop including the sequence G82-C94 taken from from the crystal structure of a domain IV of *Thermus Flavus* 5S rRNA (PDB code 36ID, chain B)

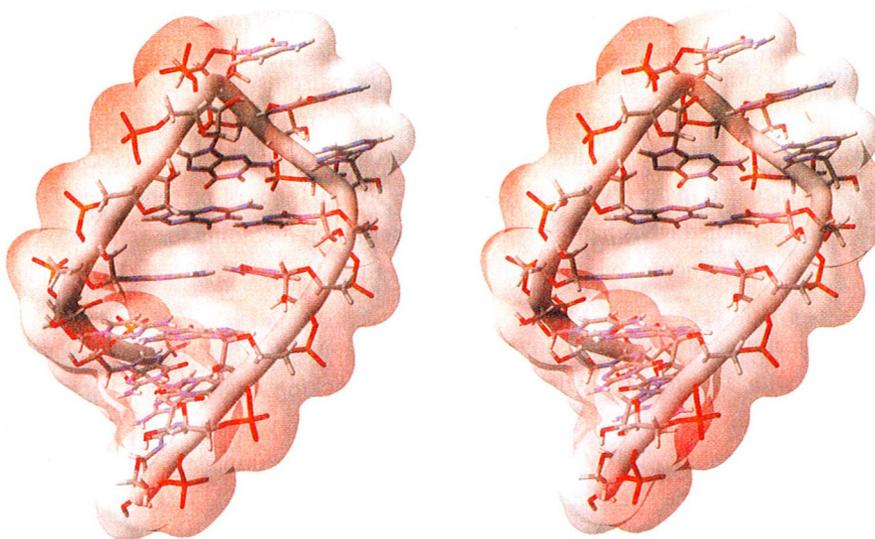


Fig. 2. A stereoview of the converged structure of the loop D (structure averaged over the 100 ps of *d350a* trajectory after 450 ps of the simulation)

Surprisingly the crystal structure of the loop D hairpin closing the helix IV of 5S rRNA from *Thermus Flavus* [7] shows conformational variability and is distinctly different in comparison to the other tetraloops determined by crystallography and NMR [8-10]. The helical part of the extended domain E of both structures were classified in general as native conformation of right handed A-RNA structure, although the stereochemistry showed several deviations from ideal values. The first two nucleotides for each strand were "flipped out" outside the double helix. Most of the conformational differences of both chains were stabilized by several direct hydrogen bonds between both RNA helices. G83-C93 base pair shows a distorted Watson-Crick interaction in molecule A in contrast to the normal base pair observed in molecule B. Intermolecular hydro-

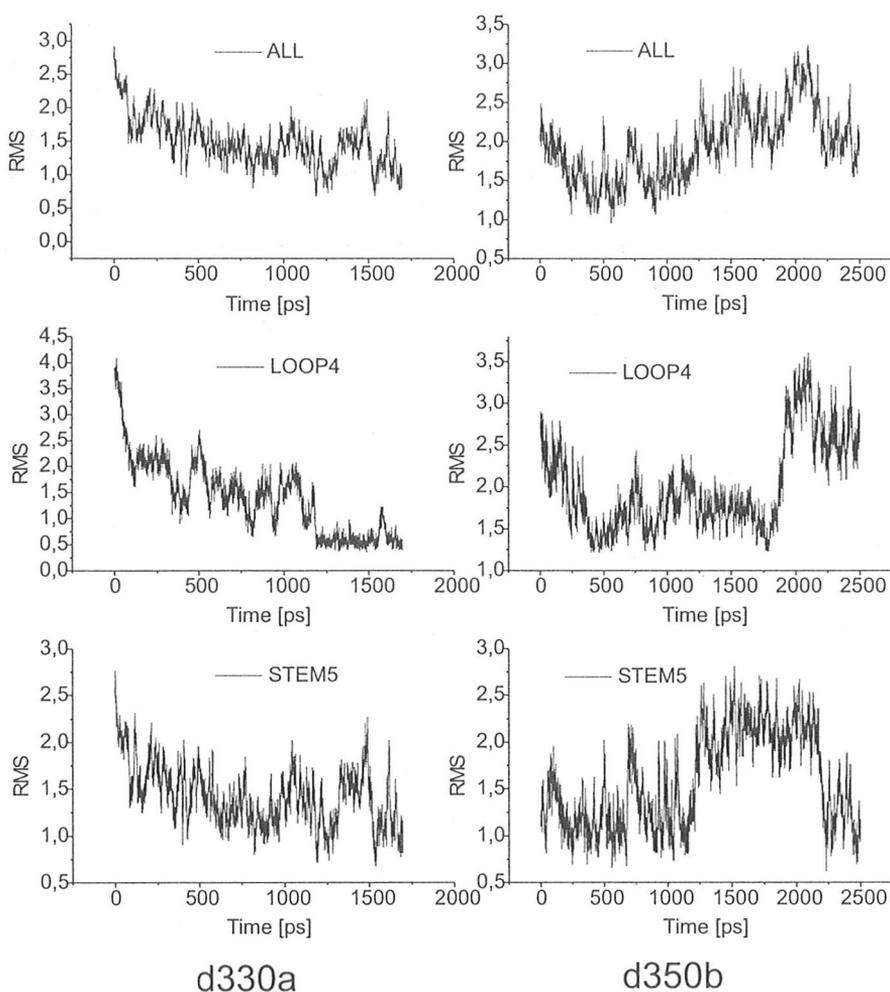


Fig. 3. Time course of the RMSD calculated in reference to the final converged structure for the simulations presented in this work, whole molecule (ALL), tetraloop (nucleotides 87-89.1; LOOP4) and helical stem (nucleotides 82-86 : 90-94; STEM5)

gen bonding lead to the loss of Watson-Crick pairing for bases C84 and G92 in molecule A. For both molecules in the crystal form G87 and A89.1 showed a distorted asymmetric hetero-purine formation G-A amino-N7, N3-amino. The terminal cytosine C88 and guanine G89 were sticking out to the solvent and were not stabilized by direct hydrogen bonds to other nucleotides.

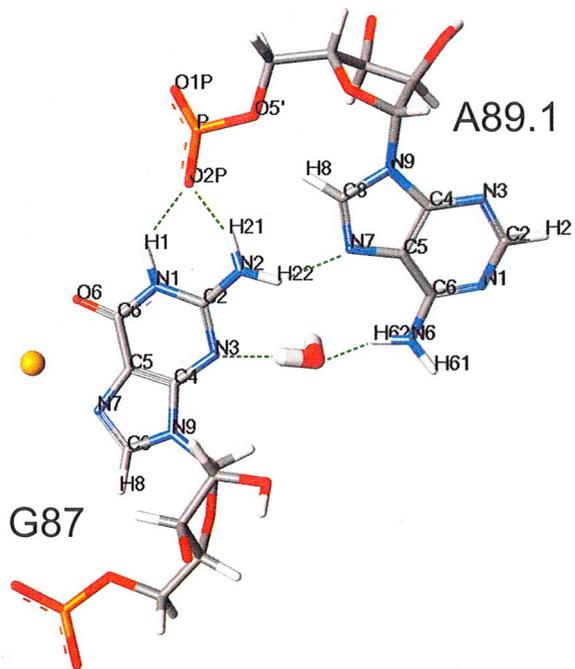
During the MD simulations of these hairpins at elevated temperature (*d330a*, *d330b*, *d350a*, *d350b*) we observed that in all trajectories RNA molecules relaxed from the crystal conformations, and in most of them the GCGA tetraloop converged toward similar conformations with the characteristic pattern of hydrogen bonds between G87 and A89.1.

Figure 3 presents the time course of the RMSD calculated in comparison to the converged structure similar to the consensus GNRA tetraloop structure. RMSD was calculated for all heavy atoms for the whole molecule (ALL), for the tetraloop motif including four apical nucleotides G87-A89.1 (LOOP4) and for the helical stem of the hairpin (STEM5). As seen from RMSD profile for trajectories *d330*, the convergence of the structure was achieved within 1000 ps of the simulation. The major rearrangement was localized within the tetraloop, which achieved final conformation in few distinct structural rearrangements, reducing the RMSD value from 4.0 Å at the beginning of the simulation to the final structure, which was remaining practically unchanged for more than 500 ps until the end of the trajectory, deviating only within 0.5 Å. In the simulation *d350b* the overall structure of the hairpin reached the consensus conformation (RMSD value around 1.2 Å) within less than 400 ps of the simulation. The tetraloop structure was close to GNRA consensus structure within 1.3 Å, RMSD of the stem was within 1.0 Å. However, after 1.2 ns of continued MD there was an unfolding event within the helical stem, followed by the unfolding of the tetraloop. which started after 1750 ps of the MD, and after more than 1 ns of the equilibration of the compact folded tetraloop structure. This simulation confirms that the conformational transformations observed by MD consist often of highly probabilistic infrequent events. Such events are difficult to be simulated in a way to correctly reflect thermodynamic conformational equilibrium of the complex systems.

In all trajectories obtained in our MD simulations the structural motif involve the G87 base making three direct hydrogen bonds, (G87)N2-H...N7(A89.1), (G87)N2-H...O1P(A89.1) and (G87)N1-H...O1P(A89.1), as well as water molecule mediating hydrogen bonding between (G87)N3 and (A89.1)N6-H - figure 4. This pattern of hydrogen bonds required major rearrangements of the initial X-ray derived structure. To accommodate a bridging water molecule the (G87)N3...N6(A89.1) distance increased to the average value of 5.7 Å. The distance and angle of the (G87)N1-H...O1P(A89.1) hydrogen bond are close to the values characteristic for this type of interaction. Water molecules involved in water-mediated interactions within the tetraloop displayed relatively long residence times. The water molecule present within GCGA tetraloop was preferentially hydrogen bonded to (A89.1)N6-H. (G87)N3 and (G87)O2' Since the water bridge between G87 and A89.1 was established, the (G87-A89.1) distances remained stable. We have observed rare exchanges of bridging water molecules while the (G87)N3...N6(A89.1) distance was kept at the value allowing to accommodate the water molecule.

In the converged structure C88 base was not stabilized by any direct hydrogen bond. During the structural rearrangements, however, several stable structures characterized by the 2' endo conformation assisted by hydrogen bonding of (C88)2'OH to N7 of neighboring G89 were observed. The (G87)2'OH...N7(G89) hydrogen bond was observed in the conformations with a water-mediated G87-A89.1 contact, and it was also present in fragments of trajectories *d330* and *d350* before the water bridge was established. It was frequently replaced by water mediated

Fig. 4. Hydrogen bond network within the tetraloop GCGA between G87 and A89.1. Bridging water molecule in the loop and sodium ion are also shown



hydrogen bonding between (G87)2'OH and O6(G89). The 2'OH groups of C88 and A89.1 were preferentially hydrogen bonded to the phosphate oxygen of the next nucleotide from the 3' side.

The dynamics observed for this hairpin was different from this found in GCAA tetraloop simulation with the OPLS force field performed by Zichi [22]. However, the hydrogen bond network between nucleotides G87 and A89.1 was the same despite the differences in the sequence and force field used for MD. The same structural motif was found previously by us in other RNA hairpin containing GUGA tetraloop [23].

The G87 base in GCGA tetraloop was involved in stacking interactions mostly with G86 from the G86-C90 base pair closing the tetraloop. Bases C88, G89 and A89.1 stacked sequentially with similar stacking interaction energies. The stacking between the A89.1 base from the loop and the C90 base from the stem was less favorable.

C88 remained in its final stacking conformation in our *d330a* MD simulation, whereas in two simulations at 350 K it shows much more flexibility. The second nucleotide in the GCGA loop is looped out into the solvent in two out of ten NMR models.

4. CONCLUSIONS

During Temperature Enhanced Molecular Dynamics (TEMED) simulations the conformational convergence of the two RNA hairpins with GCGA tetraloop was achieved. These molecules, relaxed from the intermolecular interactions imposed within the crystal network, have accommodated during MD simulations run in explicit water and ions surrounding the conformations with the pattern of hydrogen bonds and stacking interactions characteristic for GNRA loop motif as detected by NMR in solution.

Presented results prove the usefulness of TEMED simulations for the sampling of conformational space to trace possible folding pathways of RNAs. MD simulations run with explicit water and ions in a range of different temperatures, well explore conformational equilibrium in temperatures close to the melting temperature of particular structural modules. Possible conformational switches, kinetic traps in RNA folding and barriers on folding energy surfaces as well as the role of water molecules and counter ions may be efficiently identified and confronted with experimental data.

Acknowledgments

This work was supported by grant from the State Committee for Scientific Research, Republic of Poland (8 T11F002 19). Access to the Poznań Supercomputing and Networking Centre is acknowledged.

References

- [1] T. E. Cheatham, P. A. Kollman III, *Ann. Rev. Phys. Chem.*, **51**, 435-471 (2000).
- [2] J. Norberg, L. Nilsson. *Acc. Chem. Res.*, **35**, 465-472 (2002).
- [3] M. Karplus, *Acc. Chem. Res.*, **35**, 321-323 (2002).
- [4] N. B. Leontis, E. Westhof, *RNA*, **4**, 1134-1153 (1998).
- [5] N. B. Leontis, E. Westhof, *J. Mol. Biol.*, **283**, 571-583 (1998).
- [6] D. K. Treiber, J. R. Williamson, *Curr. Opin. Struct. Biol.*, **11**, 309-314 (2001).
- [7] M. Perbandt, A. Nottle, S. Lorenz, R. Bald, C. Bezel, V. A. Erdman, *FEBS Lett.*, **429**, 211-215 (1998).
- [8] H. A. Heus, A. Pardi, *Science*, **253**, 191-194 (1991).
- [9] F. M. Jucker, H. A. Heus, P. F. Yip, E. H. M. Moors, A. Pardi, *J. Mol. Biol.*, **264**, 968-980 (1996).
- [10] H. W. Pley, K. M. Flaherty, D. B. McKay, *Nature*, **372**, 68-74 (1994).
- [11] C. Simmerling, J. Miller, P. Kollman, *J. Am. Chem. Soc.*, **120**, 7149-7155 (1998).
- [12] J. L. Miller, P. A. Kollman, *J. Mol. Biol.*, **270**, 436-450 (1997).
- [13] L. Bielecki, M. Popenda, R. W. Adamiak, *Nuci. Acids Res., Suppl* 57-58 (2002).
- [14] M. R. Sorensen, A. F. Voter. *J. Chem. Phys.*, **112**, 9599-9606 (2000).
- [15] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comp. Chem.*, **4**, 187-217 (1983).

- [16] N. Foloppe, A. D. MacKerell, *J. Comput. Chem.*, **21**, 86-104 (2000).
- [17] D. A. MacKerell Jr., B. Brooks, C. L. Brooks II, L. Nilsson, B. Roux, M. Karplus in: *Encyclopedia of Computational Chemistry*, eds. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, P. R. E. Schreiner, John Wiley & Sons, Chichester, UK, 271-277 (1998).
- [18] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.*, **1479**, 926-935 (1983).
- [19] J. -P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comp. Phys.*, **23**, 327-341 (1977).
- [20] P. J. Steinbach, B. R. Brooks, *J. Comp. Chem.*, **15**, 667-683 (1994).
- [21] C. C. Correll, I. G. Wool, A. Munishkin, *J. Mol. Biol.*, **292**, 275-287 (1999).
- [22] D. A. Zichi, *J. Amer. Chem. Soc.*, **117**, 2957-2969 (1995).
- [23] J. Sarzyńska, T. Kuliński, L. Nilsson, *Biophys. J.*, **79**, 1213-1227 (2000).