A Simple Physico-chemical Mechanism of the Watson-Crick Nucleotide Base Pairs Recognition by the Proteins of the Replicative Complex

O. O. Brovarets,[1,2] L. A. Bulavin,[3] and D. M. Hovorun[2,1]

[1] Institute of High Technologies, Taras Shevchenko National University, Department of Molecular Biology, Biotechnology and Biophysics, 2 Hlushkova ave., korp.5, 03127, Kyiv, Ukraine.

[2] Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Laboratory of Computational Structural Biology, Department of Molecular and Quantum Biophysics, 150 Zabolotnoho str., 03680, Kyiv, Ukraine.

[3] Taras Shevchenko National University of Kyiv, Faculty of Physics, Chair of Molecular Physics, 2 Hlushkova ave., korp.1, 03022, Kyiv, Ukraine.

Abstract. The simple physical model of the Watson-Crick base pairs recognition in the DNA major groove by the proteins of replicative complex was firstly proposed and justifed by using modern non-empirical quantum-mechanical methods. Asparagine and Glutamine were shown to be the most credible candidates for this role. Their side chains interact with Watson-Crick base pairs by two H-bonds NH...O4/6 and C=O...HN6/4 (taking into account their rotation on 180° around single bond C-C).

Introduction

The replication of DNA *in vivo* has an extremely high precision – on the average 10^{-9} mistakes per one nucleotide [*Drake et al.*, 1969; *Alberts et al.*, 2008]. Although DNA replication plays a very important role in cell functioning, elementary physical mechanisms of this process remain poorly understood.

It was demonstrated by theoretical simulations [*Bruskov et al.*, 1974] and experimental modeling [*Morales et al.*, 2000] that *in vitro* DNA polymerase recognizes Watson-Crick nucleotide base pairs through hydrogen (H) bonds, which, on the one hand, are formed between amino acid residues of the active centre and, on the other hand, so-called invariant atomic groups of nucleotide bases and sugar-phosphate residues of nucleotides. We suppose, that such conception is incomplete, because it ignores the role of the proteins of the replication complex, which increase replication fidelity on 3-4 orders in comparison with that, obtained in the absence of these proteins in model experiments *in vitro* [*Hall et al.*, 1968]. These proteins obviously include those of the replisome [*Pomerantz et al.*, 2007], a complex of polymerases and DNA replication accessory proteins, a set of enzymes and other proteins, that support the polymerases in performing processive, accurate and rapid DNA synthesis *in vivo* (for recent reviews on replication complexes see refs. [*Loeb et al.*, 2008; *McCulloch et al.*, 2008]).

For the first time we suggested and proved the simple physical model of recognition of Watson-Crick base pairs using modern non-empirical quantum-mechanical methods. This model allows to explain the inhibition of the biosynthesis of mispairs, formed by bases in canonic tautomeric form, by the proteins of replicative complex from the side of major groove of DNA.

It was shown that only four amino acids from the 20 possible – aspartic acid (Asp), glutamic acid (Glu), asparagine (Asn) and glutamine (Gln) - can realize this function by their side chains forming intermolecular H-bonds with each of four Watson-Crick base pairs. Such H-bonds provide specificity to protein-nucleic acid interactions in the hydrophobic environment [*Dewar et al.*, 1985; *Jayaram et al.*, 2004; *Mertz et al.*, 2000]. Therefore, we have a good reason to believe that the neutral carboxyl group is more appropriate model of Asp and Glu side chains than the deprotonated one.

Objects and methods of the research

The simplest model objects we used in this work – classic Watson-Crick nucleotide base pairs and their complexes with the simplest models of amino acid side chains that belong to the proteins of the replicative complex.

The quantum-mechanical calculations of the geometrical and electronic structure were made on the DFT B3LYP/6-311++G(d,p) level of theory in the vacuum approach, which was adequate for this kind of the tasks [*Dewar et al.*, 1985].

All optimized structures were tested for the stability for the lack of imaginary frequencies in their vibrational spectra, which were calculated in the harmonic approximation.

Electronic energies of interaction between nucleotide bases in the base pairs and between the base pairs and model amino acid side chains were determined on the MP2/6-311++G(d,p)//B3LYP/6-311++G(d,p) level of theory, taking into account the so-called BSSE-correction [*Boys et al.*, 1970].

All quantum-mechanical calculations were made with the use of software package "GAUSSIAN03" for the platform Win32 [*Frish et al.*, 2004].

Intermolecular H-bonds were identified and investigated by the method of electronic density analysis [*Bader et al.*, 1990] (the so-called quantum-mechanical Bader's theory of "Atoms in molecules" (AIM)) using wave functions received at the B3LYP/6-311++G(d,p) level of theory. The topology of electronic density was analysed by the software package AIM2000. Thus energy of H-bonds was determined by the equation, offered in the work [*Espinosa et al.*, 1998].

The generally accepted numeration of atoms was used in this work.

Results and their discussion

The modelling of hypothetical center of Watson-Crick base pairs recognition by the proteins of the replicative complex was carried in two stages. First from all the 20 amino acids grounding on a high-quality spatial analysis were chosen only those of them that were able to form the simplest structures by side chains. These structures would be complementary to all of four canonical base pairs and don't suffer substantial spatial reorganization at transition from one pair to another at that. It is appeared that invariant interaction is suitable only for four amino acids – Asp and Glu which have carboxyl residue and Asn and Gln which have amide residue.

Then we took this result as a basis and learned the electronic and spatial structure of H-bonded complexes of the simplest models of these residues (HCOOH and $HCONH_2$ accordingly) with Watson-Crick base pairs of DNA.

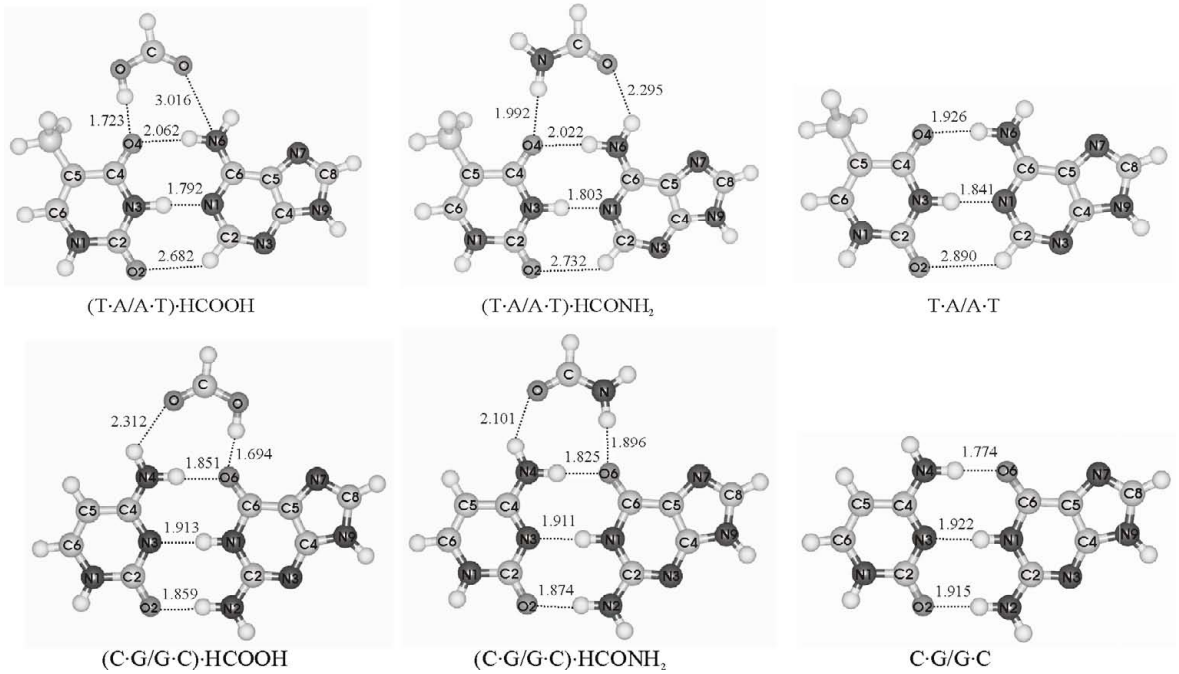

Figure 1. Geometrical structures of Watson-Crick base pairs of DNA (on the right) and their complexes with model residuals of aspartic and glutamic acids (on the left) and asparagine and glutamine (in the middle) at the B3LYP/6-311++G(d,p) level of theory. Intermolecular H-bonds are represented by the dotted lines; their lengths (the distance between atoms H and B in Å) are given near each of them; the Van der Waals contact N6…O in complex (T·A/A·T)·HCOOH is marked by the arrow.

Obtained results are presented in Figure 1 and in Tables 1, 2. Their analysis allows us to make the following discussion.

A carboxyl group HCOOH provides highest electronic energy of interaction with base pairs in complexes (T·A/A·T)·HCOOH and (C·G/G·C)·HCOOH, namely -10,46 and -15,39 kcal/mol accordingly, than amide group HCONH2 in the complexes (T·A/A·T)·$HCONH_2$ and (C·G/G·C)·$HCONH_2$, namely -7,78 and -14,50 kcal/mol accordingly.

We can conclude by comparison of the electronic energy of interaction between base pairs and residues for all investigated complexes with the total energy of the corresponding H-bonds (Tabl. 1), that contribution of the last one is dominant.

From our point of view an attraction interaction, namely – Van der Waals contact N6…O ($d_{N6...O}$=3,016 Å, ρ=0,008 a.u., Δρ=0,034 a.u.), which was detected in the complex (T·A/A·T)·HCOOH is unique. Unfortunately, the equations which would allow to estimate the energy of such interaction using their electronic topological characteristics are not reporting in literature. The amide group demonstrated noticeably higher spatial invariance of recognition center, than carboxyl group, while the last one is more energetically favourable for DNA base pairs recognition (Tabl. 1).

Table 1. The electronic-topological and energy characteristics of the intermolecular H-bonds in the investigated structures*.

Pairs of the bases and their complexes	H-bond AH..B	ρ^a	$\Delta\rho^b$	$100\cdot\varepsilon^c$	$d_{BCP-RCP}{}^d$	$E_{HB}{}^e$
(T·A/A·T)·HCOOH	OH…O4	0,039	0,134	1,27	2,74	11,25
	N6H…O4	0,019	0,069	4,93	2,02	4,01
	N3H…N1	0,045	0,094	6,23	2,35	11,47
	C2H…O2	0,006	0,020	0,22	1,47	1,14
(T·A/A·T)·$HCONH_2$	NH…O4	0,020	0,082	0,93	2,60	4,49
	N6H…O	0,012	0,047	7,88	3,02	2,46
	N6H…O4	0,021	0,076	4,79	2,04	4,55
	N3H…N1	0,044	0,093	6,31	2,31	11,04
	C2H…O2	0,006	0,018	1,12	1,42	1,03
T·A/A·T	N6H…O4	0,026	0,093	4,39	2,13	5,98
	N3H…N1	0,040	0,093	6,49	2,28	9,68
	C2H…O2	0,004	0,014	3,40	1,27	0,74
(C·G/G·C)·HCOOH	N4H…O	0,012	0,050	27,28	2,71	2,56
	OH…O6	0,043	0,137	2,34	2,68	12,57
	N4H…O6	0,031	0,108	4,84	2,14	7,83
	N1H…N3	0,033	0,087	6,81	2,18	7,54
	N2H…O2	0,031	0,105	5,45	2,18	7,51
(C·G/G·C)·$HCONH_2$	N4H…O	0,018	0,069	1,50	2,87	3,80
	NH…O6	0,026	0,101	2,45	2,59	6,28
	N4H…O6	0,033	0,114	4,28	2,14	8,63
	N1H…N3	0,034	0,087	6,83	2,15	7,63
	N2H…O2	0,030	0,101	5,56	2,17	7,14
C·G/G·C	N4H…O6	0,037	0,120	3,71	2,20	9,95
	N1H…N3	0,033	0,088	6,93	2,15	7,36
	N2H…O2	0,027	0,094	5,78	2,14	6,22

[a] The electron density value at the BCP, a.u.; [b] The Laplacian of electron density value at the BCP, a.u.; [c] Ellipticity at the BCP; [d] Distance between the bond critical point (BCP) to the ring critical point (RCP) [*Bader R. F. W. et al.*, 1990], Å; [e] H-bond energy, kcal/mol. H-bond energies were calculated by Espinosa-Molins-Lecomte method [*Espinosa E. et al.*, 1998].

Table 2. The geometrical characteristics of the H-bonds in the investigated structures*.

Pairs of bases and their complexes	H-bond AH...B	$d_{A...B}$ [a]	$d_{H...B}$ [b]	$\angle AH...B$ [c]	Δd_{AH} [d]
(T·A/A·T)·HCOOH	OH…O4	2,715	1,723	179,8	0,021
	N6H...O4	3,076	2,062	170,0	0,008
	N3H...N1	2,846	1,792	177,9	0,042
	C2H...O2	3,768	2,682	135,1	0,00013
(T·A/A·T)·$HCONH_2$	NH...O4	3,010	1,992	177,7	0,009
	N6H...O	3,304	2,295	145,1	0,003
	N6H..O4	3,037	2,022	175,0	0,008
	N3H...N1	2,857	1,803	178,1	0,041
	C2H…O2	3,818	2,732	133,6	0,00021
T·A/A·T	N6AH…O4	2,946	1,926	173,5	0,014
	N3H…N1	2,886	1,841	178,8	0,032
	C2H…O2	3,975	2,890	132,3	0,00022
(C·G/G·C)·HCOOH	N4H…O	3,320	2,312	115,4	0,003
	OH…O6	2,693	1,694	175,4	0,027
	N4H…O6	2,874	1,851	178,7	0,016
	N1H...N3	2,950	1,913	178,8	0,025
	N2H...O2	2,883	1,859	179,4	0,015
(C·G/G·C)·$HCONH_2$	N4H...O	3,113	2,101	134,5	0,007
	NH...O6	2,919	1,896	175,2	0,014
	N4H…O6	2,849	1,825	178,1	0,017
	N1H...N3	2,948	1,911	178,5	0,025
	N2H...O2	2,898	1,874	179,4	0,016
C·G/G·C	N4H...O6	2,809	1,774	178,8	0,027
	N1H...N3	2,954	1,922	177,1	0,02
	N2H...O2	2,936	1,915	178,4	0,012

[a] Distance between A (donor) and B (acceptor) atoms, Å; b Distance between H and B atoms, Å; c H-bond angle, degree; d Elongation of the chemical group AH upon H-bonding, Å.

It should be mentioned, that energies of all intermolecular H-bonds between model amino acids side chains and Watson-Crick base pairs differ one from another. In all complexes the energy of upper (from the side of the DNA major groove) H-bonds between bases decreases, while the energy of the lower (from the side of the DNA minor groove) and middle – increases in such a way that their total energy slightly differs from an analogical value for the isolated Watson-Crick base pairs (Tabl. 2).

These changes of H-bonds energy correlate well with the changes of their lengths $d_{A...B}$ and $d_{H...B}$, the elongation of the chemical bond AH Δd_{AH} and such electronic-topological characteristics as ρ and Δρ (Tabl. 1, 2): at strengthening of H-bond the first two parameters decrease, and last three - increase, and *vice versa*.

Proposed mechanism of recognition allows effectively inhibit the synthesis of mispairs G·T/T·G and A·C/C·A, using steric consideration: reason is very simple and lies in the substantial difference in geometry of these pairs from Watson-Crick base pairs [*Watson et al.*, 1953]. The aforementioned pairs can't complementary incorporate into the recognition centre, as a result it does not obtain the competent configuration for starting the chemical phase of the DNA biosynthesis.

Conclusions

For the first time we proposed and grounded the simple physical model of recognition of Watson-Crick base pairs by the proteins of replication complex in DNA major groove using modern non-empirical quantum-mechanical methods.

It was shown that the most probable candidates for this recognition are Asp and Glu, which interacts with any of four Watson-Crick base pairs through two H-bonds NH...O4/6 and C=O...HN6/4 (taking into account rotation on 180° of the amide group around single bond C-C).

Actually, Asn/Gln or Asp/Glu contacts with base pairs proposed in our model could help to eliminate the mistakes of DNA biosynthesis, arising from the formation of mispairs A·C/C·A and G·T/T·G involving the bases in the canonic tautomeric form [*Bruskov et al.*, 1974]. Also our model bears record to the adequacy of the tautomeric mechanism of point spontaneous mutations arising, firstly proposed by Watson and Crick [*Watson et al.*, 1953].

Acknowledgments. The authors sincerely thank the candidate of biological sciences Ye.P.Yurenko (Institute of Molecular Biology and Genetics of National Academy of Sciences of Ukraine) for the attention to work and corporation "Gaussian" (the USA) for software package "GAUSSIAN03" for the platform Win32, kindly granted to co-author Hovorun D.M..

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. Molecular biology of the cell,The Problems book by John Wilson, Tim Hunt. Fifth edition. New York: Garland Science. Taylor & Francis Group, 2008, 1268 p.

Bader, R. F. W. Atoms in Molecules. A Quantum Theory. Oxford: Clarendron Press, 1990, 532 p.

Bruskov, V. I., Poltev, V. I., The recognition of complementary pairs of nitrous bases by enzymes and strengthening of specificity of co-operations in the processes of matrix synthesis, *Lectures of Academy of Sciences*, 219, 231-233, 1974.

Boys, S. F., Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors, *Mol. Phys.*, 19, 553-566, 1970.

Dewar, M. J. S., Storch, D. M., Alternative view of enzyme reactions, *Proc. Natl. Acad. Sci. USA*, 82, 2225-2229, 1985.

Drake, J. W., Spontaneous mutation: comparative rates of spontaneous mutation, *Nature*, 221, 1132, 1969.

Espinosa, E., Molins, E., Lecomte, C., Hydrogen bond strenghts revealed by topological analyses of experimentally observed electron densities, *Chem. Phys. Lett.*, 285, 170-173, 1998.

Frisch, M. J. *et al.*, Gaussian 03, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

Hall, Z. W., Lehman, I. R., An in vitro transversion by a mutationally altered T4-induced DNA polymerase, *J. Mol. Biol.*, 36, 321-333, 1968.

Jayaram, B., Jain, T., The role of water in protein-DNA recognition, *Annu. Rev. Biophys. Biomol. Struct.*, 33, 343–361, 2004.

Mertz, E. L.; Krishtalik, L. I., Low dielectric response in enzyme active site, *Proc. Natl. Acad. Sci. USA*, 97, 2081-2086, 2000.

Morales, J. C., Kool, E. T., Functional hydrogen-bonding map of the minor groove binding tracks of six DNA polymerases, *Biochemistry*, 39, 12979-12988, 2000.

Pomerantz, R. T., O'Donnell, M., Replisome mechanics: insights into a twin DNA polymerase machine, *Trends Microbiol.*, 15, 156-164, 2007.

Loeb, L. A., Monnat, R. J. Jr., DNA polymerases and human disease, *Nat. Rev. Genet*., 9, 594-604, 2008.

McCulloch, S. D., Kunkel, T. A., The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases, *Cell Res.*, 18, 148-161, 2008.

Watson, J. D., Crick, F. H. C., The structure of DNA, *Cold Spring Harbor Symp. Quant. Biol.*, 18, 123-131, 1953.