Computer Simulation of Biomolecular Interactions Involved in Genetic Processes.

VALERY POLTEV, ALEXANDRA DERIABINA, EDUARDO GONZALEZ
Facultad de Ciencias Físico Matemáticas
Benemérita Universidad Autónoma de Puebla
Av. San Claudio y Río Verde, Ciudad Universitaria, Puebla, 72570
MÉXICO

Abstract: - Computational results are described of interactions between nucleic acid bases and of searching for local minima of the interaction energy. As the bases are the elementary units of genetic material, their interactions are important for understanding the molecular mechanisms of genetic processes. The results allow us to reveal three different types of energy minima, which are responsible for various functions of nucleic acids. The minima of one type, corresponding to in-plane base arrangement and inter-base hydrogen bond formation, are responsible for the specificity of base-base interactions, i.e. for the fidelity of genetic processes. The minima of second type correspond to nearly parallel base arrangement; they, together with the first minima, stabilize the double helical DNA structure. The last type of minima corresponds to nearly perpendicular mutual base positions with single hydrogen bond. These minima are important for monomer interactions with the DNA duplex and for the duplex unwinding during nucleic acid biosynthesis.

Key-Words: - Molecular Mechanics, Computer Simulations, Intermolecular Interactions, Nucleic Acids

1 Introduction

The computer simulation is a powerful tool for investigation of molecular mechanisms of biological processes. From a physical viewpoint, life can be considered as an ordered and complicated network of physical and chemical processes. The most important of them are the processes related to the storage, replication and functioning of genetic material. These and many other processes at the molecular level in living cells involve two main steps. The first step of these processes is essentially physical one; it includes the molecular recognition and the proper arrangement of the molecules in space. The next step is the chemical reaction itself with formation and disrupt of chemical bonds. The first step is the most interesting from the biophysical viewpoint. It includes specific physical interactions, which can be simulated on computers using physical approaches and computational methods.

The sequence of four nucleotides (or four nucleic acid bases, namely, two purines, adenine A and guanine G, and two pyrimidines, thymine T and cytosine C) in polynucleotide chains contains all the information necessary for cell life. Interactions between the bases and the bases with other parts of the molecular machinery of the cell determine the processes of storage, replication, repair and expression of the genetic information. It is difficult to obtain quantitative information about base interactions by experimental methods. All such quantitative experimental data refer to a limited set of base positions corresponding to energy minima of the systems considered. But the mutual position of bases alternates during nucleic acid functioning. To understand molecular mechanisms of genetic processes in details it is necessary to evaluate energy changes resulting from base rearrangement. Theoretical and computational methods provide such a possibility.

Methods widely used for such calculations are the methods of molecular mechanics. These methods include computation of interaction energy using rather simple formulas of classical physics with adjustable coefficients (potential functions) and searching for minima of the energy using minimization techniques. Several sets of potential functions (several force fields) are available for the simulation of biopolymers (see, e.g. [1-4]. A reliability of the results obtained using the method depends on the accuracy of adjustment of coefficients to experimental data. About 20 years ago, one of us suggested a set of potential functions specially adjusted to calculation of interactions in nucleic acids

[1,2]. These functions have been used in hundreds of works in various laboratories. Now, we refine these potential functions, using new experimental data, and apply them to the extended study of nucleic acid base interactions in various mutual positions.

2 Problem Formulation

The final goal of computational studies of nucleic acids is understanding of the molecular mechanisms of genetic processes in terms of intrinsic properties of nucleic acid monomers. The specific goals of this study are a searching for various minima of base-base interaction energy, using the most reliable computational method, and an explanation of biological significance of these minima.

3. Problem Solution

3.1 Method of Calculation. Atom-atom Potential Functions for Calculation of Interactions between Nucleic Acid Monomers.

The interaction energy between nucleic acid bases is calculated by the method of semi empirical atom-atom potential functions. The energy of interaction between two molecules is considered as a sum of all the pair wise atom-atom interaction energies. The dependence of the energy of interatomic interactions (E) on the distance was approximated by 1-10-12 potential (Equation 1) for interaction of hydrogen atoms bound with N or O and proton-acceptor atom (N or O) or by 1-6-12 potential (Equation 2) for all other interactions.

$$E_{ij}=e_i e_j/r_{ij}-A_{ij}^{(10)}/r^{10}+B_{ij}^{(10)}/r^{12} \qquad (1)$$

$$E_{ij}=e_i e_j/r_{ij}-A_{ij}/r^{6}+B_{ij}/r^{12} \qquad (2)$$

In these equations r_{ij} is the distance between atoms *i* and *j*, e_i and e_j are charges on atoms *i* and *j* (calculated by semi empirical methods of quantum chemistry and reproducing the experimentally determined dipole moments of molecules). The atomic charges have not been changed as compared to our early papers [1,2]. The coefficients $A_{ij}^{(10)}$, $B_{ij}^{(10)}$, A_{ij} and B_{ij} are adjustable parameters, which have been something changed as compared to our previous studies [1,2]. The adjustment of the parameters results in better agreement with experimental data on in vacuum interaction energy between bases and on interatomic distances in hydrogen-bonded base pairs (see part 3.2), as well as on distances between base planes in crystals of nucleic acid monomers (see part 3.3). No adjustment to recent quantum mechanical results [5-7] was performed, but we have rather good agreement between our interaction energy values and the most precise quantum mechanical calculations (see part 3.2).

The geometry of bases corresponds to averaged crystal structures, and has not been changed as compared to our previous studies (e.g. [1,2]), 1-methylpyrymidines and 9-methylpurines are considered to avoid H bonding via atoms, participating in glycoside bonds in nucleic acids. The energy is calculated and minimized as a function of 6 variables, corresponding to the displacement of one base with respect to another along x, y and z axes, and to rotation around these axes. Standard minimization techniques are used throughout the paper.

3.2 In plane minima of interaction energy and the specificity of base interactions.

For every twin base combinations, there are a few minima, corresponding to nearly in-plane arrangement of two bases with N-H…N and/or N-H…O hydrogen bonds (H-bonds) between them. One of these minima of each pair is the global one. We do not consider here all the minima for all the base combinations, but only those for which the experimental values of the association enthalpy were available, and some of those, related to genetic information transfer.

Table 1. Calculated values of interaction energy of bases in global minima E in comparison with experimental enthalpies ΔH and quantum mechanical results.

N	*Pairs*	E	ΔH[8,9]	E [5]	E [7]
1	*A:T*	-12.5	-13.0	-12.4	-11.9
2	*A:C*	-13.8	-13.5	-14.3	-12.7
3	*G:T*	-13.2	-14.5	-14.3	-16.3
4	*G:C*	-25.6	-21.0	-25.4	-23.9
5	*T:T*	-9.9	-9.5	-10.5	-12.0
6	*C:C*	-17.6	-17.0	-18.8	-15.8

There are quantitative experimental data for 6 base pairs only. The enthalpies of association of bases ΔH were determined by mass-spectrometric method [8,9], and the experimental accuracy varies from 1 to 2 kcal/mol. These values can be compared with the calculated interaction energies in global minima. Table

1 provides such a comparison of our data and of most precise of recent quantum mechanical results.

The table shows, that the results of our calculations are in a good agreement with experimental data, i.e. the calculated energy values of the deepest minima for every pair (except the G:C one) are within the experimental error of measured association enthalpies. The value of interaction energy for G:C pair obtained in our calculations, and in most of other computational works, is overestimated. It is due to existence of two tautomers of guanine in gas phase, the second tautomer, enol, is extremely rare in solutions and nucleic acids. This tautomer may have biological significance, but we have not here the possibility to consider it. The average deviations of our calculated energies from the

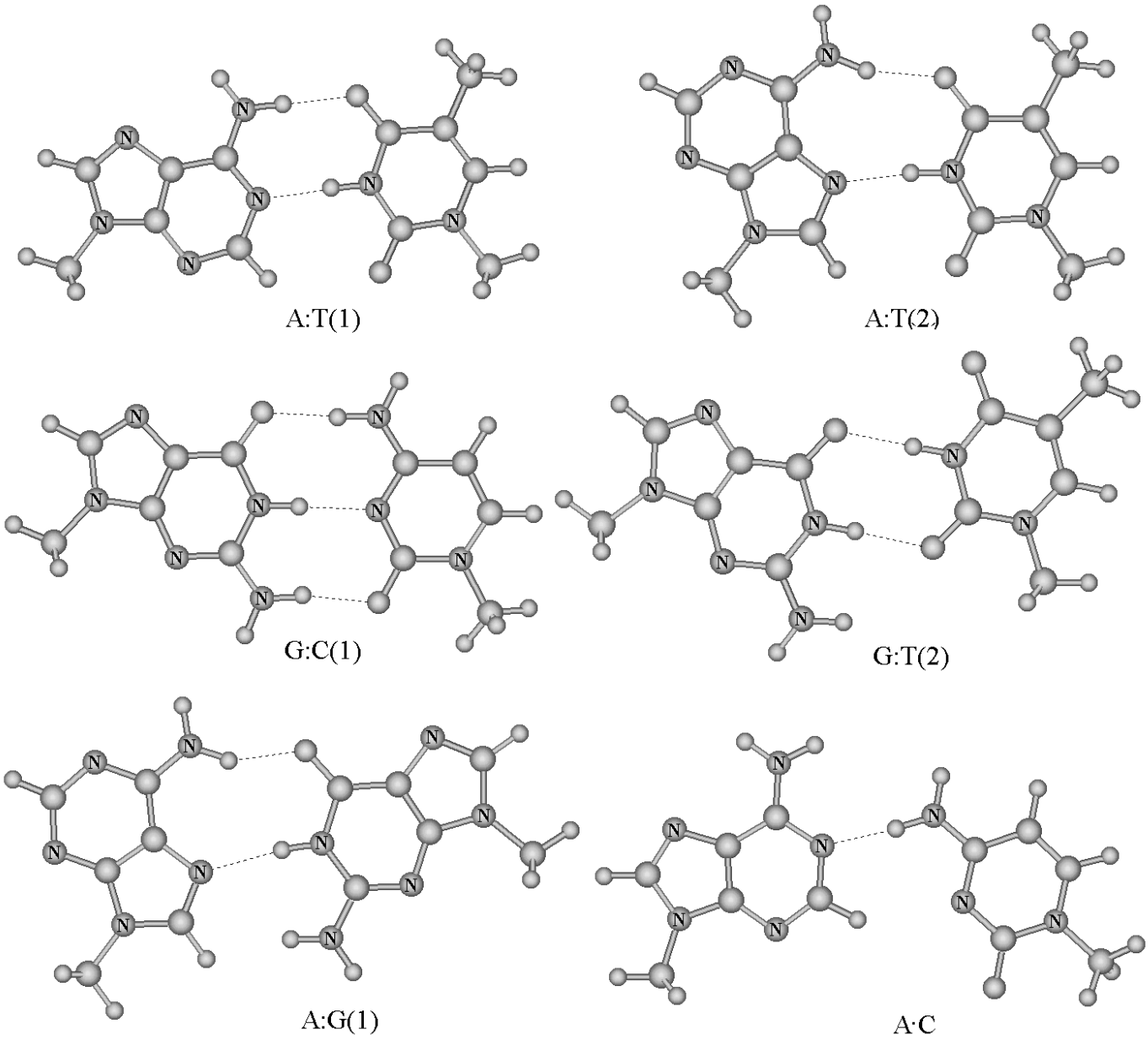

Fig.1. Some coplanar mutual positions of bases in minima of interaction energy. Nitrogen atoms are marked by N for clarity, H bonds are shown by dotted lines. A:T(1) and G:C(1) are Watson-Crick pairs. A:T(2) is the global minimum of the energy of interactions between A and T (Hoogsteen pair). G:T(2), A:G(1) and A·C are examples of pairs, which can be incorporated into DNA during biosynthesis. A·C is an example of pairs with single H bond.

experimental enthalpy values are smaller as compared to those for quantum mechanical results.

Together with rather good agreement of calculated lengths of hydrogen bonds with experimental ones [10] (numerical data not shown), and with good agreement of calculated inter-plane distances for stacked base associates (see part 3.3) with crystal data, the results of Table 1 allow to consider the method of calculation used here as a reliable one for base-base calculations.

The minima of energy for the interactions between bases, corresponding to planar base arrangements, are responsible for the coding properties of nucleic acids. Two planar pairs, so called Watson-Crick pairs, A:T(1) and G:C(1), (Fig.1) are the only normal base pairs in DNA duplex. Only these pairs are compatible with non-disturbed geometry of the polynucleotide backbone. The geometrical exclusivity of these minima provides the physical basis for the reliability of information transfer in genetic processes. The Watson-Crick pair G:C is the most stable base pair, but there are many

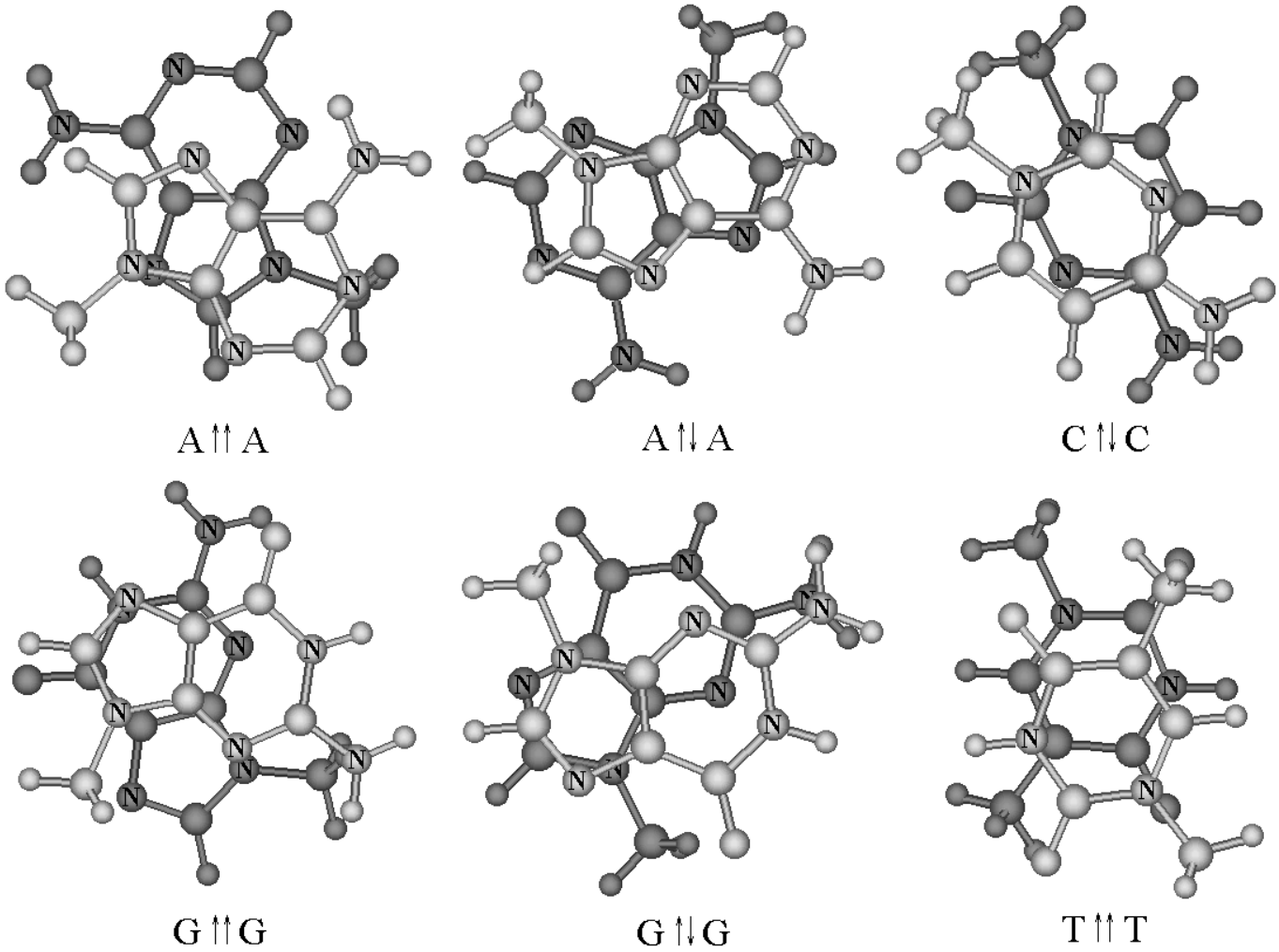

Fig.2. Some energy minima, corresponding to nearly parallel base arrangements. The view along the axis perpendicular to base plane. More dark molecules are more distant from the observer. The partial overlap of rings can be seen. The parallel arrows designate, that two molecules can be superimposed by displacements and rotation around the helix, perpendicular to molecule planes, the rotation by 180 around the axes in plane of molecule is necessary for superimpose the bases in minima designated by anti parallel arrows..

other stable planar pairs (a few of them are shown in Fig.1). Some of such pairs have more favorable interaction energy than the A:T pair, however they have distorted geometry (as compared to A:T(1) and G:C(1) pairs), and their insertion into the DNA duplex require a distortion of it. An existence, for all the twin base combinations, of interaction energy minima corresponding to a slightly disturbed backbone, provides the physical basis for point mutations, i.e. for base substitution errors during DNA biosynthesis. Such a possibility is the necessary condition for biological evolution, i.e. for the life itself.

3.3 Interaction Energy Minima Corresponding to Nearly Parallel Base Arrangement. Base Stacking and DNA Duplex Stability.

For each base combination, there are several local minima of interaction energy, which correspond to base stacking, i.e. arrangement of bases one above another in almost parallel planes. For the most of such minima, base planes form an angle of only few degrees, and the bases are partly overlapped. Some examples of such minima are shown in Fig.2. These minima are less deep, than minima with in-plane base arrangement and two H-bond formation. E.g., the energy of the deepest minimum for stacking of two adenine molecules equals to -9 kcal/mol, while the energy for deepest in-plane minimum corresponds to -11.4 kcal/mol. The distances between base planes in such minima falls in 3.3-3.5A region. Usually, these distances are somewhat

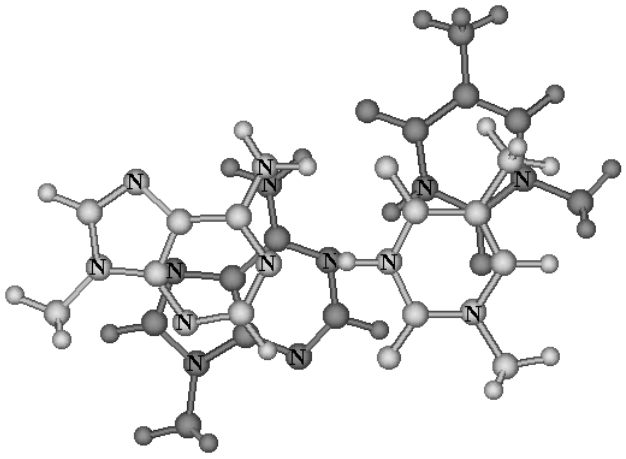

Fig. 3. Mutual arrangement of two A:T pair in one of minima of interaction energy.

greater, than in crystals of DNA monomers or in DNA helix, and the "packing effect" of other bases shorten this distance.

Base pairs in DNA double helix are arranged in similar way. In Fig.4 the position of two A:T base pair is displayed for one of the minima of pair-pair interactions. This position resembles the mutual position of two pairs in B DNA. Many other mutual positions of bases in such minima differ significantly from base arrangement in DNA helix. But the minima of this type are rather broad and shallow, i.e. a significant displacement of bases from their positions in minima is possible without great changes of energy, when the general features of a stacking arrangement are retained. As a result, the interaction of stacked base pairs is one of the main factors of conformational stability of DNA duplex.

Various base sequences in DNA have different energy of interaction between base pairs; this difference modulates the fine structure of the helix. The conformational variability of the DNA duplex, i.e. the dependence of helix parameters on base sequence, may result in specific functions of definite parts of DNA.

3.4 Interaction Energy Minima with nearly perpendicular base arrangement.

One more type of local minima of base-base interaction energy corresponds to substantially no planar base arrangement. Two bases are located in nearly perpendicular planes and form a hydrogen bond. In many cases, these minima are deeper, than minima corresponding to planar minimum of this pair with single N-H…N or N-H…O hydrogen bond. This base arrangement can be both more and less favorable than base stacking. For every base combination there are several such minima, two of them are displayed in Fig.4. Energy values for these minima are equal to –8.3 and 7.9 kcal/mol for A⊥A and A⊥T complexes respectively.

These mutual positions of bases may have biological importance. They may be involved in such processes as duplex unwinding, incorporation of new monomers into polynucleotide chain during nucleic acid biosynthesis.

Such configurations can be formed in complex systems, when other possibilities of base interactions are already used, e.g. upon interactions of monomers with DNA double helix. Perpendicular minima are of special importance for interaction of nucleic acids with such base analogs as caffeine or other conjugated molecules without H-bond donor groups.

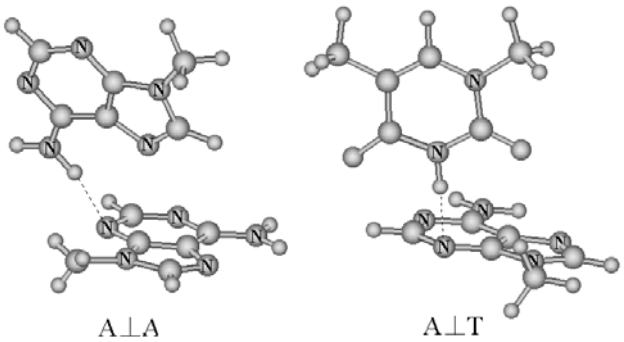

Fig. 4. Mutual positions of bases in two local minima of interaction energy corresponding to nearly perpendicular base planes. H bonds are shown by dotted lines.

4 Conclusion

Computations of interaction energy between nucleic acid monomers via molecular mechanics, and searching for minima of this energy explain some characteristic features of DNA structure and functioning. Molecular mechanics method with properly adjusted parameters allows to reproduce quantitatively available experimental data about base-base interactions in base pairs, crystals and polynucleotides.

Three types of minima of interaction energy between nucleic acid bases correspond to three main functions of DNA bases. Minima with in-plane base arrangement are responsible for molecular recognition, fidelity and possible errors of biosynthesis. The minima with stacked bases provide conformational stability and variability of the helix. Minima with nearly perpendicular base arrangement are important for interactions of the duplex with monomers, as well as for intermediate steps of helix unwinding and of pair formation.

5 Acknowledgements

This work is partially supported by the CONACyT, Mexico, grant No.35239-E and VIEP-CONACyT, Universidad Autonoma de Puebla, grant No.II28G01

References:

[1] VI. Poltev, NV. Shulyupina, Simulation of Interactions between Nucleic Acid Bases by Refind Atom-Atom Potential Functions, *Journal of Biomolecular Structure & Dynamics*, Vol.4, No.3, 1986, pp.739-765.

[2] VB. Zhurkin, VI. Poltev, VL. Florent'ev, Atom-Atom Potential Functions for Conformational Calculations of Nucleic Acids, *Molecular Biology (USSR),* Vol.14, No.5, 1980, pp.882-895.

[3] WD. Cornel, P. Cieplak, CI. Baylu, IR. Gould, K. Merz, DM. Ferguson, DC. Spellmeyer, T. Fox, JW. Caldwell and P. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules, *Journal of American Chemical Society*, Vol.117, No.19, 1995, pp.5179-5197.

[4] AD. MacKerell, Jr., J. Wiórkiewicz-Kuczera, M. Karplus, An All-Atom Empirical Energy Function for the Simulation of Nucleic Acids, *Journal of American Chemical Society*, Vol.117, No.24, 1995, pp.11946-11975.

[5] P. Hobza, M. Kabelác, J. Sponer, P. Mejzlík, J. Vondrásek, Performance of Emperical Potentials (AMBER, CFF95, CVFF, CHARMM, POLTEV), Semiempirical Quantum-chemical Methods (AM1, MNDO/M, PM3) and ab-initio Hartree-Fock Method for Interaction of DNA Bases – Comparison with Nonempirical beyond Hartree-Fock Results, *Journal of Computacional Chemistry,* Vol.18, 1997, pp.1136-1150.

[6] P. Hobza, J. Sponer, Structure, Energetics, and Dynamics of the Nucleic Acid Base Pairs: Nonempirical *Ab Initio* Calculations, *Chemical Reviews*, Vol.99, No.11, 1999, pp.3247-3276.

[7] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, E. Kaxiras, Hydrogen Bonding and Stacking Interactions of Nucleic Acid Base Pairs: A Density-functional-theory Based Treatment, *Journal of Chemical Physics*, Vol.114, No.12, 2001, pp.5149-5155.

[8] IK. Yanson, AB. Teplitsky, LF. Sukhodub, Experimental Studies of Molecular Interactions between Nitrogen Bases of Nucleic Acids, *Biopolymers*, Vol.18, No.5, 1979, pp.1149-1170.

[9] VI. Poltev, NV. Shulyupina, VI. Bruskov, AB. Teplitsky, LF. Sukhodub, IK. Galetich, Experimental and Theoretical Study of Complex Formation between Nucleic Acid Bases and Bases with Amide Group, *Journal of Biomolecular Structure & Dynamics*, Vol.9, No.1, 1991, pp.101-111.

[10] GA. Jeffrey, W. Saenger, Hydrogen Bonding in Biological Systems, Springer-Verlag New York, 1994.