

PARAMETRIZATION OF 2-AMINOPURINE AND PURINE IN CHARMM ALL-ATOM EMPIRICAL FORCE FIELD*

J. SARZYŃSKA, K. KULIŃSKA, T. KULIŃSKI

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

(Rec. 18 November 2003)

Abstract: Molecular dynamics simulations based on empirical force fields are commonly used in studies of large and complex biological systems. A new set of force field parameters complementing the CHARMM27 all atom empirical force field for nucleic acid was developed for 2-aminopurine and purine, two nucleobase analogues extensively used to probe the nucleic acids structure and dynamics.

1. INTRODUCTION

During the last decade the main goal of molecular biology became to understand the functions of biological molecules in terms of structure, interactions and processes at the atomic level. A new alternative to experimental techniques, providing an increasing amount of information with atomic resolution, is computer modeling and simulation. Despite the extremely fast progress of computing power, such calculations have to be based on a compromise between the complexity of the description of the molecular system, the number of atoms included, and the computational time required for a reliable description of the investigated process.

The most efficient and commonly used in studies of large and complex biological systems are the techniques based on the empirical description of the potential energy surface [8, 5] in the function of molecular coordinates, namely molecular mechanics (MM) and molecular dynamics (MD). The empirical fit to the potential energy surface, called the force field, may have a different functional form, and describes interactions which define the structure of the molecule as well as external interatomic interactions including van der Waals and electrostatic interactions, depending on the type of the atoms involved [1]. The force field describes entire classes of the molecules with compromised accuracy, as an extrapolation from the empirical data of a representative set of molecules.

The majority of biomolecular simulations were performed with CHARMM [2, 3] or AMBER [4] program. The force fields developed within these programs are referred to as CHARMM and AMBER force field, respectively. The force field parameters are based on the reproduction of experimental results for nucleic acid oligomers, e.g., condensed phase structural properties of DNA and RNA, and the consistency with small molecule results obtained from quantum mechanical (QM) calculations and experimental data.

* Dedicated to memory of Professor Jacek Rychlewski

The potential energy for the CHARMM force field has a form:

$$U(R) = \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\chi (1 + \cos(n\chi - \delta)) + \\ + \sum_{\text{impropers}} K_\varphi (\varphi - \varphi_0)^2 + \sum_{\text{nonbonded}} \left(\epsilon_{ij} \left[\left(\frac{R_{\min ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_D r_{ij}} \right)$$

The optimized parameters are equilibrium values of bond length, b_0 , Urey-Bradley 1,3 distance, S_0 , valence angle, θ_0 , improper torsion angle, φ_0 , and force constants of bond, K_b , Urey-Bradley, K_{UB} , valence angle, K_θ , dihedral angles, K_χ , and improper torsion angle, K_φ . The parameters n and δ in the dihedral term are the multiplicity and phase, respectively. These terms are referred to as internal parameters. Nonbonded parameters including partial atomic charges, q_i , the Lennard-Jones well depth, ϵ_{ij} , and minimum interaction radius, $R_{\min ij}$ are also optimized. The most recent CHARMM force field for nucleic acids is CHARMM27 [5, 6]. The AMBER force field for nucleic acid was developed by Cornell [7] and subtly modified by Cheatham [8]. Both force fields lead to the reliable description of the structure, energetics and dynamics of nucleic acids [9-11]. The AMBER force field does not have Urey-Bradley term describing 1-3 interactions between the atoms bonded to the common atom. Furthermore, the improper torsion is described as the standard torsion term. The major difference between force fields is in the way the parameters of the energy function are derived, although both force fields' development relies on *ab initio* and experimental target data. In the CHARMM, partial atomic charges are empirically adjusted to reproduce a set of *ab initio* calculated interaction energies between selected model compound and water molecule, whereas in the AMBER force field, partial atomic charges are derived using the restrained electrostatic potential (RESP) approach [12].

The aim of this work is to extend the CHARMM27 force field to be able to treat two nonstandard nucleic acid bases, 2-aminopurine (2AP) and purine (PUR).

2AP, the fluorescent analogue of guanine and adenine, has been extensively used to probe the nucleic acids structure and dynamics [13-15]. As an adenine isomer, 2AP forms base pair with thymine, however, it can also pair with cytosine [16, 17]. This implies that 2AP can produce mutagenic AT \leftrightarrow GC transitions.

The substitution of the natural base by 2AP and PUR changes the hydrogen donor and acceptor pattern. Substitutions of guanine or adenosine by 2AP or PUR eliminate particular interactions in the RNA tertiary structure and were used to evaluate the specific hydrogen bond energetic contribution to the structure stability [18-20].

2. METHODS

Parameterization of 2AP and PUR was performed according to the previously established protocol for the CHARMM force field [5, 21, 22]. Parameterization starts from the topology

and the initial parameters assignment for the selected model compound. For a large molecule this is usually its fragment. Parameter optimization is a multistep process involving iterative recalculations with tuned external and internal parameters. Results obtained in the CHARMM force field are compared to the target data from experiment or QM calculations and the parameters are manually adjusted to obtain the best agreement. The target data for charge optimization are the base-water minimum interaction energies and distances calculated *ab initio*, whereas the target data for bond and angle equilibrium values are experimental data or the geometry from *ab initio* optimization.

The gas phase *ab initio* calculations were carried out with the Gaussian 98 program [23]. The QM minimum interaction energies and distances between different sites of the studied bases and individual water molecules were determined at the HF/6-31G* level by optimizing the interaction distances. The intramolecular geometries were constrained to the gas phase HF/6-31G* optimized structure for the base and the TIP3P geometry for the water. The interaction energies were calculated as the total energy of the base-water supramolecular complex minus the sum of the individual monomer energies.

Molecular mechanics calculations were performed with the CHARMM program, version 27, with no truncation of nonbonded interactions and dielectric constant ϵ_d equal 1. Minimum interaction energies and geometries between bases and water molecules were determined in CHARMM by varying the interaction distances, with the intramolecular geometries constrained to the CHARMM gas phase minimized structure for the bases or the TIP3P geometry for water. The energy minimization involved 50 steps of steepest descent (SD) followed by 20 steps of Nepton-Raphson (NR) minimizer. The orientations of the individual water molecules were identical to those used in the QM calculations.

3. RESULTS AND DISCUSSION

To obtain parameters for 2AP and PUR which are compatible to the CHARMM27 force field, the same procedure as for the development of the original force field was applied. This also included the choice of HF/G-31G* level of theory for QM calculations of base-water interactions. The emphasis has been placed on the optimization of the partial atomic charges and geometries. The proper representation of the electrostatic interactions is crucial for the force field for biological macromolecules.

The structures and atom numbering of 2AP and PUR are shown in Fig. 1.

The initial topology information (connectivity, atom types) and, as a consequence, most internal parameters for 2AP and PUR was adopted from guanosine (GUA) and adenosine (ADE), respectively. For C6-H6 group, absent in GUA or ADE, the same atom type as for C2-H2 in ADE and C8-H8 in purine bases was used. The missing parameters, unique to the new bases, were assigned on the basis of the analogy to existing parameters. The preliminary partial atomic charges were also derived from GUA or ADE.

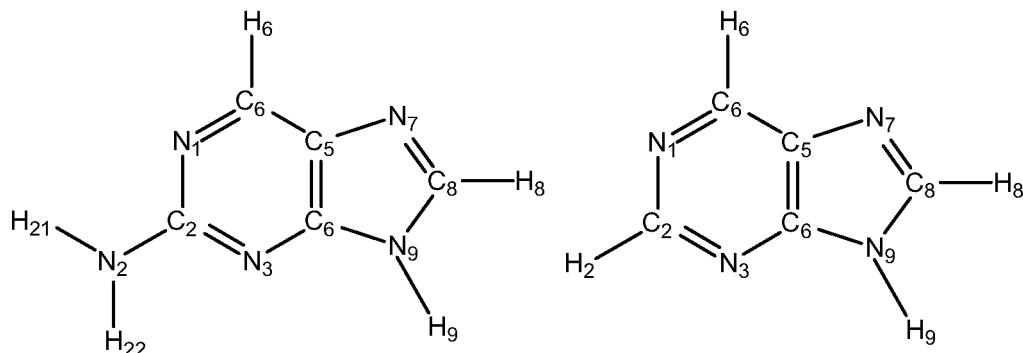


Fig. 1. Molecular structure and atom numbering of 2AP and PUR

The initial parameters were subsequently optimized *via* an iterative approach. The optimization processes started from the determination of partial atomic charges. The partial atomic charges were optimized to reproduce minimum energies and distances for individual water molecules interacting with different sites on 2AP and PUR. Different interaction orientations between bases and water molecules are shown in Fig. 2. *Ab initio* calculations predict nonplanar geometry of the amino group of isolated nucleic acids bases. The nonplanarity of 2AP is larger than that of ADE base but smaller than nonplanarity of GUA base (Table 1). In the CHARMM27 force field bases are assumed to be planar but force constants of the amino group allow deviation from planarity [5]. To be in agreement with optimization protocol for natural bases, we kept the amino group planar when geometry of 2AP was optimized for use to build complex with water. Partial atomic charges were manually adjusted to get the best agreement between the energies and the distances obtained from CHARMM and QM calculations for all base-water interactions. QM energies were scaled by 1.16 factor for all interactions prior to the comparison with CHARMM energies. [21]. The dipole moments of 2AP and PUR were also used as reference data for the partial atomic charges' optimization. The dipole moments of 2AP and PUR calculated with the CHARMM27 force field are systematically larger than the gas phase QM values, as required due to the lack of induced polarizability in the CHARMM27 force field, though well-defined scale factor is not applied (Table 2).

Table 1. Nonplanarity of the amino group of 2AP from *ab initio* calculations

Base	Method	Dihedral angle [deg]		Σ [deg]
		H22-N2-C2-N3	H21-N2-C2-N1	
2AP	HF/6-31G*	-18.2	17.6	350.4
	HF/6-31G**	-15.7	15.2	252.9
ADE[25]	HF/6-31G**	-0.5	4.6	359.6
GUA[25]	HF/6-31G**	-11.1	28.6	348.6

Σ - Sum of the HNC and HNH amino group valence bond angles

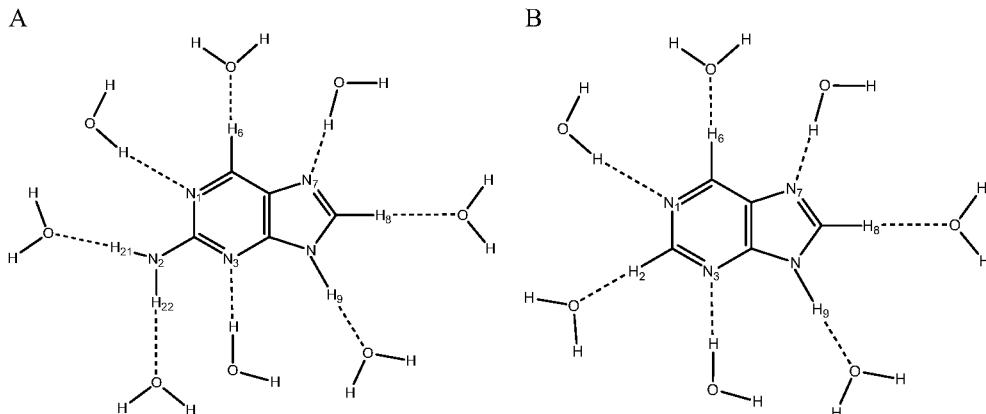


Fig. 2. Interaction orientations between bases and water used in optimization of the partial atomic charges. (A) 2AP, (B) PUR

Table 2. Dipole moments of the 2AP and PUR from CHARMM and *ab initio* gas phase calculations [Debye]

Base	CHARMM27	HF/6-31G*
2AP	3.35	3.08
PUR	4.49	3.67

The next step in the optimization process involved bond and angle terms. The geometry of the bases is primarily dictated by the bond length and angle equilibrium terms. The 2AP and PUR were energy minimized within the CHARMM27 force field using initially assigned bond and angle equilibration values and partial atomic charges obtained in the previous step. The bond and angle equilibration values for newly introduced parameters were manually adjusted to better reproduce target data. The target for geometry optimization of 2AP was the geometry extracted from the crystal structure of 9-[4-acetoxy-3-(acetoxymethyl) butyl]-2-aminopurine [24]. This geometry represents a 9-substituted 2-aminopurine. The only experimental geometry of purine was determined for the 7-H tautomeric form [23]. Because of the lack of appropriate crystallographic data, the target for optimization of the bond and angle equilibration values for PUR was the structure optimized on the MP2/6-31G* level of theory [24]. It should be noted that structures minimized in CHARMM do not have bonds lengths and angles that directly correspond to the equilibrium bond and angles parameters.

The initial attempts to optimize the 2AP geometry were made using the atom type NN3A for N1 nitrogen, the same as for ADE. The poor agreement with the target data obtained for N1-C2-N3 angle prompted us to introduce a new atom type, NN3P, for N3 and N1 atoms. The NN3P atom type was also used for N1 atom in PUR. This new atom type was necessary due to CHARMM hierarchical approach according to which no parameters already present in the force field should be changed. Following the initial optimization of the bond and angle

parameters, partial atomic charges were reoptimized, and then the geometry was rechecked. Steps within the loop were repeated until a satisfactory convergence was achieved.

Table 3A. *Ab initio* and CHARMM minimum interaction energies and distances between 2AP and water molecule

Interaction	<i>Ab initio</i>		CHARMM27		Difference	
	E_{\min}	R_{\min}^*	E_{\min}	R_{\min}	ΔE	ΔR
H8-OHH	-3.22	2.39	-3.27	2.36	-0.05	-0.03
H9-OHH	-7.21	1.82	-7.21	1.85	0.00	0.03
H21-OHH	-4.32	1.89	-4.27	1.91	0.05	0.02
H22-OHH	-4.10	1.91	-4.11	1.92	-0.01	0.01
H6-OHH	-1.74	2.44	-1.78	2.38	-0.04	-0.06
N7-HOH	-5.93	1.96	-5.96	1.92	-0.03	-0.04
N1-HOH	-6.69	1.90	-6.68	1.89	0.01	-0.01
N3-HOH	-5.02	1.94	-4.98	1.92	0.04	-0.02

Energies are in kcal/mol and distances are in Å. R_{\min}^* - minimum distances decreased by 0.2 for polar-neutral interactions, *Ab initio* energies are scaled by 1.16. See Fig. 2 for interaction orientations

Table 3B. *Ab initio* and CHARMM minimum interaction energies and distances between PUR and water molecule

Interaction	<i>Ab initio</i>		CHARMM27		Difference	
	E_{\min}	R_{\min}^*	E_{\min}	R_{\min}	ΔE	ΔR
H8-OHH	-3.82	2.35	-3.78	2.35	0.04	0.00
H9-OHH	-8.06	1.80	-8.04	1.84	0.02	0.04
H2-OHH	-1.57	2.51	-1.56	2.40	0.01	-0.11
H6-OHH	-2.06	2.41	-2.06	2.37	0.00	-0.06
N7-HOH	-5.40	1.97	-5.41	1.93	-0.01	-0.04
N1-HOH	-5.60	1.96	-5.60	1.93	0.00	-0.01
N3-HOH	-3.74	1.91	-2.77	1.95	-0.04	0.04

Energies are in kcal/mol and distances are in Å. R_{\min}^* - minimum distances decreased by 0.2 for polar-neutral interactions, *Ab initio* energies are scaled by 1.16. See Fig. 2 for interaction orientations

Table 4. Average differences, RMS differences and average absolute error between the base to water *ab initio* and CHARMM interaction energies [kcal/mol]

Base	Average Difference	RMS Difference	Average Absolute Error
2AP	-0.003	0.03	0.02
PUR	0.004	0.02	0.02

Average absolute error is the sum of the absolute values of the differences divided by n, the number of interactions of water with each base

Table 5. Statistical analysis of bond and angle parameters differences with respect to target data

Base	Bonds [Å]		Angles [deg]	
	Avg. diff.	SD	Avg. diff.	SD
2AP	0.01	0.004	0.3	0.1
PUR	0.01	0.004	0.7	0.1

SD - standard deviation. Analysis only for bonds and angles involving nonhydrogen atoms

The minimum interaction energies and distances from both the *ab initio* HF/6-31G* and the CHARMM27 calculations are presented in Table 3A and Table 3B. Small values of the average differences between base-water QM and CHARMM27 energies (-0.003 Å and 0.002 Å) ensure that the overall solvation of the bases will be reasonable, and small rms differences (0.03 Å and 0.02 Å) ensure that no individual term is too far from target data (Table 4).

Table 6. CHARMM parameters for 2AP and PUR. Force constants K_b , K_θ , K_χ and K_ϕ are in kcal/mol·Å², equilibrium bond lengths b_0 are in Å, equilibrium bond angles θ_0 , S and φ_0 are in degrees

6A) Bonds parameters		
Bond type	K_b	b_0
2AP		
CN2 NN3P	450.0	1.356
CN4 NN3P	400.0	1.290
CN4 CN5G	360.0	1.378
CN5 NN3P	350.0	1.304
PUR		
CN4 CN5	360.0	1.374

6B) Angle parameters		
Angle type	K_θ	θ_0
2AP		
CN2 NN3P CN4	90.0	120.6
NN3P CN4 CN5G	70.0	118.3
NN3P CN4 HN3	38.0	117.3
CN5G CN4 HN3	38.0	124.4
CN4 CN5G CN5	60.0	117.2
CN4 CN5G NN4	100.0	131.4
NN3P CN5 CN5G	60.0	126.2
CN2 NN3P CN5	90.0	113.5
NN3P CN2 NN3P	70.0	124.2
NN3P CN2 NN1	90.0	117.9
NN3P CN5 NN2B	100.0	129.2
PUR		
CN4 NN3P CN4	90.0	115.5
CN4 NN3P CN4	90.0	118.9
HN3 CN4 CN5	38.0	124.3
CN4 CN5 CN5	90.0	118.9
CN4 CN5 NN4	125.0	117.2
NN3A CN4 NN3P	70.0	129.2

6C) Dihedral parameters			
Angle type	K_χ	n	δ
2AP			
HN3 CN4 CN5G CN5	8.5	2	180.0
HN3 CN4 CN5G NN4	8.5	2	180.0
HN3 CN4 NN3P CN2	8.5	2	180.0
NN3P CN2 NN1 HN1	1.2	2	180.0
CN5 CN5G CN4 NN3P	1.8	2	180.0
CN4 NN3P CN2 NN3P	1.8	2	180.0
NN3P CN2 NN3P CN5	2.0	2	180.0
NN3P CN2 NN3P CN5	2.0	2	180.0
CN2 NN3P CN5 CN5G	0.2	2	180.0
CN5G CN4 NN3P CN2	0.2	2	180.0
CN4 NN3P CN2 NN1	4.0	2	180.0
CN5 NN3P CN2 NN1	4.0	2	180.0
NN3P CN4 CN5G NN4	2.0	2	180.0
CN2 NN3P CN5 NN2B	2.0	2	180.0
NN3P CN5 CN5G NN4	10.0	2	180.0
NN3P CN5 NN2B CN4	2.0	2	180.0
PUR			
HN3 CN4 NN3P CN4	8.5	2	180.0
HN3 CN4 CN5 CN5	8.5	2	180.0
HN3 CN4 CN5 NN4	8.5	2	180.0
NN3P CN4 CN5 NN4	2.0	2	180.0
NN3P CN4 CN5 CN5	0.2	2	180.0
NN3P CN4 NN3A CN5	2.0	2	180.0
NN3A CN4 NN3P CN4	1.8	2	180.0
CN5 CN4 NN3P CN4	0.2	2	180.0
6D) Improper dihedral parameters			
Angle type	K_ϕ	φ_0	
2AP			
CN4 NN3P CN5G HN3	39.0	0.0	
PUR			
CN4 NN3P CN5 HN3	39.0	0.0	

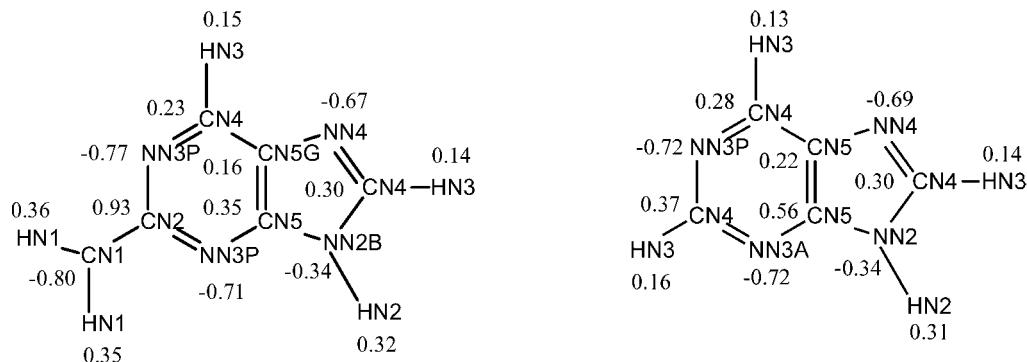


Fig. 3. CHARMM partial atomic charge and atom type parameters for 2AP and PUR

An agreement between CHARMM27 results and target data for the bond and angle equilibrium values was satisfactory when the 5-membered ring of 2AP was the same as of GUA, and of PUR the same as of ADE. The final CHARMM27 bonds and angle are in good agreement with the target data (Table 5). The bond and angle force constants, all dihedral and improper parameters, were adopted from the existing parameters for natural nucleic acids bases. The final charge parameters are shown in Fig. 3. The final CHARMM27 parameters for 2AP and PUR are presented in Table 6.

Acknowledgments

We are indebted to Nicolas Foloppe, Lennart Nilsson and Alexander D. MacKerell Jr for their helpful discussion. This work was supported by a grant from the State Committee for Scientific Research, Republic of Poland (8 T11F002 19). Access to the Poznań Supercomputing and Networking Centre has been acknowledged.

References

- [1] A. R. Leach in: *Molecular Modelling: Principles and Applications* Prentice Hall, pp. 165-245 (2001).
- [2] B. R. Brooks, R. E. Brucolieri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comp. Chem.* **4**, 187-217 (1983).
- [3] D. A. MacKerell, Jr., B. Brooks, C. L. Brooks II, L. Nilsson, B. Roux, and M. Karplus, in: *Encyclopedia of Computational Chemistry* (P. V. R. Schleyer, N. L. Allinger, T. Clark, J. Gaesteiger, P. A. Kollman, H. F. Schaefer III, and P. R. E. Schreiner, eds.), John Wiley & Sons, Chichester, UK, pp. 271-277 (1998).
- [4] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. AMBER 7, 2002. University of California, San Francisco.
- [5] N. Foloppe and A. D. MacKerell, *Journal of Computational Chemistry* **21**, 86-104 (2000).
- [6] A. D. MacKerell, N. Banavali, and N. Foloppe, *Biopolymers* **56**, 257-265 (2000).
- [7] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *Journal of the American Chemical Society* **117**, 5179 (1995).
- [8] T. E. Cheatham, P. Cieplak, and P. A. Kollman, *Journal of Biomolecular Structure & Dynamics* **16**, 845-862 (1999).

- [9] T. E. Cheatham and P. A. Kollman, Annual Review of Physical Chemistry **51**, 435-471 (2000).
- [10] J. Norberg and L. Nilsson, Accounts of Chemical Research **35**, 465-472 (2002).
- [11] W. Wang, O. Donini, C. M. Reyes, and P. A. Kollman, Annual Review of Biophysics and Biomolecular Structure **30**, 211-243 (2001).
- [12] J. M. Wang, P. Cieplak and P. A. Kollman, Journal of Computational Chemistry **21**, 1049-1074 (2000).
- [13] T. Kulinski, L. Bielecki, I. Zagorowska, and R. W. Adamiak, Collection of Czechoslovak Chemical Communications **61**, S265-S267 (1996).
- [14] T. Kulinski, L. Bielecki, M. Olejniczak, I. Zagorowska, and R.W. Adamiak, Collection of Czechoslovak Chemical Communications **2**,S191-S196 (1999).
- [15] T. Kulinski, L. Bielecki and R. W. Adamiak, Nucleic Acids Research. Supplement. **(1)**, 139-140 (2001).
- [16] L. C. Sowers, Y. Bouillard, and G. V. Fazakerley, Biochemistry **39**, 7613-7620 (2000).
- [17] E. C. Sherer and C. J. Cramer, Journal of Computational Chemistry **22**, 1167-1179 (2001).
- [18] J. SantaLucia, Jr., R. Kierzek, and D. H. Turner, Science **256**, 217-219 (1992).
- [19] D. Klostermeier and D. P. Millar, Biochemistry **41**, 14095-14102 (2002).
- [20] J.Sarzynska, L. Nilsson and T. Kulinski, Biophys.J. **85**, 3445-3459 (2003).
- [21] A. D. MacKerell (2001) in: *Computational Biochemistry and Biophysics* (O. M. Becker, A. D. MacKerell, B. Roux, and M. Watanabe, eds.), Marcel Dekker, Inc., New York, Basel, pp. 7-38.
- [22] A. D. MacKerell, http://www.pharmacy.umaryland.edu/faculty/amackere/param/force_field_dev.htm, 2000.
- [23] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian 98, 1998. Pittsburgh PA, Gaussian, Inc.
- [24] M. R. Harnden., R. L. Jarvest, A. M. Z Slavin., and D. J. Williams, Nucleosides&Nucleotides **9**, 499 (1990).
- [25] J. Sponer, J. Leszczynski, and P. Hobza, Journal of Biomolecular Structure & Dynamics **14**, 117-135 (1996).