Theor Chem Acc (1999) 101:121–125
DOI 10.1007/s00214980m124

Theoretical
Chemistry Accounts

Regular article

Molecular modeling study of an abasic DNA undecamer duplex: d(GCGTGOGTGCG)·d(CGCACTCACGC)*

L. Ayadi, C. Coulombeau

LEDSS, Laboratoire de Chimie Bioorganique, UMR CNRS 5616, Université Joseph Fourier Grenoble 1, BP 53 X, F-38041 Grenoble Cedex 9, France

Received: 1 June 1998 / Accepted: 17 September 1998 / Published online: 10 December 1998

Abstract. Molecular mechanics calculations were performed with the JUMNA program on d(GCGTGOGTGCG)·d(CGCACTCACGC) where "O" is a modified abasic site: 3-hydroxy-2-(hydroxymethyl)tetrahydrofuran. From energy minimizations, for intrahelical or extrahelical positions of the unpaired thymine, various structures with different curvatures were obtained. Dynamical properties of this abasic sequence were also investigated through the controlled studies of DNA bending. Poisson-Boltzmann calculations were used to mimic the electrostatic effect of solvent on this sequence. The lowest energy structures show an acceptable agreement with experimental data.

Key words: DNA – Abasic site – Molecular modeling – DNA curvature – DNA flexibility

1 Introduction

Apurinic/abasic sites (Ap sites) in DNA result from cleavage of N-glycosidic bonds and removal of the base. Owing to the relative instability of this bond, cleavage may occur by spontaneous hydrolysis with a relatively high frequency or enzymatically in the course of the repair of modified or abnormal bases [1, 2]. The hydrolytic process is markedly accelerated by chemical modifications of the nucleic bases (with alkylating agents, carcinogens, etc.) [3] or by physical agents (UV, γ radiations) [1, 2, 4]. Unrepaired, abasic sites can become potentially mutagenic or even lethal lesions [1, 5]. Naturally occurring Ap sites correspond to a mixture of α- and β-hemiacetals in tautomeric equilibrium with the ring-opened aldehydic form which is present less than 1% [6, 7]. This intrinsic instability of the abasic site represents a major hindrance for studying the structural and biological consequences of the presence of the lesion in the DNA strand. To avoid this instability, the 2-deoxyribose moiety "R" is replaced with a chemically stable 3-hydroxy-2-(hydroxymethyl) tetrahydrofuran [8, 10] which has been demonstrated to retain the biological properties of the natural abasic site [10]. The tetrahydrofuranyl abasic site will be named "O" (Fig. 1).

Conformational analysis is an important means for reaching a better understanding and control of the repair processes. However, a fundamental difficulty is that only a few structures of DNA duplexes containing abasic sites have already been examined by NMR and/or molecular modeling methods. These studies show that duplexes containing an unpaired purine opposite the abasic site (R or O) conserve a regular right-handed DNA geometry in which the base opposite the lesion stacks inside the helix [9–16]. Studies concerning duplexes containing unpaired pyrimidines indicate that the unpaired base may be stacked in the helix or pushed out, depending on the nature of the abasic site and the flanking sequences. Cuniasse et al. [17] showed that an unpaired thymine flanked by two cytosines and opposite the abasic site O was in equilibrium between two forms, pushed out and stacked, whose proportion depended on the temperature. For the same sequence, but with a natural abasic site R, Singh et al. [18] found that the thymine is stacked in the helix. By a combination of NMR and MD methods, Coppel et al. [19, 20] showed that an unpaired thymine flanked by two guanines and opposite the analogue O was again stacked in the helix. In this example, however, the sequence displays a kink of 30° near the lesion. We have shown by molecular modeling [21] that, for an unpaired cytosine opposite O flanked by two cytosines, the extrahelical conformation is less bent and more stable than the intrahelical form, which is in accordance with previous experimental results [17].

We present here a conformational analysis of the following sequence, where O_6 is the modified abasic site 3-hydroxy-2-(hydroxymethyl)tetrahydrofuran:

$^{5'}d(G_1C_2G_3T_4G_5O_6G_7T_8G_9C_{10}G_{11})^{3'}\cdot$
$^{3'}d(C_{22}G_{21}C_{20}A_{19}C_{18}T_{17}C_{16}A_{15}C_{14}G_{13}C_{12})^{5'}$

* Contribution to the Proceedings of Computational Chemistry and the Living World, April 20–24 1998, Chambery, France

Correspondence to: C. Coulombeau
e-mail: christian.coulombeau@ujf-grenoble.fr

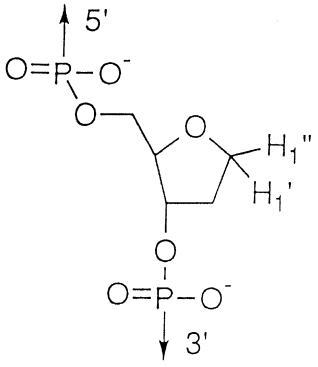

Fig. 1. The tetrahydrofuran analogue O

to elucidate the conformational perturbations introduced into DNA by the abasic lesion involving an unpaired thymine flanked by pyrimidines. The calculated results have been correlated to NMR data [17].

2 Methodology

All calculations were performed using the JUMNA (JUnction Minimization of Nucleic Acids) algorithm, which has been the subject of a number of previous publications [22, 23]. JUMNA models DNA flexibility by a combination of helicoidal parameters describing the position of each nucleotide with respect to a common helical axis system. Initial optimizations used B-DNA fiber coordinates as starting points [24]. To investigate a large conformational space and to locate different stable conformations, we used a wide variety of restraints on helical deformations (twisting or stretching) or on the sugar angles and/or amplitude.

JUMNA offers also the possibility to impose and control the curvature of DNA, by using the superhelical symmetry [25, 26]. This option can be used for obtaining a relationship between energy and DNA curvature, giving a constant which quantifies the relative flexibility of a sequence, which is an important parameter for its biological activity. This option is very useful for studying the conformational and energetic aspects of bending induced by the presence of an abasic site in the sequence. Inducing bending, while maintaining symmetry, however, requires a change from helical to superhelical symmetry. This change implies several extensions to normal helical coordinates. Superhelical geometry can be imposed on nucleic acids by fixing the radius and the pitch of the superhelical pathway. The repeating unit for superhelical symmetry corresponds to the number of base pairs per turn of the double helix or a multiple of this value. We currently use repeats of 10 base pairs for an average twist of 36°. To avoid the end-effects associated with studying oligomers, 8 base pairs on either side of the repeating unit were added. This enables to us to mimic an effectively infinite polymer. Calculations were performed here with a 18mer sequence:

d(GCGTGOGTGCGCGTGOGT)·
d(CGCACTCACGCGCACACA).

We choose as variable $C = 1000/R$ (with R in Å), which leads to a direct evaluation of the curvature angle α of the repeating unit considering that $\alpha \approx 10 \times 3.4 \times 180/R \approx 2C$ (in degrees) for a repeating unit containing 10 nucleotides separated by a average interval of 3.4 Å. Energy variations were studied as a function of the radius of curvature, allowing the DNA to rotate around its helical axis (equivalent to allowing the direction of curvature to be free).

In order to take into account environmental effects, counterion interactions are modeled by reducing the net charge on each phosphate to $-0.5e$, while solvent electrostatic damping is modeled by introducing a sigmoidal distance dependence of the dielectric function as proposed by Hingerty et al. [27]. Although this approach has led to good results concerning DNA fine structure [28], it does not include changes in solvation enthalpy. Earlier studies [29, 30] have shown that these effects can be treated more correctly using Poisson-Boltzmann electrostatics. Poisson-Boltzmann calculations were carried out here with the DelPhi program [31–34] in three "focusing" steps using successively smaller boxes around the DNA oligomer. In the first of these runs, 30% of the box is filled by the oligomer and Coulomb boundary conditions are used. Subsequent runs were performed for 60% and 90% of the box filled to achieve a good final accuracy. The internal dielectric constant of DNA was set to 2 to mimic solute polarizability and the external dielectric constant was set to 80. The probe sphere radius defining the accessible solvent surface was set to 1.05 Å. To correct the electrostatic energy E_{ES} of the pre-optimized conformations, E_{ES} was replaced by the reaction field energy E_{RF} and coulombic energy E_C calculated with DelPhi program. The total energy thus becomes:

$$E_{J-PB} = E_J - E_{ES} + E_C + E_{RF}$$

where E_J is the total JUMNA energy. Lastly, all DNA conformations were analyzed using CURVES [35, 36].

3 Results and discussion

The calculations described in the first section of the methodology were started from more than 100 differently distorted structures for intrahelical conformations. The extrahelical conformations, where both the abasic sugar (or backbone region) and the unpaired thymine on the complementary strand are protruding out of the helix, were generated using various approaches such as permutation between the unpaired base and abasic site in which the abasic sugar is shifted towards the groove. In each case we have calculated the solvent contributions using Poisson-Boltzmann electrostatics as described in the methodology.

A first large selection amongst the results was based on geometrical grounds (nonphysical distortions) or energetic grounds (energy differences greater than 15 kcal mol^{-1} with respect to the most stable conformations) and allowed many of the structures resulting from these different minimization pathways to be discarded and led to a total of 15 acceptable structures, 9

structures for the intrahelical form and 6 for the extrahelical form. Table 1 lists all the total energies of these conformations and CTC_3, CTC_6, CTC_7, CTC_{14}, and CTC_{13} correspond to minimized 11mers deduced from the 18mers obtained by the superhelical symmetry calculations.

The nine intrahelical structures mainly differ by their global curvature angle and are classified in two families according to this feature: family I (seven structures) with a global curvature angle ranging from 4° to 20° and family II (two structures) with 50° and 60° for the curvature angle. Structures within family I are slightly bent and essentially differ by the position of the abasic sugar. In all cases the unpaired thymine and all the base pairs are in very similar conformations. ΔE_{J-PB} energy differences within family I (CTC_{1-7}) range from 0.7 kcal mol^{-1} to 6.7 kcal mol^{-1} with respect to the most stable conformation.

The two structures of family II ($CTC_{8,9}$) are highly bent and present similar conformational features. Their ΔE_{J-PB} energy differences are 4.1 kcal mol^{-1} and 6.3 kcal mol^{-1} with respect to the most stable conformation. Figure 2 represents the most stable conformation within each of these families. The presence of the abasic site in the sequence induces a kink near the lesion location and the direction of curvature is toward the major groove in both cases.

The six extrahelical conformations could be grouped into two families representing two distinct positions of the unpaired thymine within the groove. Family III (CTC_{10-12}) has the unpaired thymine in the major groove and family IV (CTC_{13-15}) has the thymine in the minor groove. The most stable conformations for each family are shown in Fig. 3. For both families we observe a good stacking between the two base pairs flanking the abasic site, in agreement with the experimental results [17]. In CTC_{13} the unpaired thymine protrudes into the minor groove, making a hydrogen bond between its amino group and O3′ of G9 in the other stand. In CTC_{10} the thymine is situated in the major groove, making three hydrogen bonds between T17(HN3) and G3(O6), T17(O4) and C2(H2N), and T17(O2) and A15(H2N). This result is in line with the findings of Joshua-Tor et al. [37] and Miller et al. [38], who have demonstrated by X-ray crystallography that an extrahelical base could lie in the groove of the DNA in close proximity to one of the neighboring nucleotides.

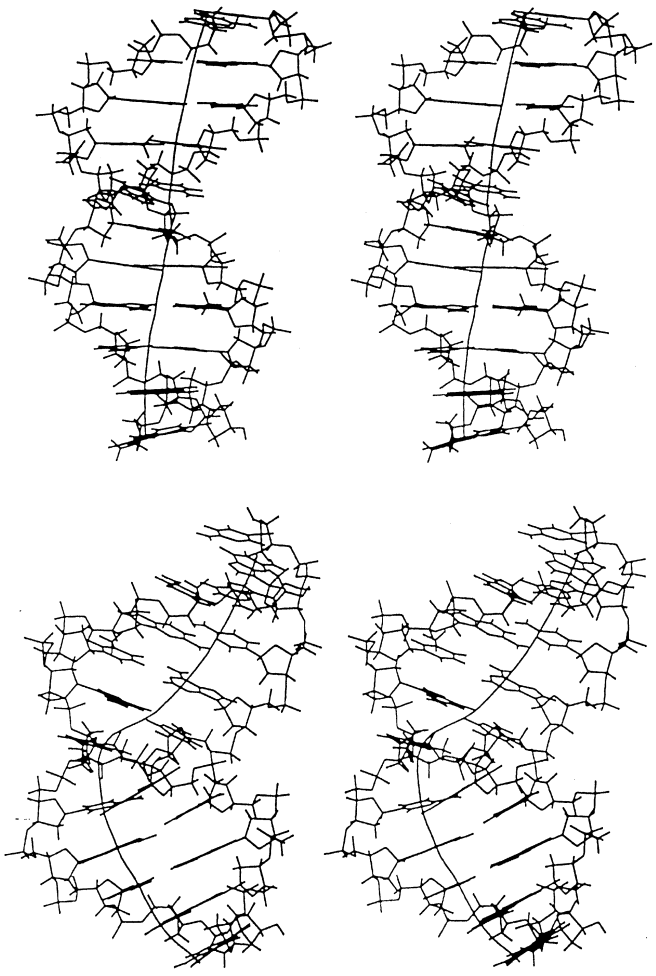

Fig. 2. Stereoviews of the two intrahelical conformations with two different degrees of curvature [39]

Table 1. ΔE_J, ΔE_{J-PB} and total curvature angle (°) of the stable conformations

Intrahelical conformations	ΔE_J	ΔE_{J-PB}	Total curvature angle (°)
Family I			
CTC_1	8.5	0.7	15
CTC_2	8.9	2.3	19
CTC_3	13.4	2.7	3
CTC_4	9.2	4.3	14
CTC_5	12.8	5.0	5
CTC_6	15.2	5.7	7
CTC_7	14.6	6.7	11
Family II			
CTC_8	7.5	4.1	50
CTC_9	10.5	6.3	60
Extrahelical conformations			
Family III			
CTC_{10}	4.6	1.6	31
CTC_{11}	5.0	3.7	11
CTC_{12}	5.2	4.6	18
Family IV			
CTC_{13}	0.0	0.0	18
CTC_{14}	3.5	2.0	21
CTC_{15}	7.8	2.9	22

If, despite the limitations of our force field and solvent model, we use a finer screening, eliminating sructures with energy differences greater than 1.6 kcal mol^{-1} with respect to the lowest energy, then only the two structures CTC_1 and CTC_{13} remain (all other being less than 4% in the mixture in terms of a Boltzmann average, assuming that the entropy contribution differences between conformers are negligible and liken the calculated differences with free energy differences). Both conformations have a small global curvature angle, 15° and 18°, and are less curved than the abasic sequence containing an unpaired thymine flanked by two guanines [20].

As a final step, the flexibility of our sequence was investigated by performing bending deformations. The

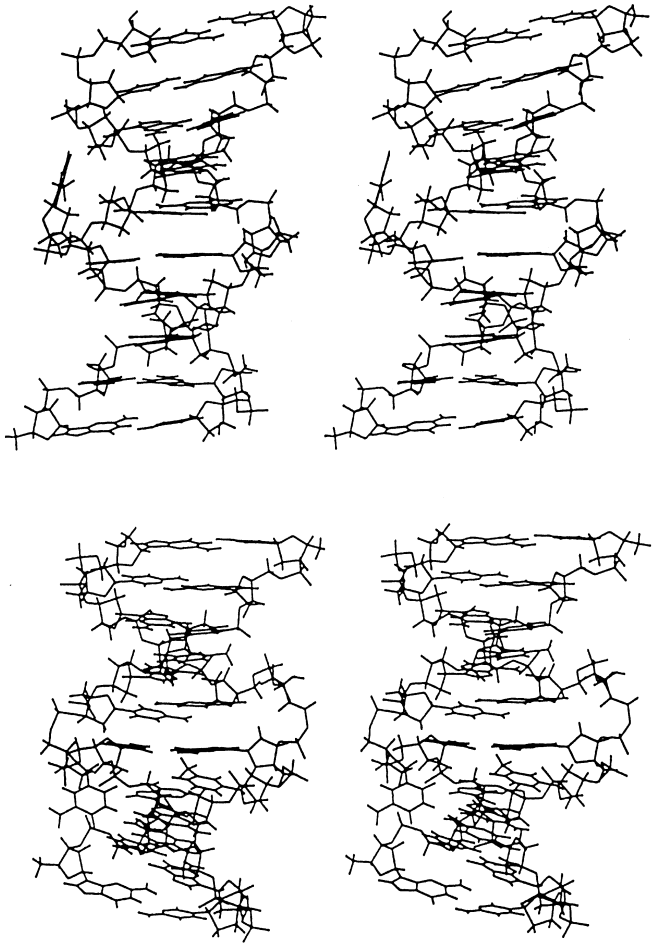

Fig. 3. Stereoviews of the two extrahelical conformations representing the unpaired thymine in the minor (*above*) or major groove (*below*). Molecular structures are displayed using Insight II 970 of Biosym Technologies, San Diego

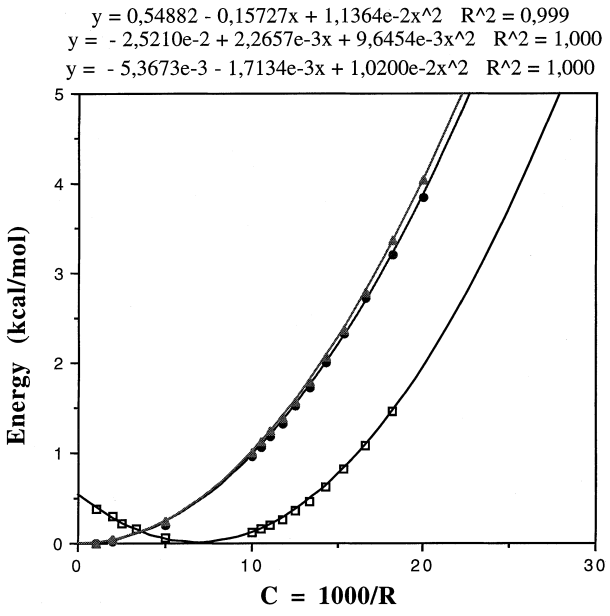

Fig. 4. Deformation energy (kcal mol^{-1}) for the intrahelical conformation (□), $(GT)_n$ (●), and $(GC)_n$ (▲) studied as a function of curvature expressed as $1000/R$ (Å^{-1})

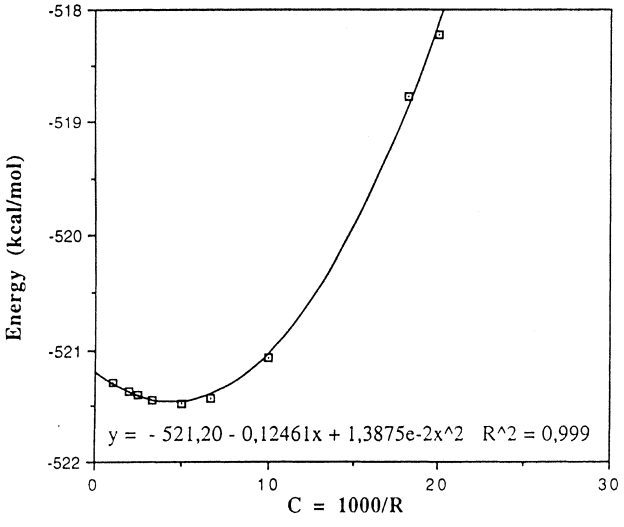

Fig. 5. Deformation energy (kcal mol^{-1}) for the extrahelical conformation studied as a function of curvature expressed as $1000/R$ (Å^{-1})

flexibility constants were obtained by fitting a quadratic function to the bending deformation curves. Results were compared to calculations performed on regular infinite sequences $(GT)_n$ and $(GC)_n$. The conformational energy of the DNA abasic sequence was examined as a function $C = 1000/R$ for an intrahelical position of the unpaired nucleotide $CTC_{3'}$ (Fig. 4) and for the extrahelical conformation $CTC_{13'}$ (Fig. 5). Our results show that, for this sequence, both the intrahelical and extrahelical forms are as or less flexible (1.1 and 1.4×10^4 kcal Å^2 mol^{-1}) than natural DNA (10^4 kcal Å^2 mol^{-1}). Thus, introduction of an abasic site facing an unpaired thymine in this case slightly reduces the flexibility of the sequence.

4 Conclusion

Molecular modeling has been used to study the conformation of a DNA containing an unpaired thymine opposite an abasic site. It is shown that this molecule can adopt several conformations with different degrees of curvature when the unpaired thymine is stacked within the helix. Extrahelical forms also exist and are found to be as flexible, slightly less kinked and as or more stable than the intrahelical forms. Using Poisson-Boltzmann electrostatics in this study to take into account solvent effects clearly improves on the results obtained with a simple distance dependent dielectric function. The conformations retained in accordance with this corrected energy criteria lead to a good correlation with the experimental observation of an equilibrium between an intra and extrahelical form.

Acknowledgements. The authors wish to thank R. Lavery for discussions and suggestions.

References

1. Loeb L, Preston B (1986) Annu Rev Genet 20:201
2. Lindahl T (1993) Nature 362:709
3. Singer B, Grunberger D (1983) Molecular biology of mutagens and carcinogens. Plenum Press, New York, pp 16–19
4. Teoule R (1987) Int J Radiat Biol 51:573
5. Boiteux S, Laval J (1982) Biochemistry 21:6746
6. Manoharan M, Ransom SC, Mazumder A, Gerlt JA, Wilde JA, Withka JM, Bolton PH (1988) J Am Chem Soc 110:1620
7. Wilde JA, Bolton PH, Mazumder A, Manoharan M, Gerlt JA (1989) J Am Chem Soc 111:1894
8. Millican TA, Mock GA, Chauncey MA, Patel TP, Eaton MA, Gunning J, Cutbush SD, Neidele S, Mann J (1984) Nucleic Acids Res 12:7435
9. Goljer I, Kumar S, Bolton PH (1995) J Biol Chem 270:22980
10. Takeshita M, Chang CN, Johnson F, Will S, Grollman AP (1987) J Biol Chem 262:10171
11. Wang KY, Parker SA, Goljer I, Bolton PH (1997) Biochemistry 36:11629
12. Withka JM, Wilde JA, Bolton PH (1991) Biochemistry 30:9931
13. Kalnik MW, Chang CN, Grollman AP, Patel DJ (1989) Biochemistry 28:3373
14. Vesnaver G, Chang C, Eisenberg M, Grollman AP, Breslauer KJ (1989) Proc Natl Acad Sci USA 86:3614
15. kalnik MW, Chang CN, Johnson F, Grollman AP, Patel DJ (1988) Biochemistry 27:924
16. Cuniasse P, Sowers LC, Eritja R, Kaplan B, Goodman MF, Cognet JAH, Lebret M, Guschlbauer W, Fazakerley GV (1987) Nucleic Acids Res 15:8003
17. Cuniasse PH, Sowers LC, Eritja R, Kaplan BE, Sowers LC (1990) J Mol Biol 213:303
18. Singh MP, Hill GC, Pÿoc'h, Rayner B, Imbach JL, Lown JW (1994) Biochemistry 33:10271
19. Fouilloux L, Berthet N, Coulombeau C, Coulombeau Ce, Dheu-Andries ML, Garcia J, Lhomme J, Vatton P (1995) J Mol Struct 330:417
20. Coppel Y, Berthet N, Coulombeau C, Coulombeau Ce, Garcia J, Lhomme J (1997) Biochemistry 36:4817
21. Ayadi L, Coulombeau Ce, Coulombeau C, Lavery R (1997) In: Cusack S (ed) Biological macromolecular dynamics. Adenine Press, New York, pp 83–87
22. Lavery R (1988) In: Olson WK, Sarma RH, Sarma MH, Sundaralingam M (eds) Structure and expression, vol 3. Adenine Press, New York, pp 191–211
23. Lavery R, Zakrzewska K, Sklenar H (1995) Comput Phys Commun 91:135
24. Arnott S, Chandrasekaran R, Birdsall DL, Leslie AGW, Ratliff RL (1980) Nature 283: 743
25. Sanghani SR, Zakrzewska K, Harvey SC, Lavery R (1996) Nucleic Acids Res 24: 1632
26. Sanghani SR, Zakrzewska K, Lavery R (1996) Biological Structure and Dynamics, Vol 2. Sarma RH, Sarma MH (eds) Adenine Press, New York, pp 267–278
27. Hingerty BE, Ritchie RH, Ferrell TL, Turner JE (1985) Biopolymers 24: 427
28. Poncin M, Hartmann B, Lavery R (1992) J Mol Biol 226: 775
29. Zakrsewska K, Madami A, Lavery R (1996) Chem Phys 204: 263
30. Bernet J, Zakrzewska Z, Lavery R (1997) J Mol Struct 398: 473
31. Gilson MK, Sharp KA, Honig B (1987) J Comput Chem 9: 327
32. Gilson MK, Honig B (1988) Proteins Struct Funct Genet 4: 7
33. Sharp KA, Hong B (1990) Biopolymers 32: 145
34. Biosym/MSI (1995) DelPhi and Solvation user guide. Biosym/MSI, 9685 Scranton Road, San Diego, USA
35. Lavery R, Sklenar H (1989) J Biomol Struct Dynam 6: 655
36. Lavery R, Sklenar H (1988) J Biomol Struct Dynam 6: 63
37. Joshua-Tor L, Rabinovich D, Hope H, Frolow F, Appella E, Sussman JL (1998) Nature 334: 82
38. Miller M, Harrison RW, Wlodawer A, Appella E, Sussman JL (1988) Nature 334: 85