

**AN OPTIMIZED PROTOCOL FOR *IN VACUO* MOLECULAR  
DYNAMICS SIMULATION AND TRAJECTORY  
ANALYSIS OF MODIFIED DNA DUPLEXES**

**Ryszard W. Adamiak\* and Lukasz Bielecki**

*Institute of Bioorganic Chemistry, Polish Academy of Sciences,  
61-704 Poznań, Noskowskiego 12/14  
adamiakr@ibch.poznan.pl bielecki@ibch.poznan.pl*

**ABSTRACT**

This paper presents an optimized protocol for *in vacuo* molecular dynamics studies of short DNA duplexes containing modified nucleosides. The example studied is an 11 base pair DNA duplex modified with 1,N<sup>6</sup>-ethenodeoxyadenosine (dεA) located opposite deoxyguanosine. A complex molecular dynamics trajectory was subjected to statistical cluster analysis. Groups (clusters) of similar conformations, which can be statistically identified after energy minimization of the trajectory results, have been analyzed with particular regard to the estimation of intrinsic stability of the duplex (hydrogen bonding and base stacking). The modified residue was found to form two hydrogen bonds within the dεA/dG base pair, which stabilize the duplex, thus allowing its conformation to remain close to regular B-DNA.

**INTRODUCTION**

During last few decades, X-ray crystallography and NMR spectrometry have been the principal methods for the evaluation of structural details of the nucleic acids and their interactions with proteins and small ligands [1], Nevertheless, it has been generally accepted that for the study of most conformational equilibria in macromolecules and their complexes other methods ought to be applied, among them computer simulation of molecular dynamics, which allows probing of conformational transitions. Combined results of all those methods should provide a stereodynamic picture of molecules in question.

Our laboratory is engaged in structural research of DNA and RNA molecules containing modified residues. One of the objects currently studied is a fluorescent 1,N<sup>6</sup>-ethenodeoxyadenosine (dεA) nucleoside which might serve as a conformational probe to study dynamic properties of nucleic acids. On the other hand this nucleosidic fluorophore is of importance from the biological point of view. 1,N<sup>6</sup>-ethenodeoxyadenosine has been proved to be formed by chemical mutagenesis when deoxy-adenosine contained in DNA is exposed to vinyl chloride or its direct metabolite chloroacetaldehyde [2], Considerable data have been collected with regard to possible biological consequences of this transition and numerous papers on

\* Author to whom correspondence should be addressed.

chemistry of related nucleosides, coenzymes and polynucleotides have been published [3], Only recently, structural data on short DNA duplexes were presented with an analysis of their NMR spectra [4] as well as an X-ray crystallography analysis [5]. In both cases formation of non-classical base pair between 1,N<sup>6</sup>-ethenoadenine and guanine was observed.

Therefore, in addition to our thermodynamic and time-resolved spectrofluorimetry studies [6], we decided to investigate molecular dynamics of DNA duplexes modified with dεA.

Recently, a physical rationale behind the application of the molecular dynamics simulations using the force field concept has been explained in the previous issue of this journal [7], Therefore now we would like to present our current protocol for computer simulations of *in vacuo* molecular dynamics of complex biomolecular systems [8,9] on the example of one helix turn of DNA duplexes modified with 1,N<sup>6</sup>-ethenodeoxyadenosine.

To evaluate the influence of the modification with the dεA residue on the overall and local structure of the duplex, sixteen double-stranded B-like DNA molecules of the following sequences:



were studied using simulation of molecular dynamics. The duplexes differed in the nucleoside opposite dεA (X = dA, dC, dG or T) and the initial orientation (syn/anti) of the N-glycosidic bond in the nucleosides of the dεA/X pair. Additionally, a similar unmodified sequence:



has been included as reference.

In this paper we would like to present the protocol of molecular dynamics simulations applied to all the above seventeen DNA molecules as well as the protocol of subsequent analysis of the dynamics trajectory. The procedure includes statistical cluster analysis of the resultant structures and will be exemplified by a DNA duplex containing the syn-dεA/anti-dG nucleoside pair, i.e. the case we have studied most thoroughly.

## METHODS

The simulations were performed using the Discover 2.9.5 program which is part of the InsightII software package (Biosym Technologies/Molecular Simulations). The Amber 4.0 force field [10], designed for nucleic acids and proteins was applied. All calculations were carried out on a Cray YMP-2 EL machine at the Poznań Supercomputing and Networking Center.

The initial structures were built as canonical B-DNA duplexes using the Biopolymer module of InsightII. As the simulations were performed *in vacuo*, i.e. without explicit solvent molecules, the partial charges on the  $PC_6^-$  oxygens were increased to -0.32 and a distance-dependent dielectric constant ( $\epsilon = 1/4r$ ) was used to mimic the presence of solvent in the calculation. No cutoffs for nonbond interactions were introduced.

The 1,N<sup>6</sup>-ethenodeoxyadenosine has been constructed from the deoxyadenosine unit. The partial charges for the atoms in the modified residue have been calculated using the DMol program (Biosym Technologies/Molecular Simulations) for a 1,N<sup>6</sup>-etheno-9-methyladenine model. The methyl group was applied as a substitute of the deoxyribose moiety in order to decrease the deserved CPU time. The modified residue was inserted into the duplex using the Biopolymer module. The figure 1 shows the structural parameters of the  $\epsilon$ Ade unit as calculated using DMol. The regular numbers show bond lengths in the molecule (in Å) and the numbers in italics represent the partial charges on the atoms.

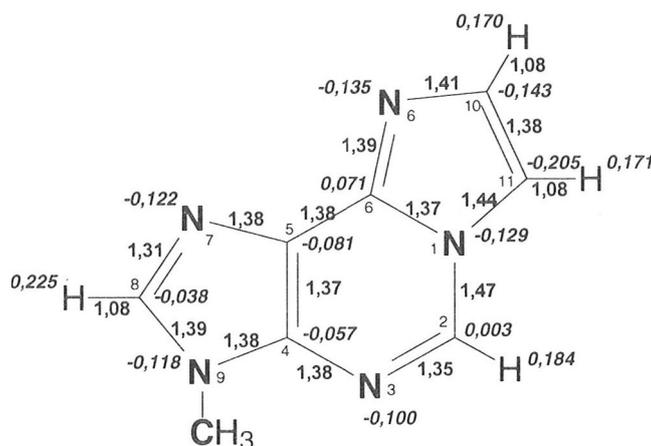


Fig. 1. 1,N<sup>6</sup>-etheno-9-methyladenine calculated with DMol.

All structures were energy minimized using 100 steps of steepest descent algorithm followed by 5000 steps of conjugate gradients. The dynamics simulation protocol consisted of two stages. During the first stage, distance restraints were imposed on Watson-Crick hydrogen bonds in two base pairs at each terminus. The restraints value was 20 kcal/mol\*Å<sup>2</sup> for each distance. The restrained MD stage consisted of 5 ps initialization (the atom velocities were assigned from a Maxwell-Boltzman random distribution in velocities at 300 K) and 95 ps simulation at 300K. The time step of 1.0 fs was used. The subsequent unrestrained stage comprised 100 ps simulation at 300 K and was followed by 5000 steps of conjugate gradients energy minimization.

This protocol is the optimized version of several protocols we have tested and it was routinely applied in our simulations. We have also tried other periods of initialization as well as iterated initializations. Surprisingly, we have found that the protocol

with 5 ps initialization provided more reproducible results than that based on iterated 0.1 ps initializations. The distance restraints stabilizing the hydrogen bonds in terminal pairs were apparently necessary to help the structure equilibrate during the first dynamics period which lasted usually 10-80 ps. We found that the restraints value of 20 kcal/mol\*Å<sup>2</sup> was optimal for stabilization of the structure. After the restrained stage, the molecular system showed much more tendency to retain its stability than it could be observed when protocols without any initial restraints were applied.

The resultant structures were visualized in InsightII at a Silicon Graphics Iris Indigo2 workstation and analysed using the Analysis module of InsightII, the Curves program [11] and our own procedures described below. The latter included statistical cluster analysis of energy minimized structures extracted from the trajectories, as well as evaluation of the stacking interactions most relevant to analysis of time-resolved spectrofluorimetry data.

In order to identify possible alternative conformations of selected duplexes, with special regard to their middle, modified regions, 101 structures from the dynamics trajectory (i.e. 2 ps snapshots) have been energy minimized for further statistical and conformational analysis. The minimization procedure consisted of 5000 steps of conjugate gradients algorithm.

Statistical cluster analysis of the minimized set has been performed in order to find groups (clusters) of similar conformers. The criterion applied to measure the similarity of two conformers was the RMS deviation of atomic positions of the sixth (modified) nucleotide pair of molecules. This procedure included calculation of the RMSD parameter for all the pairs within the set and further creation of the "lineage tree" of the structures according to the Unweighted Pair Group Method using arithmetic Averages (UPGMA) given in [12].

The parameter applied to estimate the intrinsic stability of duplexes was the consistency of stacking interactions between subsequent nucleobases. We propose this to be calculated according to the following formula:

$$S = D_1 + D_2 + D_3 + 2(|D_1 - D_2| + |D_2 - D_3| + |D_1 - D_3|) - 15.3$$

$D_1$  - distance between the N<sup>9</sup> (purines) or N<sup>1</sup> (pyrimidines) atoms

$D_2$  - distance between the N<sup>7</sup> (purines) or C<sup>5</sup> (pyrimidines) atoms

$D_3$  - distance between the C<sup>5</sup> (purines) or N<sup>3</sup> (pyrimidines) atoms

These distances are exemplified on fig. 2 on a trinucleotide motif.

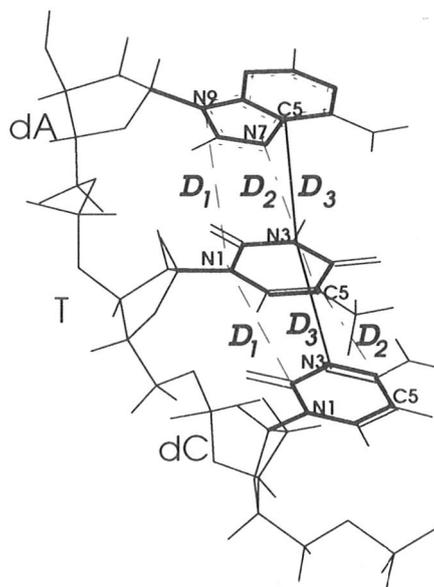


Fig. 2. Distances:  $D_1$ ,  $D_2$ ,  $D_3$  defining the stacking parameter.

This stacking parameter is close to zero when the neighboring base planes are aligned parallelly in the helix and increases with the increase of the angle between the base planes or of the distance between them. During the dynamics, the stacking interactions seem to be maintained as long as the parameter value is in range  $\pm 3$ . For greater values, the stacking is no longer observed.

## RESULTS AND DISCUSSION

### General conformational analysis

This analysis was based on the following parameters:

- hydrogen bonds geometry,
- measure of the stacking interactions within the duplex (see above),
- pseudorotation angles of the deoxyribose rings,
- duplex helical parameters defined according to the CURVES program (Lavery and Sklenar [11]).

In most of the seventeen cases studied, the resultant structures appeared to be severely bent and/or kinked, usually at the site of modification. Within each dεA/dX group of four structures, the arrangement following the syn-dεA/anti-dX pattern was the most stable one in regard to generally accepted criteria, such as maintenance of hydrogen bonds and formation of stacking base-arrays. Structures characterized by the lowest values of root mean square (RMS) deviation of atomic positions during the simulation were selected for further evaluation of conformational parameters.

### *Hydrogen bonding*

From all the sixteen modified structures calculated, the syn-dεA/anti-dG case appeared to be the most stable duplex and showed also the formation of two hydrogen bonds between the modified base and the guanine residue. This finally confirmed previous results of NMR [4] and X-ray [5] analyses.

This case has been chosen for further detailed conformational analysis, performed on two separate sets of conformations: the trajectory of 401 0.5 ps snapshots from dynamics and a parallel set of 101 minimized structures derived from every fourth of them. The former set gives us the insight into dynamic conformational changes within the molecule, whereas the latter one lets us identify definite conformers which are likely to be stable.

### *Base stacking*

In most cases studied, stacking interactions were lost in a middle region of the duplex. The loss of stacking concerned not only the modified pair, but usually also the adjacent pairs and correlated with the occurrence of a kink of the molecule. In several cases also the bases in the terminal pairs showed a tendency to lose their stacking with their neighbors. Only in the unmodified reference duplex and in the syn-dεA/anti-dG case the stacking was maintained during the entire simulation within the whole molecule.

### Sugar puckering

We found a general correlation between the degree of structural distortions in the duplexes (axis kinking, losing of stacking or hydrogen bonding) and the tendency of deoxyribose rings to switch from the typical for DNA C2'-endo conformation to C3'-endo sugar puckering. The latter is typical for RNA. This was observed usually in the terminal base pairs and the regions of kinking. In little distorted or undistorted duplexes most nucleosides maintained their usual C2'-endo conformation.

### Duplex helical parameters

A detailed analysis of the duplexes studied has been performed using the CURVES program written by Lavery and Sklenar [11]. The results of this analysis exceed the size of this paper and will be presented elsewhere [6].

### Conformational analysis of the DNA duplex containing the syn-dεA/anti-dG base pair

The "lineage tree" of the 101 minimized structures, resulting from the statistical cluster analysis is shown on fig. 3.

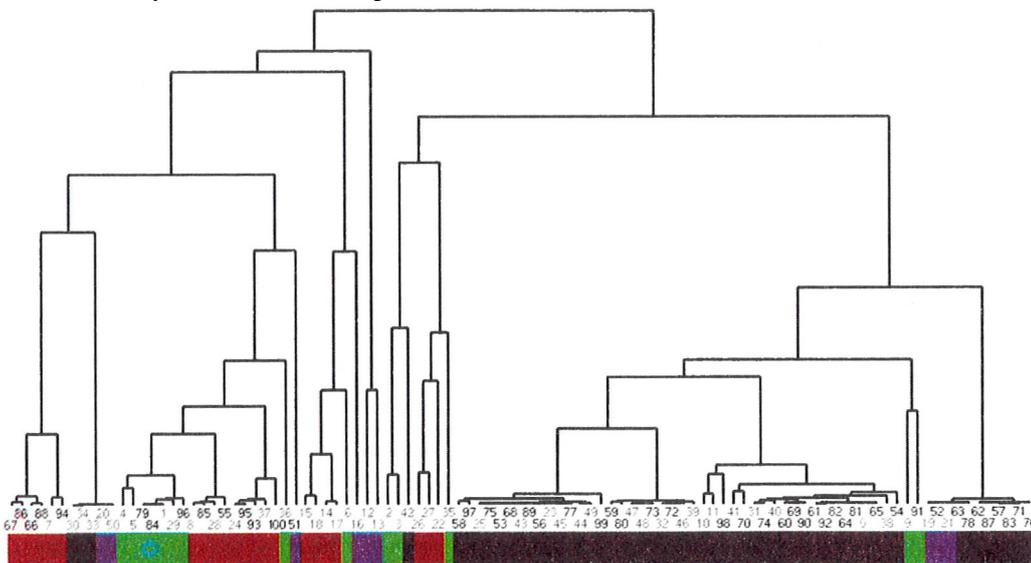


Fig. 3. The energy minimized structures statistically grouped in clusters. Three clusters which were most represented are distinguished by coloured strips under numbers responding to those structures.

The colors used refer to the images of those conformations displayed on fig. 5.

The green color marks all structures that do not belong to any of the three main clusters.

The graph on the fig. 4 shows the values of three distances between atoms in dεA and dG where hydrogen bond formation is possible. The thick lines refer to the set of minimized structures whereas the thin lines concern the original dynamics trajectory. 2.5 Å is considered as the limit of the presence of a hydrogen bond.

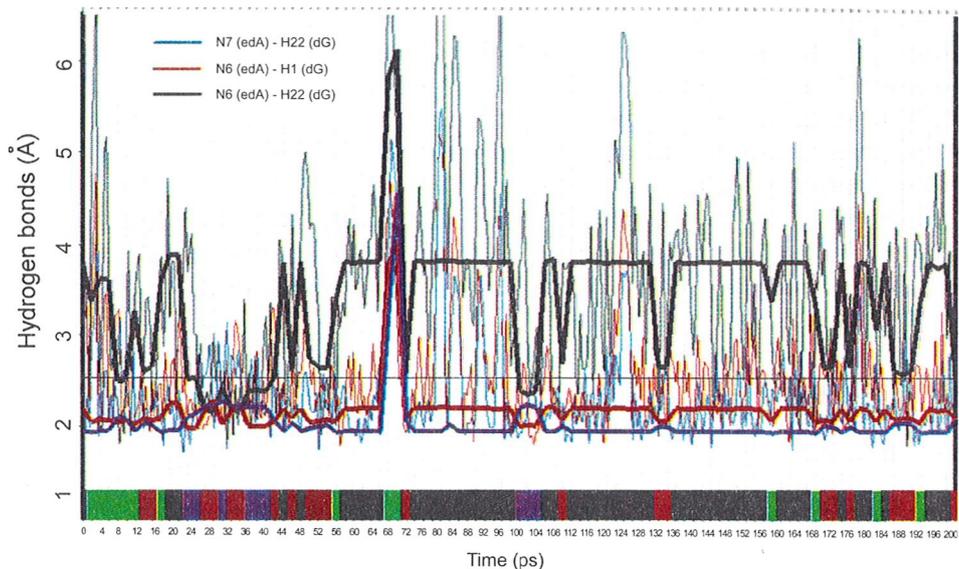


Fig. 4. Hydrogen bonds during the dynamics simulation (thin lines) and after energy minimization (thick lines).

It can be clearly seen that although during the dynamics these distances vary within a relatively broad range, after minimization they tend to converge to a small number of definite values. They represent mostly three alternative alignments of the middle region in the molecule, which can also be identified as clusters on the "lineage tree" (fig. 3). The distribution of the clusters is shown on the coloured strip along the x-axis. The black colour represents the prevailing cluster (black molecules on the fig. 5).

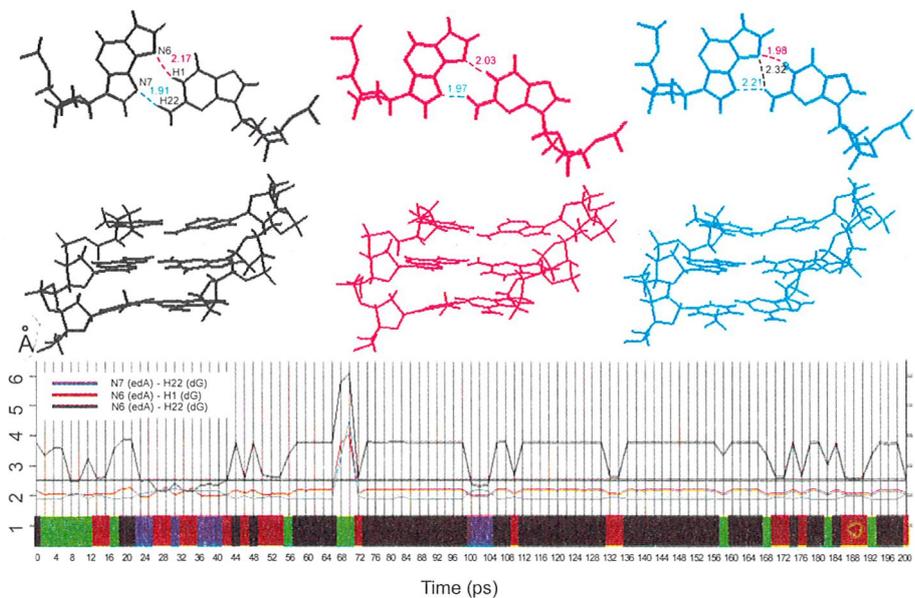


Fig. 5. Prevailing conformations within the energy minimized set.

Three conformers representing the most frequent clusters are shown on fig. 5. The images of the modified pair and the middle three nucleotide pairs are presented. The hydrogen bonds formed are shown in colors responding to the lines on the graph below. The graph depicts the distances between the respective atoms in the 101 minimized structures which are also marked with the color of the cluster they belong to (on the strip below the graph).

Nearly 60% of the minimized structures belong to one "black" cluster (represented by the left image on fig. 5). These conformers are characterized by helicoidal parameters very close to those in a canonical B-DNA helix. Two hydrogen bonds are formed.

Interestingly, the structures derived from the first period of dynamics are far more diverse and show little tendency to converge to the clusters. This would suggest that the equilibration of the dynamical molecular system was not completed until ca. 50 ps of the MD run.

The graphs on the fig. 6 show the values of the stacking parameter for all the base steps in the duplex during the dynamics (red lines) and for the respective energy minimized structures (black lines). It can be clearly seen that this parameter is close to zero for nearly all minimized conformations and during the dynamics it varied within a relatively narrow range. Alterations to the eAde stacking are clearly visible but of small amplitude. This proves that the base stacking was well preserved in this modified DNA duplex.

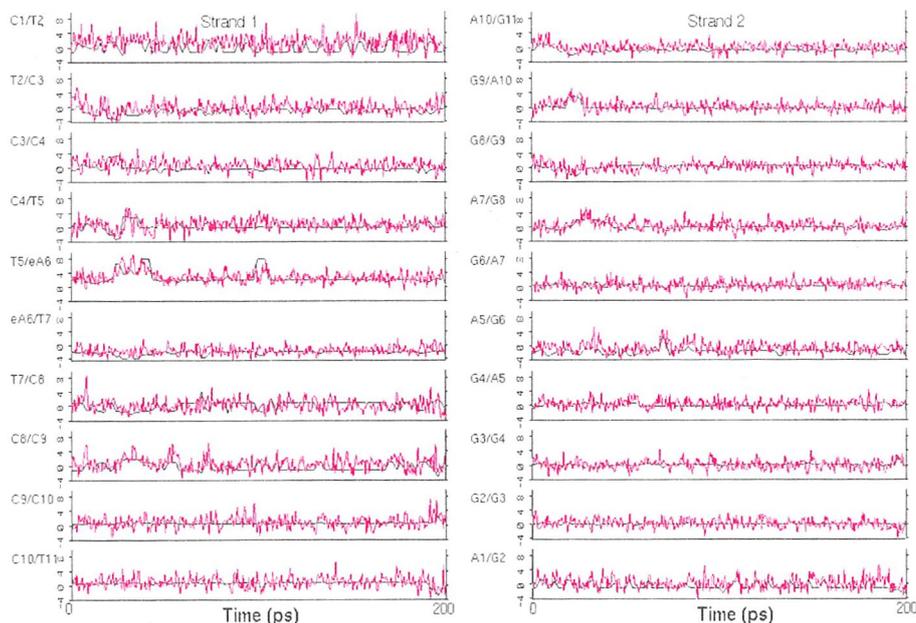


Fig. 6. Base stacking within the duplex.

Also the conformation of the sugar rings in the duplex presented does not significantly differ from the B-DNA model (fig. 7). In most of the nucleosides the

canonical C2'-endo sugar pucker is preserved, both in the dynamics (red lines) as in the minimized structures (black lines).

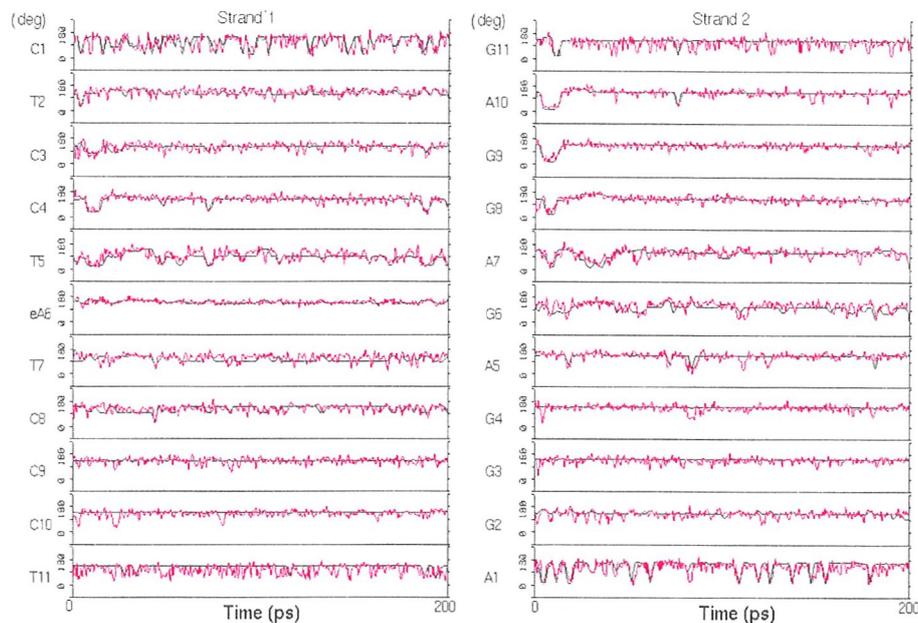
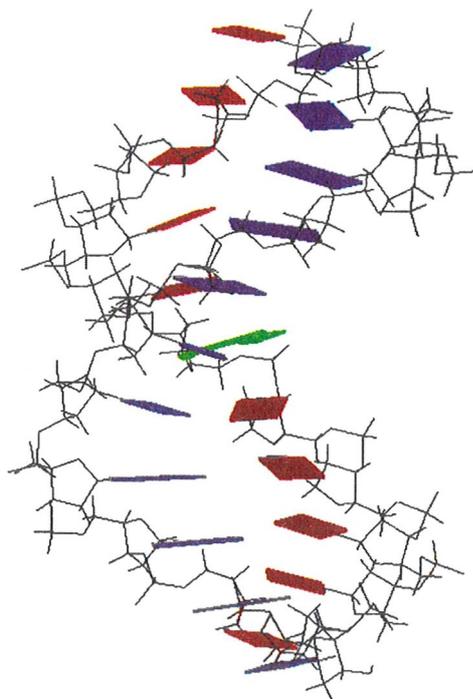


Fig. 7. Pseudorotational angles (sugar pucker) in the duplex.

The results of the conformational analysis of this duplex prove that the syn-oriented  $\epsilon$ A residue may form a stable pair with deoxyguanosine, thus stabilizing the modified duplex and allowing it to maintain the shape close to regular, typical for the B-DNA helix. All the conformational parameters tested are comparable with those of an unmodified duplex (data not shown).

The overall view of an example conformer from the main cluster is shown on fig. 8. The bases are schematically represented by filled rectangles in order to emphasize their stacking alignment. The regularity of the duplex can be clearly observed.

Fig. 8. Overall view of the duplex. The bases from the first (polypyrimidine) strand are shown in red and the purines from the second strand in blue. The  $\epsilon$ Adenucleobase is shown in green.



## ACKNOWLEDGEMENTS

This work was supported by KBN grant No. 8T11F01008P01. Authors thank Professor Bogdan Lesyng for the information on phosphate charge distribution and help in the initial DMol calculations and M. Popenda for discussion. Support of Foundation for Polish Science and Poznan Supercomputing and Networking Center is acknowledged.

## REFERENCES

- [1] Saenger, W. "Principles of Nucleic Acids Structure", Springer Verlag, Heidelberg (1984)
- [2] Singer, B., Antocchia, A., Basu, A. K., Dosanjh, M. K., Frankel-Conrat, H., Gallagher, P. E., Kusmierk, J. T., Qiu, Z. H. & Rydberg, B. (1992) *Proc.Natl. Acad. Sci. USA* **89**, 9386-9390
- [3] Leonard, N. J., (1992) *Chemtracts - Biochemistry & Mol. Biol.* 3, 273-297
- [4] de los Santos, C., Kouchakdjian, M., Yarema, K., Basu, A., Essigman, J. & Patel, D. J. (1991) *Biochemistry* **30**, 1828-1835
- [5] Leonard, G. A., McAuley-Hecht, K. E., Gibson, N. J., Brown, T., Watson, W. P., & Hunter, W. N. (1994) *Biochemistry IS*, 4755-4761
- [6] Zagórowska, I., Skalski, B., Bielecki, L., Popenda, M., Biala, E., Paszyc, S., Verrall, R. & Adamiak, R. W., in preparation
- [7] Kuliński, T. (1996) *Computational Meth. Sci. Tech.* **1**,43-54
- [8] van Gunsteren, W. F. & Berendsen, H. J. C. (1990) *Angew. Chem. Int. Ed. Engl.* **29**, 992-1023
- [9] Kumar, S., Duan, Y., Kollman, P. A. & Rosenberg, J. M. (1994) *J. Biomol. Str. Dyn.* **12**, 487-525
- [10] Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comp. Chem.* 7, 230-252
- [11] Lavery, R. & Sklenar, H. (1988)./ *Biomol. Str. Dvn.* 6, 63-91
- [12] Walkowiak, R. & Podlesny, A. (1995) *Rocz. AR Pozn. CCLXXVIII. Algor. Biometr. Stat.* **17**, 73-88