Theor Chem Acc (2000) 104:284–289
DOI 10.1007/s002140000140

Theoretical
Chemistry Accounts

Regular article

Molecular modeling study of DNA abasic sites

Leila Ayadi, Damien Forget, Alain Martelli, Jean-François Constant, Martine Demeunynck, Christian Coulombeau

LEDSS, Chimie Bioorganique, UMR CNRS n° 5616, Université Joseph Fourier Grenoble 1,
BP 53X, 38041 Grenoble Cedex 9, France

Received: 16 September 1999 / Accepted: 3 February 2000 / Published online: 12 May 2000

Abstract. We use molecular modeling calculations to study the structure and the flexibility of abasic (AP sites) and for the design of anticancer drugs targeted against AP sites. For either adenine or cytosine on the opposing strand within the same sequence context, the results are in line with experimental data which show that the two unpaired bases lead to intrahelical forms, but with differences in induced curvature. Results on flexibility, indicate that the two duplexes have the same bending rigidity for cytosine. In previous work a series of polyfunctional molecules, such as ATAc, were designed to selectively recognize and cleave abasic sites in DNA. The nitrobenzamide group which was added to the ATAc molecule to obtain a new molecule, termed ATAc4, can induce a second lesion under irradiation in close proximity to the abasic site. The different conformations of ATAc4 interacting with a DNA oligomer containing a stable analog of the abasic site were compared to the photoinduced cleavage pattern observed experimentally.

Key words: AP sites – DNA – Molecular modeling – Flexibility – Curvature

1 Introduction

Abasic sites (AP sites) which are created by the loss of a nucleic base (purine or pyrimidine) are the most frequent lesions encountered in DNA. These lesions may arise from a variety of routes, including chemical modifications (protonation or alkylation of a purine base) [1], modification by physical agents (UV radiation) [2–4] or enzymatic repair of modified or abnormal bases [2, 3].

An abasic site corresponds to a mixture of α and β hemicetals in tautomeric equilibrium with the ring-opened aldehyde form (less than 1%) [5, 6].

If unrepaired, this lesion may be mutagenic or lethal for the cell [2, 7]. The repair of AP sites is thus essential to cell viability. Structural knowledge of such sites is important for understanding how repair enzymes recognize and repair these lesions. Due to its intrinsic instability, which hinders the study of the biological and structural properties, the 2-deoxyribose moiety is replaced by the model 3-hydroxy-2-(hydroxymethyl) tetrahydrofuran in many experiments [8, 9]. Structural studies [10–24] of the DNA duplexes containing abasic sites indicated several intrahelical and extrahelical conformations for the natural or the modified abasic site as well as for the unpaired base. In a previous report [23] we described the deformation energy of eight sequences obtained with the JUMNA program without any experimental constraints and found good agreement with available experimental data. We propose applying the same procedure to two new sequences experimentally described by Berger and Bolton [24]. They used spectroscopic data in conjunction with constrained molecular dynamics to obtain detailed conformations for an 11-mer containing an abasic site opposite unpaired adenine (TAT) and unpaired cytosine (TCT) between two pyrimidines (TT) opposite a natural abasic site (F).

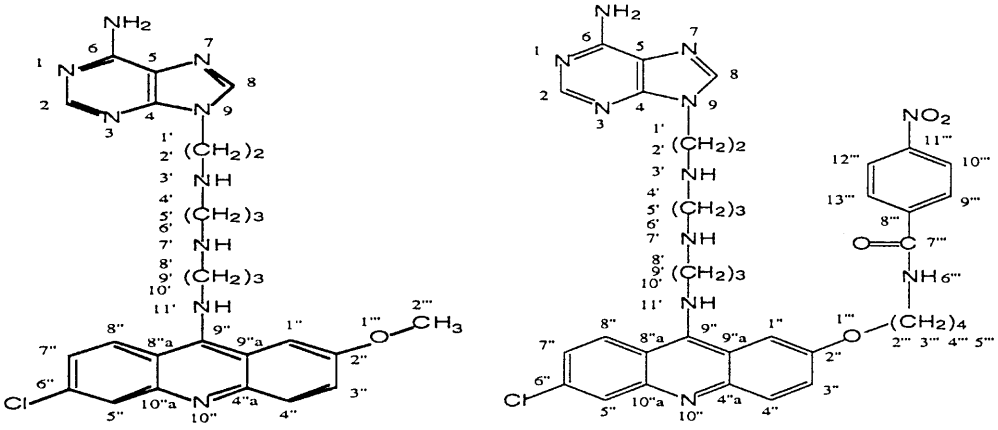
TAT C_1 G_2 C_3 G_4 A_5 F_6 A_7 C_8 G_9 C_{10} C_{11}
G_{12} C_{13} G_{14} C_{15} T_{16} A_{17} T_{18} G_{19} C_{20} G_{21} G_{22}

TCT C_1 G_2 C_3 G_4 A_5 F_6 A_7 C_8 G_9 C_{10} C_{11}
G_{12} C_{13} G_{14} C_{15} T_{16} C_{17} T_{18} G_{19} C_{20} G_{21} G_{22}

The striking difference between the two oligomers is that the abasic site exists predominantly in the β hemiacetal form in the case of the unpaired cytosine, whereas in the case of the unpaired adenine the β and α forms are equally present. Beside the structural perturbations introduced by the presence of the abasic site, flexibility of DNA is of crucial importance in many processes involving the recognition of DNA by proteins or drugs. Recently, Marathias et al. [25] found that damaged DNAs are more flexible than the corresponding undamaged duplex and suggested a general or global recognition of damaged DNAs on the basis of their flexibility by DNA repair. They also found that the TCT duplex is more flexible than

Correspondence to: C. Coulombeau

Fig. 1.



the TAT duplex. We performed the calculations on these sequences with the O6 fragment, which is an abasic site analog, 3-hydroxy-2-(hydroxymethyl) tetrahydrofuran.

In previous work [26–31], a series of polyfunctional molecules, such as ATAc (Fig. 1), containing a nucleic base (adenine) linked by a diamino chain to an intercalator (9-amino-6-chloro-2-methoxyacridine) were designed to selectively recognize and cleave abasic sites in DNA. A structural analysis of the complex between ATAc and a DNA oligomer (GTG) containing the stable analog of the abasic site "O" was performed by NMR and molecular modeling studies.

GTG	C	G	C	A	C	O	C	A	C	G	C
	G	C	G	T	G	T	G	T	G	C	G

This study revealed that the base moiety docks in the abasic site, possibly interacting with the unpaired thymine in a Hoogsteen mode, the acridine ring being exclusively intercalated on the 5′ side of the abasic site, and that the linker was located in the minor groove [19, 32]. The production of multiple DNA damages in tracks of a few base pairs leads to repair problems for the cell. For this reason, the synthesis of molecules able to induce such lesions can be an interesting strategy in anticancer drug design. As the abasic site, resulting from the loss of a base either spontaneously or after the action of physical or chemical reagents, is one of the most frequent lesions, it can be considered as a good target for this new strategy. The ATAc synthesis was modified in order to obtain ATAc4 (Fig. 1). In this new molecule, the added nitrobenzamide group can induce a second lesion under irradiation in close proximity to the abasic site [33].

The complex conformations of ATAc4 with a DNA oligomer containing a stable analog of the abasic site were studied by molecular modeling. The minimized structure obtained for ATAc from NMR experiments was used as a starting point. The aliphatic chain bearing the nitrobenzamide group was added and its interaction with DNA was allowed either in the minor or in the major groove. The different resulting conformations were compared to the photoinduced cleavage pattern observed in molecular biology studies.

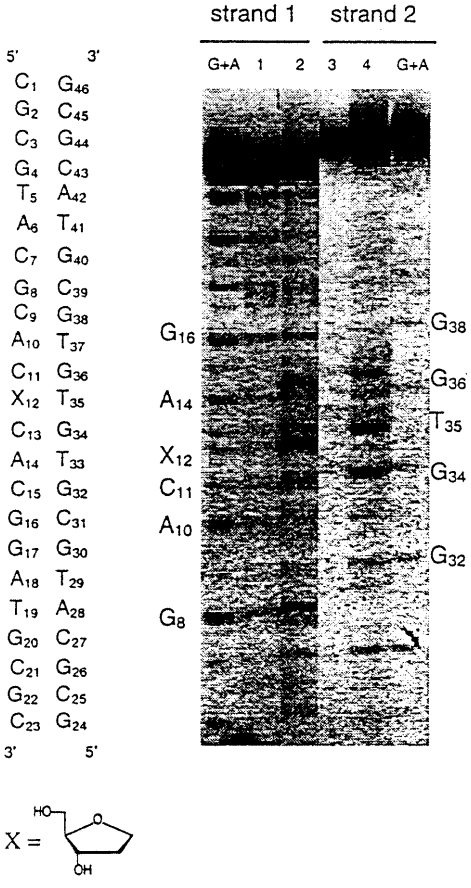


Fig. 2. Autoradiogram of 20% denaturing polyacrylamide gel showing the photocleavage of 5′-^{32}P end-labeled 23-mer oligonucleotide containing an analog of the abasic site by ATAc4. *G + A* is a sequencing method for localization of purines. *Lane 1*: abasic site containing strand (*strand 1*) alone (control); *lane 2*: strand 1 + ATAc4; *lane 3*: opposite strand (*strand 2*) alone (control); *lane 4*: strand 2 + ATAc4

2 Experimental

2.1 Materials

The molecule ATAc4 was synthesized and its abasic site recognition was tested on a 23-mer duplex DNA containing the same stable analog of the apurinic site as in the NMR studies. In these condi-

tions, we observe the photoreactivity of the nitro group of ATAc4 [34]. The oligomer structure is shown in Fig. 2. Each strand was successively 5′-^{32}P-labeled using ^{32}P-γ-ATP T4 kinase. The oligonucleotide was incubated with ATAc4 and irradiated ($\lambda > 310$ nm) at 4 °C for 3 h. After irradiation, the damaged sites were revealed by hot piperidine treatment. The results are presented in Fig. 2. A major cleavage site appears at C11 but other cleavages occur at C13, A14, G34, T35, and G36.

2.2 Methods

Calculations were carried out and structures visualized on an IBM RS 6000-39H computer running JUMNA [35, 36] version 10 and InsightII version 98.0.[1] The screening effect of the solvent was simulated using a sigmoidal distance-dependent dielectric function and counterions were simulated by reducing the charge on each phosphate group to $-0.5e$. Other solvent effects were included by the Poisson–Boltzmann electrostatic calculations carried out with the DelPhi program [37–39] as previously described [40]. The total energy of one conformation is deduced from the relation

$$E_{TOT} = E_J - E_{ES} + E_{RF} + E_C$$

where E_J is the JUMNA energy, E_{ES} the electrostatic term in JUMNA, E_{RF} the reaction field energy and E_C the Coulombic energy term in DelPhi.

The structures of the two oligomers (TAT and TCT) were obtained using the procedure described in detail in previous publications [21–23]. The flexibility calculations were performed with the superhelical symmetry option [41, 42].

ATAc4 was modeled with a total charge of $2.0e$ with either N10″—N7′ or N10″—N3′ amino groups protonated. Furthermore, to account for the conjugation of the Acr-C9″—N11′ bond, quadratic penalty terms were used to maintain the coplanarity of the acridine moiety and the NH—C group. The DNA 11-mer GTG corresponds to the central part of the experimental 23-mer with an abasic site analog in which a hydrogen replaces the base.

3 Results and discussion

3.1 AP site structure of TAT and TCT

The most stable conformations for the TAT and TCT duplexes are shown in Fig. 3. The TAT conformation is intrahelical and moderately curved with an angle of curvature of 25°. The TCT conformation is intrahelical and moderately curved with an angle of curvature of 35°. The lowest energy extrahelical conformations in the TAT and TCT duplexes are, respectively, 7.3 and 7.7 kcal · mol^{-1} less stable with respect to the most stable conformation. For both conformations curvature is toward the major groove at the level of the abasic site. All these finding are in agreement with the experimental results. Berger and Bolton [24] found that the unpaired adenine and the unpaired cytosine facing an abasic site are in the helix and that TCT is structurally more distorted than the TAT sequence. The unpaired adenine between two thymines seem to produce almost the same sequence effects on the conformation as the unpaired adenine between two cytosines; however, the presence of unpaired cytosine flanked by two thymines does not seem to produce the same sequence effects on the conformation as between two cytosines. In previous work, using NMR experiments [16] or theoretical calculations [23], the unpaired cytosine between two cytosines was found in an extrahelical conformation in the case of CCC.

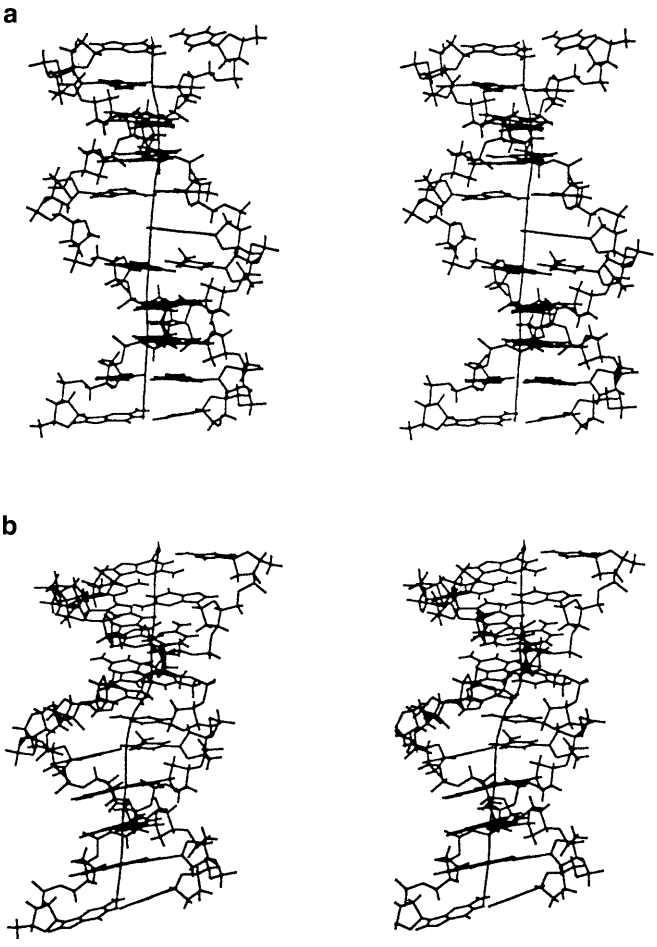


Fig. 3. Stereoviews of the most stable conformations representative of **a** the TAT and **b** the TCT duplexes

3.2 Flexibility of TAT and TCT duplexes

Since curvature in the abasic sequences was seen to depend on the sequence context, the superhelical symmetry was used to deform infinitely long DNA polymers to any desired radius of curvature. Calculations were performed with nearly straight intrahelical conformations as starting points for the TCT and TAT sequences. Calculations were also performed starting from extrahelical conformations. The energy curves obtained for the two intrahelical structures are shown in Fig. 4. For the TAT and TCT duplexes the minimum energy was found at the curvature angles of 25° and 33°, respectively (Fig. 4). For both duplexes, the stable conformations found with the superhelical option closely resemble those found for the 11-mer. The variation of the deformation energy with the curvature also allows the calculation of a bending rigidity constant by fitting a quadratic function to bending deformation curve.

The TAT sequence (2.9×10^{-19} erg cm) is as flexible as the TCT sequence (2.9×10^{-19} erg cm). The value of

[1]Molecular structures are displayed using Insight II 970 of Biosym Technologies San Diego

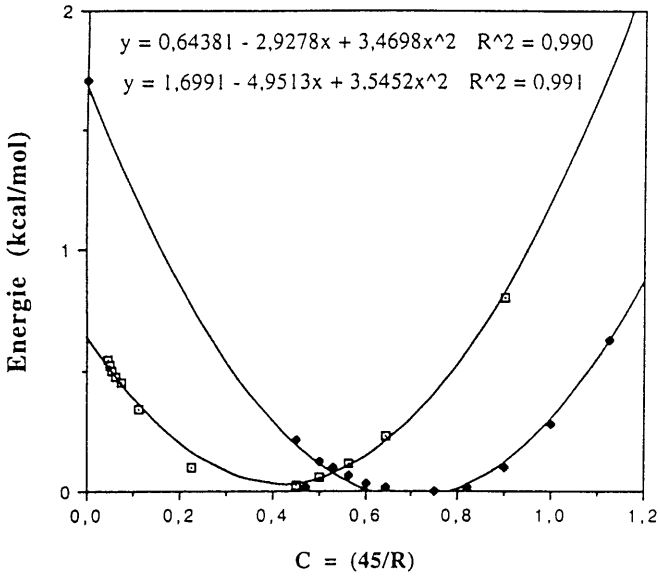


Fig. 4. Deformation energy (kcal/mol) for the TAT and TCT conformation studied as a function of curvature expressed as $45/R$ (Å^{-1})

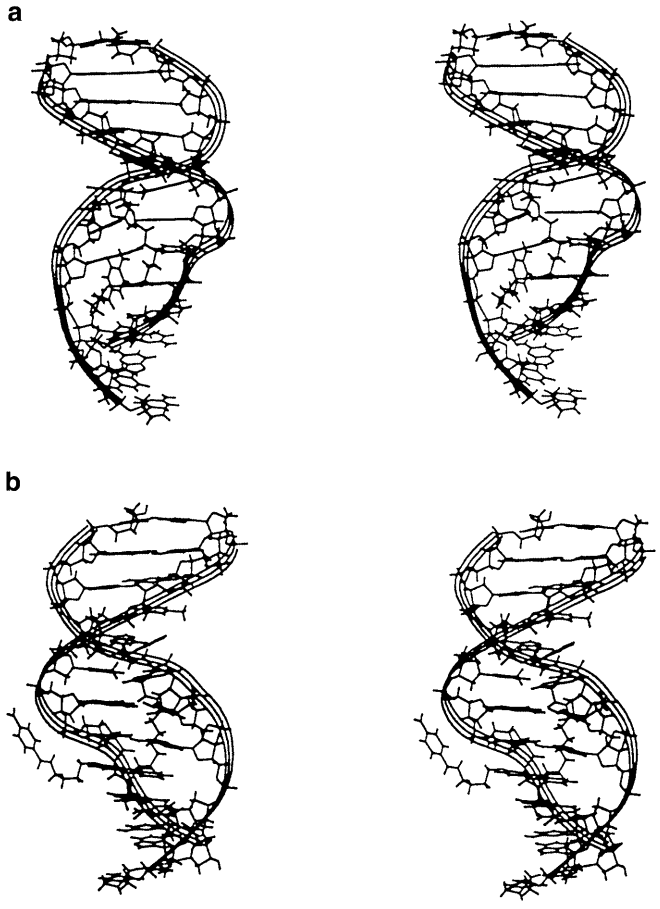


Fig. 5. Stereoviews of the most stable conformations representing the folded form: **a** in the minor groove, **b** in the major groove

2.9×10^{-19}erg cm found for the two sequences is smaller than for the natural sequences [45]. The difference between the angles of curvature should be responsible of the difference between the values obtained from recent NMR measurements of diffusion constants [25].

3.3 Conformation of the complex between the abasic duplex and ATAc4

The pK values of the amino groups in the linker are critical for cleavage activity since they affect both the ionic binding to DNA and the reaction mechanism. It has been shown that the diprotonated species is the major form (70%) at the physiological pH [43, 44], with the aminoacridine and one nitrogen of the linker protonated. Two cases will be considered for the protonated linker. Various initial and final positions can be described as coiled, stretched or folded forms. The different energy contributions and the classification related to the most stable conformation are given in Tables 1 and 2 for the protonated nitrogens N10″—N7′ and in Tables 3 and 4 for the protonated nitrogens N10″—N3′. In both cases, the stretched forms are the most stable, but do not correspond to any of the observed cleavage sites. For N10″—N7′, in the folded form in the minor groove (Fig. 5), which is only disfavored by 1 kcal/mol, one oxygen of the nitro group is 3.14 Å from H4′ of C11. This cytosine is located on the 5′ side of the abasic site, which is the most reactive position for photocleavage. Another folded form in the major groove (Fig. 5) has one oxygen of the nitro group 4.0 Å from H3′ and 3.4 Å from H5′1 of G36. The energy gap is 3.9 kcal/mol in relation to the reference and could correspond to the second reactive position. These forms are a potentially good representation because they correspond to the previously proposed form for ATAc with the nitrogen unprotoned to produce a β-elimination process.

For N10″—N3′, the energy of active form is unfavorable: the folded form in the minor groove is 6.6 kcal/mol higher in energy than the stretched conformation with an oxygen of the nitrobenzamide group 4.0 Å from H3′ and 3.5 Å from H5′2 of the abasic sugar. The other folded conformation in the major groove with a 3.8 kcal/mol energy gap has an oxygen 4.0 Å from H2′1 of G36 and a coiled conformation in the major groove with a 3.1 kcal/mol energy gap has one oxygen 2.7 Å from H4′ of C13 and is the only one able to react with the bases located on the 3′ side of the abasic site. However, the protonated N3′ atom is not well positioned to allow the cleavage reaction of ATAc. On the whole, agreement with experiment is much better for the N10″—N7′ than for the N10″—N3′ type.

The most significant result is the impossibility with our starting position to bring the nitro group and the A14 or the G34 nucleotide to close proximity: another complex must be considered for ATAc4, for example, a conformation in which acridine is intercalated in the 3′ abasic site side (unlike ATAc) or just groove-bounded, leaving more degrees of freedom to the chain.

4 Conclusion

Molecular modeling has been used to study first the conformation and the flexibility of DNA containing

Table 1. ATAc4 protonated in N10″—N7′ in the minor groove

	Jumna energy	Electrostatic energy	Coulombic energy	RF energy	Total energy	ΔE	Forms
A1	−650.15	−763.40	−7147.32	−765.25	−4578.9	+1.0	Folded
A2	−646.20	−758.95	−7152.90	−758.30	−4578.6	+1.3	Stretched
A3	−647.66	−759.46	−7152.08	−757.77	−4578.4	+1.5	Stretched
A4	−651.16	−758.54	−7143.40	−750.20	−4576.5	+3.4	Folded

Table 2. ATAc4 protonated in N10″—N7′ in the major groove

	Jumna energy	Electrostatic energy	Coulombic energy	RF energy	Total energy	ΔE	Forms
C1	−647.81	−763.46	−7149.57	−768.70	−4579.9	+0	Coiled
C2	−655.10	−774.10	−7154.10	−768.40	−4579.1	+0.8	Stretched
C3	−645.76	−762.80	−7147.22	−769.35	−4577.5	+2.4	Coiled
C4	−641.90	−762.99	−7147.87	−772.97	−4576.0	+3.9	Folded

Table 3. ATAc4 protonated in N10″—N3′ in the minor groove

	Jumna energy	Electrostatic energy	Coulombic energy	RF energy	Total energy	ΔE	Forms
B1	−645.49	−757.39	−7144.97	−770.54	−4582.0	+6.6	Folded
B2	−644.38	−756.96	−7144.98	−769.71	−4580.8	+7.8	Folded
B3	−644.41	−757.89	−7146.10	−766.92	−4578.9	+9.7	Stretched
B4	−645.83	−757.60	−7143.42	−765.26	−4578.1	+10.5	Folded

Table 4. ATAc4 protonated in N10″—N3′ in the major groove

	Jumna energy	Electrostatic energy	Coulombic energy	RF energy	Total energy	ΔE	Forms
D1	−659.17	−761.42	−7143.75	−766.67	−4588.6	0	Stretched
D2	−656.89	−759.96	−7141.66	−766.65	−4586.6	+2.0	Stretched
D3	−656.05	−767.43	−7151.73	−768.74	−4585.5	+3.1	Coiled
D4	−653.72	−763.95	−7149.97	−767.48	−4584.8	+3.8	Folded

an abasic site and second the complex conformations of ATAc4 with a DNA oligomer containing a stable abasic site. Results show that the TAT and the TCT conformations are both intrahelical and moderately curved, the TCT conformation being more distorted than the TAT conformation in accordance with experimental results; however, TAT is as flexible as the TCT duplex, in disagreement with experimental results.

Various conformations were obtained for the complex ATAc4–DNA. Results show that for the ATAc4 protonated in N10″—N7′, two structures, which are close in energy, are relevant to the observed photocleavage on C11 and G36. For the ATAc4 protonated in N10″—N3′ in the minor groove, one structure is relevant to the observed photocleavage on C13. The most significant result is the impossibility with our starting position to bring the nitro group and the A14 or the G34 nucleotide to close proximity.

References

1. Singer B, Grunberger D (1983) Molecular biology of mutagens and carcinogens. Plenum, New York, pp 16–19
2. Loeb L, Preston B (1986) Annu Rev Genet 20: 201
3. Lindahl T (1993) Nature 362: 709
4. Teoule R (1987) Int J Radiat Biol 51: 573
5. Manoharan M, Ransom SC, Mazumder A, Gerlt JA, Wilde JA, Withka JM, Bolton PH (1988) J Am Chem Soc 110: 1620
6. Wilde JA, Bolton PH, Mazumder A, Manoharan M, Gerlt JA (1989) J Am Chem Soc 111: 1894
7. Boiteux S, Laval J (1982) Biochemistry 21: 6746
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. Takeshita M, Chang CN, Johnson F, Will S, Grollman AP (1987) J Biol Chem 262: 10171
10. Goljer I, Kumar S, Bolton PH (1995) J Biol Chem 270: 22980
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. Kalnik MW, Chang CN, Johnson F, Grollman AP, Patel DJ (1988) Biochemistry 27: 924
15. Cuniasse P, Sowers LC, Eritja R, Kaplan B, Goodman MF, Cognet JAH, Lebret M, Guschlbauer W, Fazakerley GV (1987) Nucleic Acids Res 15: 8003
16. Cuniasse PH, Sowers LC, Eritja R, Kaplan BE, Sowers LC (1990) J Mol Biol 213: 303
17. Singh MP, Hill GC, Pÿoc'h D, Rayner B, Imbach JL, Lown JW (1994) Biochemistry 33: 10271
18. Fouilloux L, Berthet N, Coulombeau C, Coulombeau Ce, Dheu-Andries ML, Garcia J, Lhomme J, Vatton P (1995) J Mol Struct 330: 417

19. Coppel Y, Berthet N, Coulombeau C, Coulombeau Ce, Garcia J, Lhomme J (1997) Biochemistry 36: 4817
20. Ayadi L, Coulombeau Ce, Coulombeau C, Lavery R (1997) In: Cusack S (ed) Biological macromolecular dynamics. Adenine, New York, pp 83–87
21. Ayadi L, Jourdan M, Coulombeau C, Garcia J, Lavery R (1999) J Biomol Struct Dyn 17: 245
22. Ayadi L, Coulombeau C (1999) Theor Chem Acc 101: 121
23. Ayadi L, Coulombeau C, Lavery R (1999) Biophys J 77: 3218
24. Berger RD, Bolton PH (1988) J Biol Chem 273: 15565
25. Marathias VM, Jerkovic B, Bolton PH (1999) Nucleic Acids Res 27: 1854
26. Constant JF, Fkyerat A, Demeunynck M, Laval J, O'Connor TR, Lhomme J (1988) Nucleic Acids Res 16: 2691
27. Constant JF, Fkyerat A, Demeunynck M, Laval J, O'Connor TR, Lhomme J (1990) Anticancer Drug Des 5: 59
28. Fkyerat A, Demeunynck M, Constant JF, Michon P, Lhomme J (1993) J Am Chem Soc 115: 9952
29. Fkyerat A, Demeunynck M, Constant JF, Lhomme J (1993) Tetrahedron 49: 11237
30. Berthet N, Boudali A, Constant JF, Decout JL, Demeunynck M, Fkyerat A, Garcia J, Laayoun A, Michon P, Lhomme J (1994) J Mol Recognit 7: 99
31. Belmont P, Boudali A, Constant JF, Demeunynck M, Fkyerat A, Lhomme J, Michon P, Serratrice G (1997) New J Chem 21: 47
32. Coppel Y, Constant JF, Coulombeau C, Demeunynck M, Garcia J, Lhomme J (1997) Biochemistry 36: 4831
33. Nielsen PE, Jeppesen C, Egholm M, Buchardt O (1988) Biochemistry 27: 6338
34. Martelli A (2000) Thesis. Joseph Fourier University of Grenoble
35. Lavery R (1988) In: Olson WK, Sarma RH, Sarma MH, Sundaralingam M (eds) Structure and expression Vol 3. Adenine, New York, pp 191–211
36. Lavery R, Zakrzewska K, Sklenar H (1995) Comput Phys Commun 91: 135
37. Gilson MK, Sharp KA, Honig B (1987) J Comput Chem 9: 327
38. Gilson MK, Honig B (1988) Proteins Struct Funct Genet 4: 7
39. Biosym/MSI (1998) DelPhi and Solvation user guide. Biosym/MSI, 9685 Scranton Road, San Diego, USA
40. Ayadi L (1999) Thesis. Joseph Fourier University of Grenoble
41. Sanghani SR, Zakrzewska K, Harvey SC, Lavery R (1996) Nucleic Acids Res 24: 1632
42. Sanghani SR, Zakrzewska K, Lavery R (1996) In: Sarma RH, Sarma MH (eds) Biological structure and dynamics. Adenine, New York, pp 267–278
43. Berthet N, Constant JF, Demeunynck M, Michon P, Lhomme J (1997) J Med Chem 40: 3346
44. Thomas F, Michon J, Lhomme J (1999) Biochemistry 38: 1930
45. Ayadi L, Coulombeau C, Lavery R (1999) J Biomol Struct Dyn 17: 645