CBPF-NF-074/97

# Influence of the Hypermodified Y Base on the A•U Pairing in Codon-Anticodon Interaction

Marçal de Oliveira Neto[a], Myriam Segre de Giambiagi[b], Mario Giambiagi[b]

[a] *Departamento de Química, Universidade de Brasília, Campus Universitário Asa Norte, 70910-900 Brasília D.F. Brazil*

[b] *Centro Brasileiro de Pesquisas Físicas – CBPF/CNPq, Rua Dr. Xavier Sigaud, 150 22290-180 – Rio de Janeiro, RJ – Brazil*

Abstract

The presence of a hypermodified basis adjacent to the 3' anticodon terminal in tRNAs plays an important role in the genetic code reading mechanism. We employ the AM1 Hamiltonian for our calculations. We show here that, when the first uracil of codon UUC binds to adenine 36 of anticodon GAA, dramatic conformational variations occur in the side chain of the hypermodified Y base in position 37 of $tRNA^{Phe}$. A hydrogen bond is found to be built between one of the carboxylic oxygens in the Y side chain and one of the amino group hydrogens in adenine; this confers to A•U the stabilization energy of G•C. The calculated group-group indices and molecular valences agree with these features, yielding a simple chemical pattern. It is hence suggested that this significant stabilization of the A•U pair, due to the presence of the Y base, may prevent misreading in the genetic code.

*Key-words:* Hypermodified base Y; Codon-anticodon pairing; $tRNA^{Phe}$.

## Introduction

The fundamental role of hypermodification in the nucleotide adjacent to the 3' end of the anticodon on transfer RNA has been first discussed by Jukes in a very elegant paper [1]. These ideas were based on detailed investigations showing the presence of a hypermodified purine nucleotide following the anticodon sequence in all major tRNAs that read codons beginning by adenine (A) or uracil (U) nucleic bases, with the exception of *Escherichia Coli* $\mathrm{tRNA}^{fMet}$ [2]. It was therefore suggested that this anticodon-adjacent hypermodification stabilizes the relatively weak A•U/U•A base pairings and thereby prevents misreading of the genetic code; A•U is energetically weaker than G•C, as it contains only two hydrogen bonds and G•C three hydrogen bonds [3, 4]. In Ref. [3] it has been postulated that, so far as tRNA is concerned, the genetic code is read as four, rather than three, letters. In this work the remarkable regularity features that are observed among particular modified nucleosides adjacent to the tRNA anticodon and the recognition mechanism of the corresponding mRNA, are discussed.

Transfer RNAs which recognize codons starting by U almost always contain hydrophobic modified nucleosides, such as *$N^6$-isopentenyladenosine ($i^6A$)*. On the other hand, tRNAs which recognize codons starting by A contain hydrophylic modified nucleosides, such as *N-(9-(β-D-ribofuranosyl)purin-6-ylcarbamoyl)threonine ($t^6A$)*. However, no similar consistency appears in tRNAs which recognize a codon starting by C or G. In these tRNAs, the rather simple methylated purine nucleosides or unmodified adenosine are present.

Several experimental works have shown that the presence of a hypermodified nucleotide adjacent to the 3' side of the anticodon in tRNAs prevents misreading of the genetic code improving the recognition of fidelity the corresponding mRNA codon [5–7].

A recent theoretical work using *ab initio* quantum mechanics to obtain electrostatic potentials has shown the remarkable electrostatic effects produced on the environment of the A•U pair in the codon-anticodon pairing, by the presence of the hypermodified Y base in position 37 of $\mathrm{tRNA}^{Phe}$ [8]. These results indicate that the molecular structure of the Y base in this position may by determinant in the energetics of the codon-anticodon interaction for the stabilization of A•U pairs.

We propose here an additional step towards the elucidation of this mechanism. We study the interactions arisen from the presence of a Y base in the anticodon position 37, adjacent to the 3' side of the $\mathrm{tRNA}^{Phe}$ adenine 36. In addition to general features of the relevant energies, we report three-center (3c) hydrogen bond indices [9] and the pertinent group-group indices [10]. We show that the dramatic geometric modification in the Y side chain, as a consequence of uracil introduction, is in agreement with the values obtained for bond indices and molecular valences [10].

## Geometry modification under introduction of uracil

Fig. 1 shows the molecular structure of the hypermodified base Y. For the anticodon loop geometry we have employed the one originated from a Molecular Dynamics protocol [8] in order to allow the relaxation of the crystallographic structure [11]. Two distances in this configuration are relevant for our discussion; they involve the hydrogen atom of the amino $N_6$ group in adenine 36 (A36) which is *not* involved in the hydrogen bond between adenine and the uracil $O_2$ atom of the corresponding codon. Let us label this hydrogen H'$_6$(A). The obtained distances between it and the $O_{17}$ and $O_{22}$ in the Y base are respectively 4.6Å and 3.3Å.

The aim of this work is to verify whether the hydrogen bonds formation between A36 and the corresponding uracil in the codon-anticodon pairs are or not accompanied by significant conformational changes in the Y side chain.

If a uracil molecule is manually constrained to interact with adenine 36 in the presence of the Y37 base (within the geometry obtained through molecular dynamics [8]), the calculations show that $O_2$ in uracil is only 1.7Å from $O_{17}$ in the Y base. Now, when the geometry of Y is optimized (with the AM1 method) in the present of A•U, the strong repulsion between these two oxygens leads to dramatical conformational changes in the Y side chain; the interatomic $O_2$(U)-$O_{17}$(Y) distance goes to 3.9Å, as the torsional angle $O_{17}$-$C_{16}$-$C_{15}$-$N_{20}$ suffers a variation from $-99.63^0$ to $-24.34^0$. In turn, as a consequence, the torsional angle $C_{15}$-$N_{20}$-$C_{21}$-$O_{22}$ changes from $-14.21^0$ to $23.18^0$. The resulting distances H'$_6$(A)-$O_{17}$(Y) and H'$_6$(A)-$O_{22}$(Y) go to 5.9Å and 2.1Å respectively. So, this conformational change in the Y side chain allows H'$_6$(A) of the A36 amino group to form a hydrogen bond with the Y oxygen atom $O_{22}$. This hydrogen bond is not linear, the corresponding bond angle $N_6$(A)-H'$_6$(A)-$O_{22}$(Y) being $126^0$.

Fig. 2a displays the Y base within the tRNA$^{Phe}$ structure, as obtained from molecular dynamics procedures. Fig. 2b shows the resulting structure after AM1 optimization of the A•U pair in the presence of the stacked Y base. The most significant atoms involved in the conformational change of the Y base side chain are indicated, as well as the $C_{15}$-$N_{20}$-$C_{21}$-$O_{22}$ torsional angle $\tau$. In Fig. 3a, the A36 and Y37 structures in tRNA$^{Phe}$ are shown before the A•U pairing and Fig. 3b displays the (A•U)Y complexation.

## Interaction energies

As mentioned above, we have used in this work the semiempirical AM1 Hamiltonian. The original MNDO approximation suffers from its inability to deal with hydrogen bonds; Dewar and co-workers introduced AM1 in order to overcome this drawback, improving the core-core repulsion term for H···O and H···N interactions [12]. In a variety of hydrogen bonding problems for nonbiomolecular systems, AM1 has provides reasonable hydrogen

bond geometries [13, 14]; it has been successful in reproducing the geometry of organic systems [15]. When applied to biomolecules, AM1 has also reproduced satisfactorily experimental results [16]. Although the PM3 Hamiltonian reproduces the structures of the hydrogen bonding of nucleotide base pairs in nuclei acids, it understimates hydrogen bond energies and bond lengths for Watson-Crick pairs nucleosides [17]. According to our results, under minimization of PM3 energies, the amino hydrogens in nuclei acid base pairs stabilize in a pyramidal configuration $sp^3$ instead of the planar configuration $sp^2$.

We have chosen hence AM1, which in addition furnishes greater stabilization than PM3 for the base pair A•U in the plane of hydrogen bonding interaction. Moreover, the presence of the hypermodified Y base does not alter the coplanarity of the A•U base pair after AM1 minimization. Nevertheless, it must be remarked that AM1, as PM3, underestimates base pair interaction energies; instead, it overestimates the hydrogen bond lenghts in about 0.1−02 Å relative to experimental data [17]. Our goal being to perform comparative studies rather than details of interaction energies and bond lenghts, AM1 has proven to be a very useful tool serving our purpose.

Table 1 reports the resulting AM1 energies for the isolated molecules, and after entering the (A•U)Y complex. Table 2 shows the similar quantities for the base pairs. Our total minimization energy associated with the complex formation is −5296.04 kcal/mole, indicating −7.9 kcal/mole of stabilization energy in relation to the isolated molecules total energy, $E_I(A) + E_I(U) + E_I(Y) = -5288.15$ kcal/mole. In fact, we have seen that the flexibility of the Y base side chain makes possible for its $O_{22}$ atom to join a hydrogen bonded complex through the $H'_6$(A) (see Fig 3b). Let us point out a significant adenine stabilization in the complex. As uracil and the Y base are destabilized, the resulting stabilized complex may be due to electronic charge transfer from them to A36 (see next section).

The base pair interaction energy is the difference between the energy of the pair and the sum of the energies of each isolated base; for A•U, the tables give −4.99 kcal/mole, i.e. stabilization under pair formation. Besides, A•U in the complex is stabilized in −9.28 kcal/mole, yielding thus an overall stabilization energy of −14.27 kcal/mole. This is strikingly close to the base pair interaction energy for G•C, which we have calculated to be −13.67 kcal/mole. The complexation of the Y base with A36 in $tRNA^{Phe}$ codon-anticodon pairing seems thus to confer to the A•U pair a stabilization energy near to that of G•C. The greater energy obtained for the A−Y interaction (−13.76 kcal/mole) in relation to that of the A•U base pair in the complex (−9.88 kcal/mole) is perhaps not significant, since the Y base returns to its original conformation after codon reading.

The above results agree with those obtained in a previous work about codon-anticodon pairing interactions in $tRNA^{Phe}$ based on the analysis of the electrostatic potential created by this system in the plane containing the complementary base pair [8]. The electrostatic

potentials were calculated from *ab initio* electron density distributions of each nucleic base of the $tRNA^{Phe}$ anticodon loop; comparative studies were developed for the alternatives of a Y base or a guanine in position 37. As Y is a guanine derivative, guanine was chosen for the comparison. These results show that, in the first case (Y base), a strong negative electrostatic potential is created in the codon-anticodon interaction environment. Instead, a weaker negative potential is generated by the set of the anticodon nucleic acid bases if there is a guanine in the 3' side of the anticodon loop (position 37) of $tRNA^{Phe}$. Hence, hypermodification in this position may be important in codon-anticodon pairing. Recent results obtained using the *ab initio* quantum mechanical procedure of Ref. [18] indicate that a strong electrostatic field of 0.85 volt/Å is produced in the uracil plane of adenine 36, in the position of the deepest potential calculated in the presence of a Y base; if the base is guanine, a weaker field of 0.29 volt/Å appears. These fields interacting with the macroscopic dipole of uracil allow an estimate of the respective electrostatic energies. A greater interaction energy between the GAA anticodon loop bases of $tRNA^{Phe}$ and the corresponding UUC codon is then expected to be found in presence of a Y base.

The presence of the hypermodified Y base in position 37 of $tRNA^{Phe}$ seems thus to enhance the adjacent AU base pair interaction in the course of the codon-anticodon recognition mechanism, in agreement with experimental results [5–7]. The striking stabilization of the adenine base taking into account the additional hydrogen bond formed with the Y base chain is due mainly to a charge transfer mechanism, as we shall see next through the bond indices.

## Multicenter and group-group indices

In the linear combination of atomic orbitals (LCAO) framework, the wavefuntion $\psi_i$ of the i-th molecular orbital (MO) is defined in the atomic basis $\{\phi_a\}$ ($\phi_a$, atomic orbital centered on atom A) as

$$\psi_i = \sum_a x^{ia} \phi_a \tag{1}$$

In a non-orthogonal basis, the overlap matrix $\mathbf{S}$ between atomic orbitals $\phi_a$ and $\phi_c$ ($\phi_c$ may be centered on A or on any other atom in the system) is the scalar product

$$S_{ac} = (\phi_a \ , \phi_c) \tag{2}$$

$\mathbf{S}$ allows contravariant (as in eq. 1) coefficients and covariant ones $x_{ic}$ to relate:

$$x_{ic} = \sum_d S_{cd} x^{id} \tag{3}$$

The first-order density matrix $2\mathbf{\Pi}$ for closed-shell systems is actually the representation of a mixed tensor [19]

$$2\Pi_a^b = 2 \sum_i x_{ia} x^{ib} \tag{4}$$

For orthogonal bases, there is no distinction between covariant and contravariant coefficients and $2\mathbf{\Pi}$ becomes a symmetrical matrix. Due to the idempotency of $\mathbf{\Pi}$, a bond index $I_{AB}$ for the bond between atoms A and B may be defined as [19, 20]:

$$I_{AB} = 4 \sum_{\substack{a \in A \\ b \in B}} \Pi_a^b \Pi_b^a \tag{5}$$

This index is the generalization of the Wiberg bond index [21] to non-ortogonal bases. The valence $V_A$ of atom A is in turn [22].

$$V_A = \sum_{B \neq A} I_{AB} \tag{6}$$

Atomic charge $q_A$ may be divided into self-charge and active charge [19, 20, 23]

$$q_A = (I_{AA} + V_A)/2 \quad ; \quad \sum_A q_A = N \tag{7}$$

where N is the number of electrons in the systen. In the last equation, $q_A$ is identical to Mulliken's charge; however, the partition is very different from the classical population analysis.

Let us consider a group G within a molecular system, such as $G = \{A_1, A_2, \cdots A_L\}$. It is straightforward to extend equation (5) in order to introduce a group-group bond index $I_{GG'}$ [10] between groups G and G' in the system considered

$$I_{GG'} = \sum_{\substack{A \in G \\ B \in G'}} I_{AB} \tag{8}$$

In these formulae, $A$ and $B$ refer to any pair of atoms in the systen, independently from the formal linkage between them. The group valence $V_G$ is [24]

$$V_G = \sum_{\substack{A \in G \\ B \notin G}} I_{AB} \tag{9}$$

That is, the sum of the bond indices between atoms respectively inside and outside the group.

Since the idempotency of $\mathbf{\Pi}$ holds for any power, i.e.

$$N = 2\ Tr\mathbf{\Pi} = 2\ Tr(\mathbf{\Pi})^L \tag{10}$$

this allows us to define a multicenter bond index [9, 25]

$$I_{ABC\cdots L} = 2^L \sum_{\substack{a \in A \\ b \in B, \cdots, l \in L}} \Pi_a^b \Pi_b^c \cdots \Pi_l^a \tag{11}$$

For $L = 3$, the $I_{ABC}$ index has been shown to be particularly suitable as a measure of hydrogen bonds. Besides highlighting the distinction between strong and normal hydrogen bonds, it predicts that the peptide bond is of the same order of magnitude as strong hydrogen bonds [9].

For the group-group indices and 3c-bond ones in Table 3, we used the PM3 approximation, leading to more expressive values than AM1. The behaviour of the $I_{GG'}$s in the Table is in agreement with the previous discussion and with chemical expectation: $I_{AU} \sim 3I_{AY}$ corresponds qualitatively to the existence of two normal hydrogen bonds in A•U. The A−Y hydrogen bond is appreciably lower; it is a weaker non-linear bond, its angle being $126^0$ and the H···O distance 2.1Å. It would be an example of unusually *bent* hydrogen bond NH···O = C (see Fig. 3b), where the angle is less than $140^0$ and the H···O distance grater than 1.9Å [26]. Consistently, $I_{PHQ}$ for AY is less than the indices in A•U. It may instead be questioned whether $I_{NHO}$ should be greater or less than $I_{NHN}$. Although some results for the hydrogen bond distances conform to our values, the energies conflict with them [17].

The $I_{PHQ}$ are negative, as we have consistently obtained for hydrogen bonds [9]. Let us remark that $I_{ABC}$ may assume positive or negative values; it can be shown that it corresponds to the correlation betweeen the fluctuations of $q_A$, $q_B$ and $q_C$ from their average values. The three fluctuations are not likely to be in the same sense, if two are in one sense and one in the opposite sense, either positive or negative values are expected to be found, with no "a priori" distinction [27].

In eq. (8), $G$ and $G'$ may be two molecules, as in the isolated A•U pair. In this case, $I_{AU}$ defines an index between them and we have also

$$I_{GG'} = V_G = V_{G'} \tag{12}$$

For the isolated A•U pair, $V_A = V_U = 0.1387$. From the more general definition (9), if in the complex each molecule is considered to be a group:

$$V_Y = I_{AY} + I_{UY} = 0.0540 \ ; \ V_A = I_{AU} + I_{AY} = 0.1812 \ ; \ V_U = I_{AU} + I_{UY} = 0.1472$$

The molecular valences in the complex are in agreement with the other magnitudes obtained. Thus, valences of $A$ and $U$ are both enhanced in the complex, $V_A$ more than $V_U$.

Let us now consider the carbonyl group in $Y$ joining the the NH···O bond in A−Y. For the isolated $Y$ base, $I_{CO}$ is 1.7066; in the complex, it decreases to 1.6727. Now, we have

$$I_{OH} = 0.0182 \ \ ; \ \ I_{ON} = 0.0094, \text{ so that}$$
$$I_{OH} + I_{ON} = 0.0276 \text{ compared with the decrement of } 0.0339 \text{ in } I_{CO} \ .$$

In eq. (7), $q_A$ may be written as

$$q_A = (I_{AA} + \sum_{B \neq A} I_{AB})/2 \tag{13}$$

The bond index $I_{AB}$ is then the electronic distribution along the $AB$ bond (half from $q_A$, half from $q_B$) [20, 25]. Now, the decrement in $I_{CO}$ is greater than the charge spent in the hydrogen bond. Thus, under complexation, charge transfer takes place from the $Y$ base towards the $AU$ pair.

## Conclusion

A hydrogen bond found between one of the carboxylic oxygens in the Y side chain and one of the amino group hydrogens in adenine confers to A•U the stabilization energy of G•C and may be important in the recognition code mechanism involving $tRNA^{Phe}$. The calculated group-group bond indices and molecular valences agree with these features.

## Figure captions

**Fig. 1** - Structure and atomic labelling of the hypermodified Y base.

**Fig. 2** - (a) Base Y structure not taking into account the A•U interaction in the codon-anticodon pairing.
(b) Base Y conformational change due to A•U interaction in the codon-anticodon pairing.

**Fig. 3** - (a) Structure of A36 and Y37 before A•U interaction in the codon-anticodon pairing.
(b) Structure of A36 and Y37 after A•U interaction in the codon-anticodon pairing.

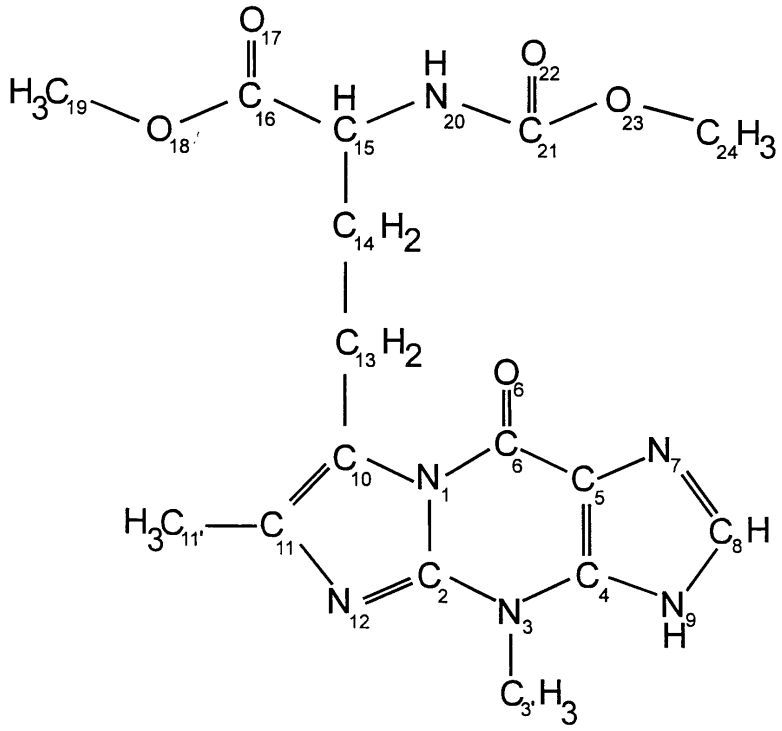
O17
H3C19
O18'
C16
H
C15
N20
H
O22
C21
O23
C24H3
C14H2
C13H2
O6
C10
N1
C6
C5
N7
H3C11'
C11
C8H
N12
C2
N3
C4
N9
H
C3'H3

Fig. 1

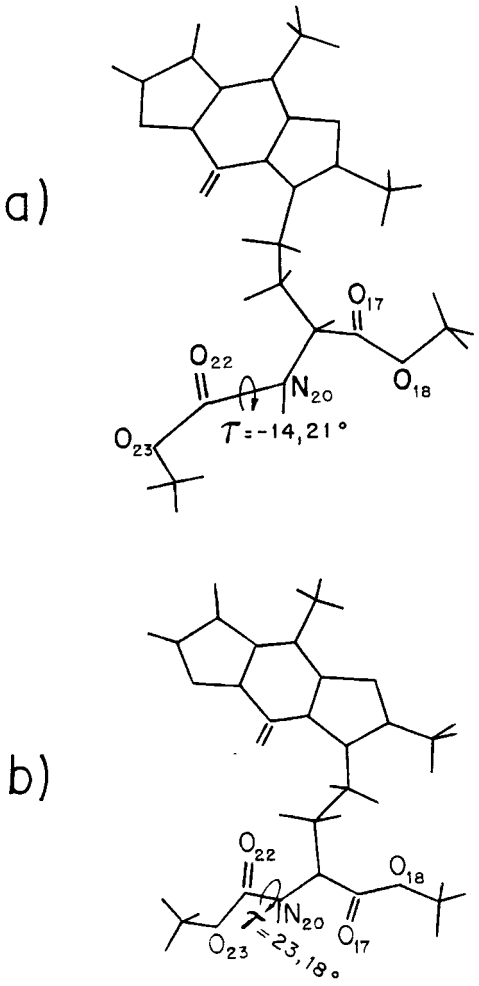


Fig. 2

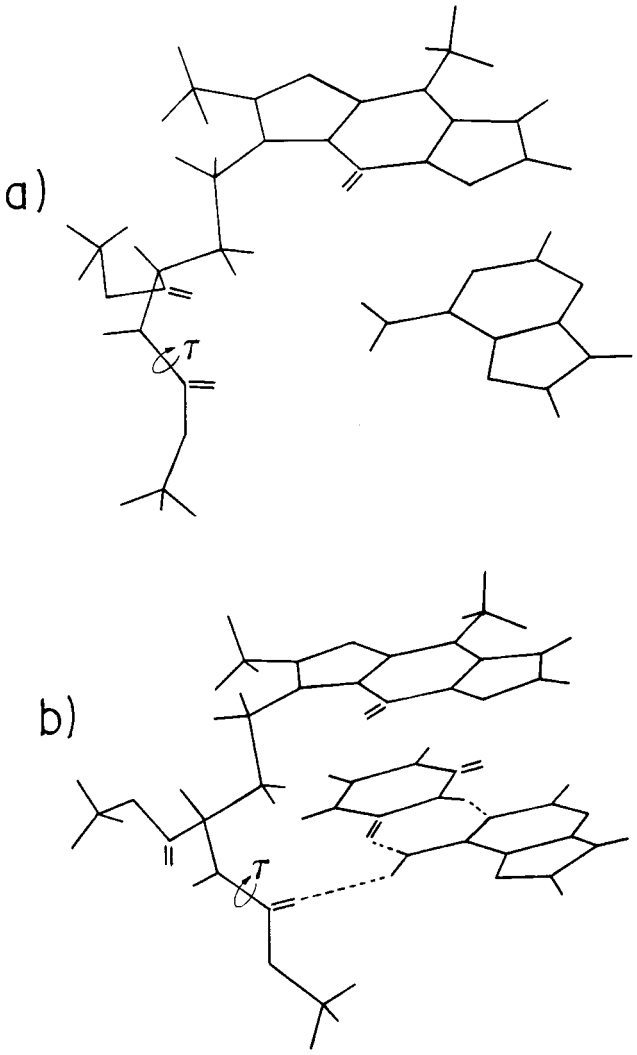


Fig. 3

Table 1: AM1 resulting energies E, in kcal/mole. $E_i$, for isolated systems, $E_c$ in the complex.

Table 1

| | $E_i$ | $E_c$ | $\Delta E$ |
|---|---|---|---|
| A | $-1593.21$ | $-1606.34$ | $-13.13$ |
| U | $-1291.02$ | $-1288.42$ | $+2.60$ |
| Y | $-2403.92$ | $-2398.67$ | $+5.25$ |

Table 2: AM1 resulting energies E, in kcal/mole. $E_i$, for isolated systems, $E_c$ in the complex.

Table 2

| | $E_i$ | $E_c$ | $\Delta E$ |
|---|---|---|---|
| A•U | $-2889.22$ | $-2898.50$ | $-9.28$ |
| A−Y | $-4000.50$ | $-4014.26$ | $-13.76$ |
| U−Y | $-3694.49$ | $-3676.17$ | $+18.32$ |

Table 3: PM3 group-group indices ($I_{GG'}$) and 3c-hydrogen bond indices $I_{PHQ}$.

Table 3

| | $I_{GG'}$ | $I_{PHQ}$ |
|---|---|---|
| A−Y | 0.0440 | −0.0099 (NHO) |
| A•U | 0.1372 | −0.0205 (NHO)<br>−0.0301 (NHN) |
| U−Y | 0.0100 | |

# References

[1] T.H. Jukes, Nature, (1973), 246, 22.

[2] O, Uhlenbeck, J. Baller and P. Doty, Nature, (1970), 255, 508.

[3] J.A. McCloskey and S. Nishimura, Acc. Chem. Res. (1977), 10, 403.

[4] S. Nishimura in Transfer RNA: Structure, Properties and Recognition, Eds. P.R. Shimmel, D. Soll and J.N. Abelson (Cold Spring Harbor Lab. Cold Spring Harbor, NY), 1980, pp. 59-79.

[5] D.W.E. Smith, A.L. MacNamara, J.L. Mushinski and D.L. Hatfield, J. Biol. Chem, (1985), 260, 147.

[6] D.W.E. Smith and D.L. Hatfield, J. Mol. Biol., (1986), 189, 663.

[7] R.K. Wilson and B.A. Roe, Proc. Natl. Acad. Sci. USA, (1989), 86, 409.

[8] A.S. Werneck, M. Oliveira Neto and E.R. Maia, J. Mol. Struct. (Theochem), (1997), 00, 000.

[9] M. Giambiagi, M.S. de Giambiagi and K.C. Mundim, Struct. Chem., (1990), 1, 423.

[10] M. Giambiagi, M.S. de Giambiagi and K.C. Mundim, J. Chem. Soc. Far. Trans., (1992), 88, 2995.

[11] E. Westhof, P. Dumas and D. Moras, Acta Cryst., Sect. A, (1988), 44, 112.

[12] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy and J.J.P. Stewart, J. Am. Chem. Soc., (1985), 107, 3902.

[13] J.J. Dannenberg and L.K. Vinson, J. Phys. Chem., (1988), 92, 5635.

[14] L.K. Vinson and J.J. Dannenberg, J. Am. Chem. Soc., (1989), 111, 2777.

[15] A.C. Pavão, L.A. Soares Neto, J. Ferreira Neto and M.B.C. Leão, J. Mol. Struct. (Theochem), (1995), 337, 57.

[16] M. Khalil, R.J. Woods, D.F. Weaver and V.H. Smith, J. Comp. Chem., (1991), 12, 584.

[17] T.N. Lively, M.W. Jurema and G.C. Shields, Int. J. Quantum Chem., Quantum Biol. Symp., (1994), 21, 95.

[18] M. Oliveira Neto, A.S. Werneck, private communication.

[19] M.S. de Giambiaig, M. Giambiagi and F.E. Jorge, Z. Naturforsh., (1984), 39a, 1259.

[20] M. Giambiagi, M.S. de Giambiagi, D.R. Grempel and C.D. Heymann, J. Chim. Phys., (1975), 72, 15.

[21] K. Wiberg, Tetrahedron, (1968), 24, 1083.

[22] D.R. Armstrong, P.G. Perkins and J.J.P. Stewart, J. Chem. Soc. Dalton Trans., (1973), 838, 2273; N.P. Borisova and S.G. Semenov, Vestn. Leningr. Univ., (1973), 16, 119.

[23] C. Trindle, J. Am. Chem. Soc., (1969), 91, 219.

[24] K.C. Mundim, M. Giambiagi and M.S. de Giambiagi, Nvo. Cim. (1990), 12D, 765.

[25] M.S. de Giambiagi, M. Giambiagi and M. de Souza Fortes, J. Mol. Struct. (Theochem) (1997), 391, 141.

[26] G.A. Jeffrey and W. Saenger, Hydrogen Bonding in Biological Structures, Springer-Verlag, Berlin, 1991, p. 224.

[27] K.C. Mundim, M. Giambiagi and M.S. de Giambiagi, J. Phys. Chem., (1994), 98, 6118.