ISSN 0006-2979, Biochemistry (Moscow), 2007, Vol. 72, No. 3, pp. 328-331. © Pleiades Publishing, Ltd., 2007.
Published in Russian in Biokhimiya, 2007, Vol. 72, No. 3, pp. 402-406.

"False" Thymine–1H-Enol Guanine Base Pair. Low Misinsertion Rate by DNA Polymerase Explained by Computational Chemistry Consideration

E. Seclaman[1,2], L. Kurunczi[1,2], and Z. Simon[2]*

[1]*University of Medicine and Pharmacy "Victor Babes" Timisoara, Eftimie Murgu 2, 300041 Timisoara, Romania; fax: 40256-220-479; E-mail: eddie@acad-icht.tm.edu.ro; dick@acad-icht.tm.edu.ro*
[2]*Romanian Academy, Institute of Chemistry Timisoara, Mihai Viteazul 24, 300223 Timisoara, Romania; fax: 40256-491-824; E-mail: zsimon@acad-icht.tm.edu.ro*

Received September 18, 2006
Revision received November 2, 2006

Abstract—Formation of correct TA and GC and "false" thymine–1H-enol guanine (TGenol) base pairs is here considered to control nucleotide insertion into DNA via low substrate concentration Michaelis–Menten controlled kinetics. Contributions of base pairing to formation of Gibbs free energies in water solution, $\Delta\Delta G$, are calculated for the correct and false base pairs with the semi-empiric MNDO/PM3 method for base pairing energies in vacuum and the BEM method for hydration effects. The results for $\Delta\Delta G$ indicate equal insertion rates for correct base pairing and a 10^{-3}-10^{-4} error probability for false insertion controlled by the TGenol false pair.

DOI: 10.1134/S000629790703011X

Key words: point mutation rates, misinsertion rates, enzymatic Michaelis–Menten kinetics, molecular energy in aqueous solution calculation, PM3

Fidelity in nucleotide insertion by DNA polymerase has been considered for half a century to be based largely on the Watson–Crick thymine–adenine (TA) and cytosine–guanine (CG) base pair formation. Error rates for single base substitutions due to proofreading deficient DNA polymerases vary from 10^{-3} to 10^{-6} and, in normal DNA replication, are further decreased by various error correction processes (proofreading, mismatch repair) [1]. The thymine–1H-enol guanine (TGenol) base pair (three internal hydrogen bonds) is very similar in stereochemistry to the TA and CG base pair and, in aqueous solution, at least for inosine, there is a near to 1 ratio of the keto and enol tautomers [2]. This is a serious challenge to the Watson–Crick pairing based hypothesis for securing fidelity in nucleotide insertion ([3, 4] and literature quoted therein).

Based on considerations related to first order Michaelis–Menten kinetics, the misinsertion rate via the false thymine–1-H-guanine base pair can be related to differences, $\Delta\Delta G$, of base pair Gibbs free energies of formation in aqueous solution of correct TA and CG and the false TGenol base pair [4] (see also Appendix). These differences, calculated by us, explain the low 10^{-3} to 10^{-6} misinsertion rate by DNA polymerase via the false TGenol pair.

There are several publications concerning computations of base pairing energies (see for example [5]) and also more recent computations with elaborate *ab initio* methods [6, 7], but neither imply the above considered base pairs in aqueous solution. Also, direct experimental studies concerning the corresponding equilibria do not seem to be available. Previously, we performed calculations of formation energies for these three and also other false base pairs (implying also enol tautomers of bases) by the semi-empiric PM3 method, but for vacuum (gas phase) interactions and with a reduced level of accuracy [8]. More recently, hydration energies of DNA bases were calculated by a Monte Carlo method [9].

METHODS OF INVESTIGATION

Insertion of a nucleotide in the growing DNA chain is an enzymatic process and the base pairing of the incoming nucleoside triphosphate with the base of the templating nucleotide is expected to be essential for the

Abbreviations: TGenol) thymine–1H-enol guanine base pair.
* To whom correspondence should be addressed.

formation of the corresponding Michaelis–Menten substrate–enzyme complex. For low substrate concentrations (first order kinetics) the quotient of insertion rates of a false and correct nucleotide, implicitly the error rate, can be related to the difference of formation free energies, $\Delta\Delta G$, of the false and the correct substrate–enzyme complex [4]. A 10^{-3} to 10^{-6} error rate requires differences in $\Delta\Delta G$ values of 4-8 kcal/mol (see Appendix).

We consider here the base pair formation equilibria T + A ↔ TA, C + G ↔ CG, and T + G ↔ TGenol in aqueous solution as models for the formation of substrate–enzyme complexes for DNA polymerase catalyzed nucleotide insertion. The $\Delta\Delta G$ formation free enthalpy of these base pairs will be calculated by methods of computational chemistry.

The formation free energy of the substrate–enzyme complex contains also other terms, not related to the base pair, in aqueous solution. We consider these terms equal for formation of all three complexes because the base pairs differ only in hydrogen bonding.

In *ab initio* methods molecular energies are calculated starting from isolated nuclei and electrons; for molecules of the size of nucleic acid bases, these energies amount to 10^5 kcal/mol. Base pair formation energies are small differences between large numbers. An accuracy of about 1 kcal/mol, as required by our problem, corresponds to an accuracy of 10^{-5} in *ab initio* energy calculations.

The MNDO/PM3 method implemented in HyperChem 5.11 Pro [10] is considered most appropriate for energy calculations for complexes with hydrogen bonds [11], and it was successfully used for calculations of stability and structure of 1,4-dioxane–water cluster [12]. Buz'ko et al. [12] use periodic boxes of 30-60 randomly arranged water molecules to study the 1,4-dioxane–water interaction.

We used the PM3 method for calculation of enthalpies ΔH_0 (actually these are energies) of isolated molecules out of elements in standard state and also for base pairs. Hydration free energies, ΔG_{hyd}, are calculated by the BEM method [13-16], which we used with good results for calculation of stability of ionic bonds in aqueous solution [17]. Within the BEM method, $D_{out} = 78.5$ was used for the polarized solvent (water), $D_{in} = 2.0$ for the solvated molecule ($D_{in} = 2.0$; $D_{out} = 1.0$ in vacuum). The assignments of appropriate atomic partial charges were performed using the QuACPAC software [18]. AM1BCC charges start with partial charges derived from the AM1 wave function. In a second stage, bond–charge corrections (BCC) are applied to the partial charges on each atom to generate the final partial charges. For our calculation, we consider a methyl group attached to N_1 (pyrimidine) or N_9 (purine) instead of the pentose phosphate moiety.

As an example, for calculation of formation free energy, $\Delta\Delta G$, for the TA base pair, the following equation was used:

$$T + A \rightarrow TA, \quad (1)$$

$$\Delta\Delta G(TA) = [\Delta H_0(TA) - \Delta H_0(T) - \Delta H_0(A)] + [\Delta G_{hyd}(TA) - \Delta G_{hyd}(T) - \Delta G_{hyd}(A)], \quad (2)$$

where ΔH_0 is the enthalpy of the pair formation and ΔG_{hyd} is the hydration free energy.

Neither zero point energy corrections nor thermodynamic (internal heat content at 25°C) corrections were performed. But as we are interested here in the differences

$$\Delta\Delta G_1 = \Delta\Delta G(TGenol) - \Delta\Delta G(TA),$$

$$\Delta\Delta G_2 = \Delta\Delta G(TGenol) - \Delta\Delta G(CG),$$

due to the general structural similarity of the three base pairs we assume that such corrections do not depend much upon the specific base pair, and will not significantly influence the $\Delta\Delta G_1$ and $\Delta\Delta G_2$ differences.

As our results are dependent upon ΔH_0 differences between G and Genol, *ab initio* calculations at the 6-31G** level (HyperChem 5.11) were performed by us for Genol and G; the difference $\Delta H_0(Genol) - \Delta H_0(G) = +0.04$ kcal/mol hereby obtained is rather near to the +0.83 kcal/mol value obtained by us with the PM3 method.

RESULTS

Formation enthalpies (ΔH_0) (out of standard state elements) and hydration free energies (ΔG_{hyd}) are listed in Table 1 for the bases and base pairs and compared with results from other authors [6-9]. Some structural characteristics of TA, CG, and TGenol base pairs are given in Table 2 in order to demonstrate the stereochemical similarity of the "false" TGenol pair with the Watson–Crick TA and CG pairs.

Base pair formation energies ($\Delta\Delta H_0$) out of isolated bases, hydration free energy differences ($\Delta\Delta G_{hyd}$), and formation energies in water solution ($\Delta\Delta G$) according to Eqs. (1) and (2) are listed in Table 3. The TGenol pair was considered to be formed out of T + G, as the keto G tautomer results as more stable than the 1H-enol tautomer (see Table 1).

As can be seen, reduction of hydration energy ($\Delta\Delta G_{hyd}$) in base pair formation is larger in CG and TGenol pairs with three hydrogen bonds than in TA (two hydrogen bonds) and almost equal formation free energies ($\Delta\Delta G$) are obtained for both the TA and CG pairs. The false TGenol pair has a formation $\Delta\Delta G$ increased with about 5 kcal/mol over those of the TA and CG pairs. These $\Delta\Delta G$ values represent the contribution *per se* of base pairing to the affinity responsible for formation of the enzyme–substrate complexes.

Table 1. Formation enthalpies ΔH_0 (MNDO/PM3) and hydration free energies ΔG_{hyd} (BemCalc) for 1-methylated pyrimidine derivatives and 9-methylated purine derivatives

Base/base pair	ΔH_0, kcal/mol			ΔG_{hyd}, kcal/mol		
	a	b	c	a	c	d
A	54.44	57.11	–	−10.71	–	−39.28
T	−77.09	−76.03	–	−10.83	–	−40.63
C	−14.28	−12.43	–	−14.17	–	−40.58
G	6.08	9.88	(0.00)	−17.02	(0.00)	−47.99
Genol	6.91	8.00	0.09	−14.65	3.10	–
TA	−28.29	−26.58	–	−11.58	–	–
CG	−20.36	−18.66	–	−14.46	–	–
TGenol	−77.98	−76.20	–	−11.15	–	–

Note: Genol stands for the 1H-enol guanine tautomer; a) computational results from this paper; b) results from Seclaman et al. [8] (non-methylated bases); c) results from Hanus et al. [6] (values relative to guanine, the (1,9) tautomer); d) results from Monajjemi et al. [9] with OPLS, TIP3-MonteCarlo variant.

Table 2. Structural characteristics of the TA, CG, and TGenol base pairs

Base pair	Distance, Å		
	$C(N_1)$–$C(N_9)$	N_3–N_1	C_4–C_6
TA	10.99	2.82	4.19
CG	10.80	2.80	4.20
TGenol	10.96	2.79	4.15

Note: $C(N_1)$–$C(N_9)$, distances between the carbons of the CH_3 groups attached to N_1 (pyrimidine) and N_9 (purine); N_3–N_1, distances between the N_3 (pyrimidine) and N_1 (purine) atoms; C_4–C_6, distances between C_4 (pyrimidine) and C_6 (purine) atoms. All data are PM3 results.

Concerning comparison with results of [8] from Tables 1 and 3, formation enthalpies in vacuum (ΔH_0) as well as dissociation enthalpies in vacuum ($\Delta\Delta H_0$) are similar and present identical trends. Hydration free enthalpies (ΔG_{hyd}) of DNA bases calculated by Monajjemi et al. [9] are larger than our results but present similar trends.

DISCUSSION

According to our calculations, the TGenol false pair is about 5 kcal/mol less stable than the correct TA and CG pairs, which is near to the 4-8 kcal/mol required to account for the quoted low insertion rates. The formation $\Delta\Delta G$ values for the TA and CG pairs are almost equal allowing from the standpoint of low substrate concentration Michaelis–Menten kinetics, equal incorporation rates for both correct base pairs. Further discrimination of the false TGenol pair could be due to the hypothetic mechanism of Kubitschek and Henderson [3]: DNA polymerase should interact also with groups in the major helix DNA groove of base pairs (correct pairs present both hydrogen donating and accepting groups, TGenol only hydrogen accepting groups). The available structur-

Table 3. Characteristics of pair formation energies ($\Delta\Delta H_0$ is difference between formation enthalpies ΔH_0 in vacuum (see Table 1); $\Delta\Delta G_{hyd}$ is difference in hydration free energies; $\Delta\Delta G$ is base pair formation free energy in aqueous solution)

Process	$\Delta\Delta H_0$, kcal/mol			$\Delta\Delta G_{hyd}$, kcal/mol	$\Delta\Delta G$, kcal/mol
	a	b	c	a	a
T + A → TA	−5.64	−7.66	–	+9.96	+4.32
C + G → CG	−12.16	−16.11	−21.64	+16.73	+4.57
T + G → TGenol	−6.97	−8.17	–	+16.70	+9.73

Note: See Table 1 for significance of (a) and (b); c) results from Nir et al. [7] (HF/6-31G(d,p) level with zero point energy correction).

al data for DNA polymerase–DNA interaction are insufficient for modeling such a hypothetic type of interaction.

We thank Dr. M. Mracec (Institute of Chemistry, Timisoara) for access to the HYPERCHEM packages and Dr. T. Sulea (BRC-Canada) for fruitful discussion concerning calculation of hydration energies.

This work was supported by the Romanian Ministry of Education and Research through a "Research of Excellence" grant CEEX 2979/11.10.2005.

APPENDIX

If the incorporation rate, v_{cor}, of the correct nucleotide is substantially larger than that of the rates, v_f, for the false ones, the probability p_{er} of incorporation of a false nucleotide (to a DNA template chain position) is $p_{er} \approx v_f/v_{cor}$. At low substrate concentrations, first order enzymatic incorporation rate will be:

$$v_f = \frac{k_2}{K_{mf}}[S_f],\ v_{cor} = \frac{k_2}{K_{mcor}}[S_{cor}],$$

where $[S_{cor}]$ and $[S_f]$ are here the correct and false substrate concentrations, K_{mf} and K_{mcor} the Michaelis–Menten constants for formation of enzyme–substrate complexes (here due to Watson–Crick hydrogen bonding between the template nucleotide base and the base of the incoming nucleoside triphosphate). If nucleoside triphosphate $[S_f]$ and $[S_{cor}]$ concentrations are approximately equal, the quotient p_{er} between v_f and v_{cor} depends only upon the quotient of the Michaelis–Menten constants K_{mcor}/K_{mf}. Thus, the $\Delta\Delta G$ differences between complex formation Gibbs free energies will be added to the activation free energy, $\Delta G^{\#}$, of both v_f and v_{cor}. The more thermodynamically favored the enzyme–substrate complex formation, the higher the incorporation rate. Thus:

$$p_{er} \approx \frac{K_{mcor}}{K_{mf}} = \exp\left(-\frac{\Delta\Delta G_f - \Delta\Delta G_{cor}}{RT}\right).$$

With 2.3RT = 1.38 kcal/mol at 37°C, a $p_{er} = 10^{-5}$ error level requires a $\Delta G_f - \Delta\Delta G_{cor} \cong +7.0$ kcal/mol difference.

REFERENCES

1. Kunkel, T. A., and Bebenek, K. (2000) *Annu. Rev. Biochem.*, **69**, 497-529.
2. Wolfenden, R. V. (1969) *J. Mol. Biol.*, **40**, 307-310.
3. Kubitschek, H. E., and Henderson, T. R. (1966) *Proc. Natl. Acad. Sci. USA*, **55**, 512-519.
4. Simon, Z., Seclaman, E., and Pacureanu, L. (2005) *Rev. Roum. Chim.*, **50**, 311-321.
5. Orozco, M., Cubero, E., Barril, X., Colominas, C., and Luque, F. J. (1999) in *Computational Molecular Biology (Theoretical and Computational Chemistry)* (Leszczynski, J., ed.) Elsevier Science, Amsterdam, pp. 119-166.
6. Hanus, M., Ryjacek, F., Kabelac, M., Kubar, T., Bogdan, T. V., Trygubenko, S. A., and Hobza, P. (2003) *J. Am. Chem. Soc.*, **125**, 7678-7688.
7. Nir, E., Janzen, Chr., Imhof, P., Kleinermanns, K., and de Vries, M. S. (2002) *Phys. Chem. Chem. Phys.*, **4**, 740-750.
8. Elenes, F., Fara, D., Seclaman, E., Kurunczi, L., and Simon, Z. (2001) *Rev. Roum. Chim.*, **46**, 303-308.
9. Monajjemi, M., Ketabi, S., Hashemian Zades, M., and Amiri, A. (2006) *Biochemistry (Moscow)*, **71**, S1-S8.
10. HyperChem 5.11 Pro, ChemPlus 1.6, HyperCube Inc., Gainesville, FL, www.hyper.com, 2006.
11. Stewart, J. J. P. (1989) *J. Comp. Chem.*, **10**, 209-220.
12. Buz'ko, V. Y., Sukno, I. V., Kovaleva, I. A., and Panyushkin, V. T. (2005) *Rus. J. Phys. Chem.*, **79**, 1618-1626.
13. Purisima, E. O., and Nilar, S. H. (1995) *J. Comput. Chem.*, **16**, 681-689.
14. Purisima, E. O. (1998) *J. Comput. Chem.*, **19**, 1494-1504.
15. Chan, S. L., and Purisima, E. O. (1998) *J. Comput. Chem.*, **19**, 1268-1277.
16. Sulea, T., and Purisima, E. O. (2001) *J. Phys. Chem. B.*, **105**, 889-899.
17. Sulea, T., and Simon, Z. (2002) *Internet Electron J. Mol. Des.*, **1**, 59-63.
18. QuACPAC, version 1.1.1, OpenEye Scientific Softwate, Inc., Santa Fe, NM, USA, www.eyesopen.com, 2006.