Theor Chem Acc (2005) 113: 249–256
DOI 10.1007/s00214-005-0633-4

REGULAR ARTICLE

Yaying Zhao · Lixin Zhou

Computational study of hydrogen-bonded complexes between the most stable tautomers of L-leucine and bases of RNA

Received: 6 October 2004 / Accepted: 10 January 2005 / Published Online: 23 May 2005

Abstract A total of 16 hydrogen-bonded complexes between the lowest energy tautomers of L-leucine and each base of RNA have been characterized at the DFT level of theory. The most stable complexes are formed by L-leucine and guanine. Considering backbone of RNA, the affinity order between the L-leucine and bases is guanine> cytosine> adenine≈uracil in perfect accordance with the experimentation. The interplay between the transformed H–X bonds' structural parameters from two monomers to the dimer accompanying with the shifts of the frequency for the H–X stretching mode and interaction energies has been discussed.

1 Introduction

Hydrogen bonds (H-bonds) determine the 3-D structures of biomacromolecules, for that reason, it is one of the key interactions in molecular biology. Of course, binding of proteins to DNA or RNA plays an important role in the regulation and control of gene expression, especially on the problem of the life-origin. Recently, it has been proved that proteins are capable of specific recognition of DNA sequences with extremely high precision (4-8 bp) [1]. H-bonds between peptide bonds or hydrophilic side chains of amino acid and DNA bases are among the most important interactions responsible for the amazing specificity of protein binding, which may be one of the keys to the life-origin. Therefore, basic knowledge concerning the interactions between the building blocks of proteins and DNA or RNA — amino acids and nucleic acid bases, is very interesting. Although these systems are definitely simpler than real biochemical targets, quantitative information regarding the interactions between amino acids and nucleobases can provide insight into biochemical problems.

Y. Zhao · L. Zhou (✉)
Department of Chemistry,
Jinan University,
Guangzhou, Guangdong 510632,
P. R. China
E-mail: tlzhou@jnu.edu.cn

Ponnamperuma [2] made a conclusion that the order of the affinity between the L-leucine and nucleoside monophosphate is G>C>A=U by experimentation. And a few studies have been done to analyze the role of the amide group in the course of untwisting of the DNA double helix [3,4]. On the theoretical side, the computational studies of hydrogen bonds that develop between nucleobases have recently been reviewed, [5] and several studies on the interactions between nucleobases and water molecules have been reported [6–11]. The involvement of a proton acceptor with the smaller proton affinity was an expected finding, which indicated that the strength of a hydrogen bond might be more sensitive to the acidity of a proton donor than to the basicity of proton acceptor [8].

So far, there is a relative paucity of information about interactions between amino acids and nucleobases. The early studies concentrated on the interaction of proteins with nucleobase pairs at the Hartree–Fock (HF) level of theory, [12, 13] it was displayed that whether the external hydrogen bonds stabilize or destabilize the base pairs mainly depends on both the type of interacting residues and the site of the interaction. And the interaction between single- and double-stranded B-DNA helices and polyglycine has been studied at the HF level too, [14] which illustrated that the most stable configuration is represented by realizing their symmetry agreement. For the recent theoretical study, free energies of the interaction between the hydrophilic side chain of asparagines and nucleobase pairs have been calculated by extensive conformational sampling using a molecular force field [15]. The differential affinity of asparagines toward A–T and G–C was shown, the role of both structural flexibility of the side chain and entropic interactions was emphasized as well. More recently, the interaction of hydrogen-bonded complexes between the most stable tautomers of glycine and uracil has been studied both at the semi-empirical level (PM3) and the density functional level of theory (DFT) [16,17]. It was found that the formation of a stable structure with two hydrogen bonds requires not only a favorable two-body interaction but also a favorable topological match of the proton donor and acceptor [18].

Different tautomers of bases are obtained considering different hydrogen positions around the base [19]. However, tautomers are rarely observed in oligonucleotide crystals [20] and for most biochemical processes probably only major tautomers of bases are involved. So it is reasonable to assume that the minor tauomers were eliminated during evolution processes to ensure the stability of the genetic code. This prompted us to consider only the most stable structure of bases in the current study [5].

In order to identify decisive factors responsible for the affinity between L-leucine and bases of RNA, we make our current computational effort to report the results of electronic structure calculations concerning the L-leucine&nucleobase complexes formed by the most stable tautomers of L-leucine and bases, i.e., the dimer of L-leucine and guanine, cytosine, adenine, uracil.

2 Computational method

As other reports on complexes between nucleic acid bases and water, [21,22] pyridine and water, [23] and glycine and uracil, [18] which have demonstrated the usefulness of the approach in studying systems with intermolecular hydrogen bonds, we adopted the same approach in our work. We applied primarily PM3 [24] and then the DFT method with a hybrid B3LYP functional [15–17] and 6-311++G** basis set [25,26] to study structure and stability of the L-leu&base complexes. In addition to those familiar hydrogen bonds involving two highly electronegative atoms (N or O) on the part of bases, we explored complexes with the CH group of bases acting as a proton donor too.

The stability of L-leu&base complexes is measured in terms of E_{stab}, H_{stab}, and G_{stab}. E_{sta} is defined as a difference in electronic energies of the monomers and the dimer with the electronic

$$E_{\text{stab}} = E^{B}(\text{Geom}^{B}) + E^{\text{L-leu}}(\text{Geom}^{\text{L-leu}}) - E^{\text{l-leu\&}B}(\text{Geom}^{\text{L-leu\&}B}), \quad (1)$$

energy E^{x}(X = L-leu, Bases (A, U, C, G), or L-leu&Bases) computed for the coordinates determining the optimal geometry of X (i.e., the geometry where E^{x} is at the minimum). E_{stab} can be decomposed as [27]

$$E_{\text{stab}} = E^{B}_{\text{dist}} + E^{\text{L-leu}}_{\text{dist}} + E^{\text{L-leu\&}B}_{\text{int}}, \quad (2)$$

Where E^{x}_{dist} is a repulsive one-body component related to a distortion of the monomer X (X = L-leu or Bases) in the dimer

$$E^{X}_{\text{dist}} = E^{X}(\text{Geom}^{X}) - E^{X}(\text{Geom}^{\text{L-leu\&}B}), \quad (3)$$

and $E^{\text{L-leu\&}B}_{\text{int}}$ is a two-body interaction energy between the distorted monomers [28]

$$E^{\text{L-leu\&}B}_{\text{int}} = E^{B}(\text{Geom}^{\text{L-leu\&}B}) + E^{\text{leu}}(\text{Geom}^{\text{L-leu\&}B}) - E^{\text{L-leu\&}B}(\text{Geom}^{\text{L-leu\&}B}). \quad (4)$$

The $E^{\text{L-leu\&}B}_{\text{int}}$ component was corrected for basis set superposition (BSSE) by adopting the counterpoise method of Boys and Bernardi [29,30]. In this method, the energy of each monomer is evaluated in the basis set of the dimer. While the values of the E^{X}_{dist} terms were calculated with monomer centered basis sets [31]. As for the stabilization enthalpy H_{stab}, it results from correcting E_{stab} for zero-point vibration terms; thermal contributions to energy from vibrations, rotations, and translations; and the pV terms. Finally, the stabilization Gibbs free energy G_{stab} was the consequence of supplementing H_{stab} with the entropy term. The values of H_{stab} and G_{stab} discussed in Sect. 3 were obtained for T = 298 K and p = 1 atm.

All calculations were carried out with the Gaussian 98 code [32].

3 Results and discussion

3.1 Selection of hydrogen-bond

Both L-leucine and bases of RNA belong to the class of molecules having several proton donor and acceptor centers capable of forming hydrogen bonds of various strengths. These are O7, O4 and N10 for L-leucine, N**n**, O**n**, and C**n** for bases (different **n** from base to base); see Fig. 1. Here, we mainly focused on complexes with two intermolecular hydrogen bonds, as it is more difficult for L-leu to overcome a topological mismatch than glycine, who's distortion of energy approaches 40 kJ/mol [18]. And in the same paper, we found that except the form of O7 acceptor&O10H donor of amino acid interacting with bases, the others forms have been proved to be unstable in terms of energy, or unstable in terms of free energy. Considering our main goal, in this work, we neglect those unstable forms and mainly deal with the form of O7 acceptor&O4H donor interacting with bases.

3.2 Relative stability of the L-leu&bases complexes

The B3LYP/6-311++G** values of H_{stab} and E_{stab} for those hydrogen-bond complexes are plotted in Fig. 3. The almost parallelism between E_{stab} and H_{stab} indicates that the contributions to H_{stab} arising from rotations and vibrations are of similar magnitude for each kind of complex. The values of E_{int}, E_{stab}, H_{stab}, and G_{stab} collected, while E_{int} and E_{stab} were corrected for BSSE using the counterpoise procedure of Boys and Bernardi [29]. The values of BSSE were found to be much small at the B3LYP/6-311++G** level as the counterpoise estimates are in a range from −2.3 kJ/mol to −3.5 kJ/mol.

The most stable families of complexes are L-leu&Gn (see Fig. 3) with the carbonyl (O7) and hydroxyl (O10H) groups of L-leucine interacting with the proton donor and acceptor centers of guanine (see Fig. 2). Based on the data shown in Table 1 and the plot shown in Fig. 3, it is clear that the L-leu&G2 structure is the most stable, followed by L-leu&G1 and L-leu&G3, which was obviously consistent with

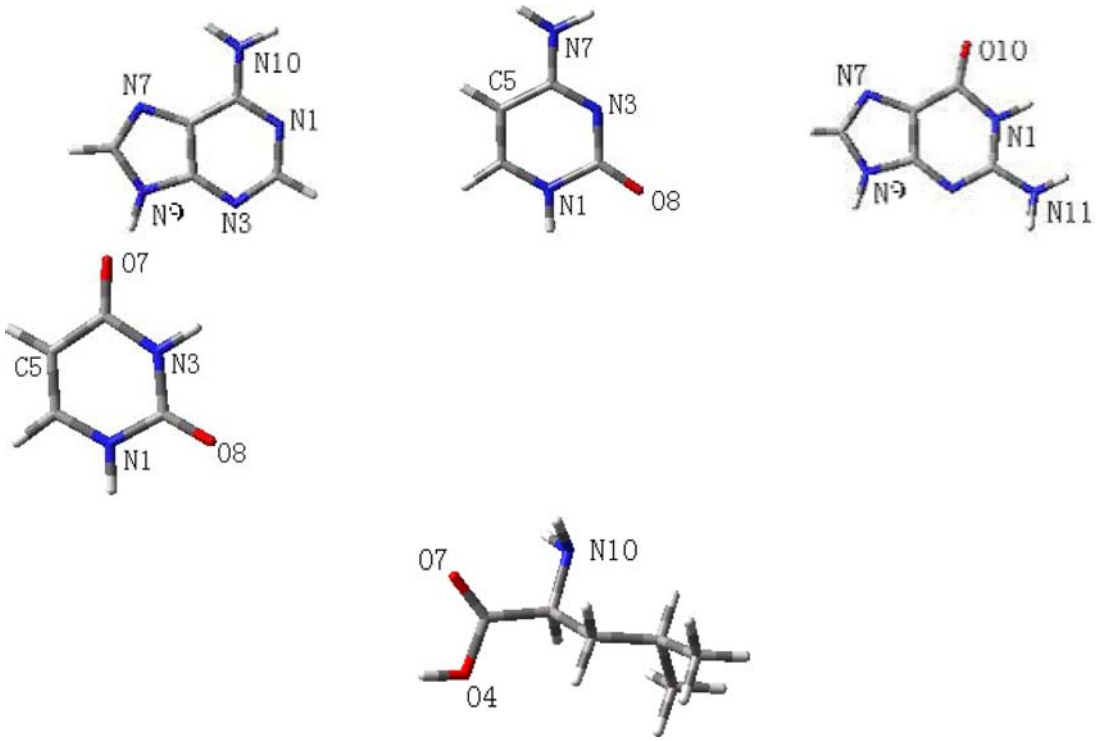

Fig. 1 Lowest energy tautomers and conformers of L-leucine and bases of RNA

the order of their structural parameters (see Table 2). These three structures have two strong hydrogen bonds, and the values of E_{stab} for L-leu&Gn span a range of 71.1–53.7 kJ/mol (except the structure of L-leu&G4), which provides ca. 35.6–26.9 kJ/mol per hydrogen bond. These stabilization energies are typical for dimers forming ring-like structures, such as the formic acid dimer (E_{stab} = 63.5 kJ/mol [33]) or the formamide dimer (E_{stab} = 60.2 kJ/mol [34]). The values of G_{stab} are positive for these three structures indicating a themodynamic preference to form the L-leu&Gn dimer. Herein, the L-leu&G4 structure is the most unstable, whose G_{stab} is negative, because of the proton donor of guanine – C8H, and it is obvious that the C8H$\cdots$O7 hydrogen bond is much weaker than other hydrogen bond with the largest interatomic distance of H-bond, 2.348 Å and the smallest valence angle, 124.8°.

As far as the next stable families — L-leu&Cn concerned, the proton acceptor and donor sites of cytosine are N3&N7H and O8&N1H (see Fig. 2), correspondingly display a different stability (see Fig. 3). The L-leu&C2 structure is the most stable in this family; it has an almost similar value of E_{stab} and G_{stab} with L-leu&G1. One thing that needs to be pointed out is that the corresponding order to these two values is just in reverse – E_{stab} (L-leu&G2) > E_{stab}(L-leu&C2); G_{stab} (L-leu&G2) < G_{stab}(L-leu&C2), as the value of entropy for the former is larger than that for the latter. Whereas, the L-leu&C1 structure has two hydrogen bonds too, which provide ca. 31.5 kJ/mol per hydrogen bond and is weaker than the hydrogen bond of L-leu&C2 structure that corresponds to their geometrical parameters of H-bonds too.

The third (L-leu&An) families, in which the proton acceptor and donor sites of adenine are N7&N10H$_2$, N7&C6H, N3&N9H, N3&C2H, N1&C2H, and N1&N10H$_2$, respectively (see Fig. 2), exhibit a different stability. Obviously the L-leu&A4 structure possess the almost similar values of E_{stab} with the L-leu&C1, and it provides the unique positive G_{stab} in this family, and its value of H_{stab} is much larger than any other structure of L-leu&An as well. As to the L-leu&A1 and L-leu&A6 structure, they have a much similar values of E_{stab} and H_{stab}, whereas the value of G_{stab} for the latter is larger than the former, which might be attributed to the chelation of the ring of adenine. As regards to the weak CnH$\cdots$O7 hydrogen bond involved in the L-leu&A2, L-leu&A3, L-leu&A5 structure, it is a much weaker H-bond too, which is even longer than the CnH$\cdots$O7 of the L-leu&G4, as a result, its contribution to the stability of the complexes is much smaller. Although the L-leu&A5 provides the most stable CnH$\cdots$O7 hydrogen bond with the shortest distance in this family, 2.418 Å (see Table 2), and whose the value of E_{stab} is 40.2 kJ/mol, its value of G_{stab} is negative yet.

Finally, the L-leu&Un families, with the N1H&O8, C5H&O7, N3H&O7, and N3H&O8 of uracil acting as a proton donor and acceptor, severally (see Fig. 2), is even more weakly bound with the values of H_{stab} in a range of 55.5-31.8 kJ/mol. Except for the L-leu&U3, which has the largest values of E_{stab} of 61.2 kJ/mol, all the other structures in this family are characterized by negative values of G_{stab}. Herein, the order of the complexes' stability, which is L-leu&U3 > L-leu&U1> L-leu&U2> L-leu&U4, within this family is similar with the order which was reported for a water molecule

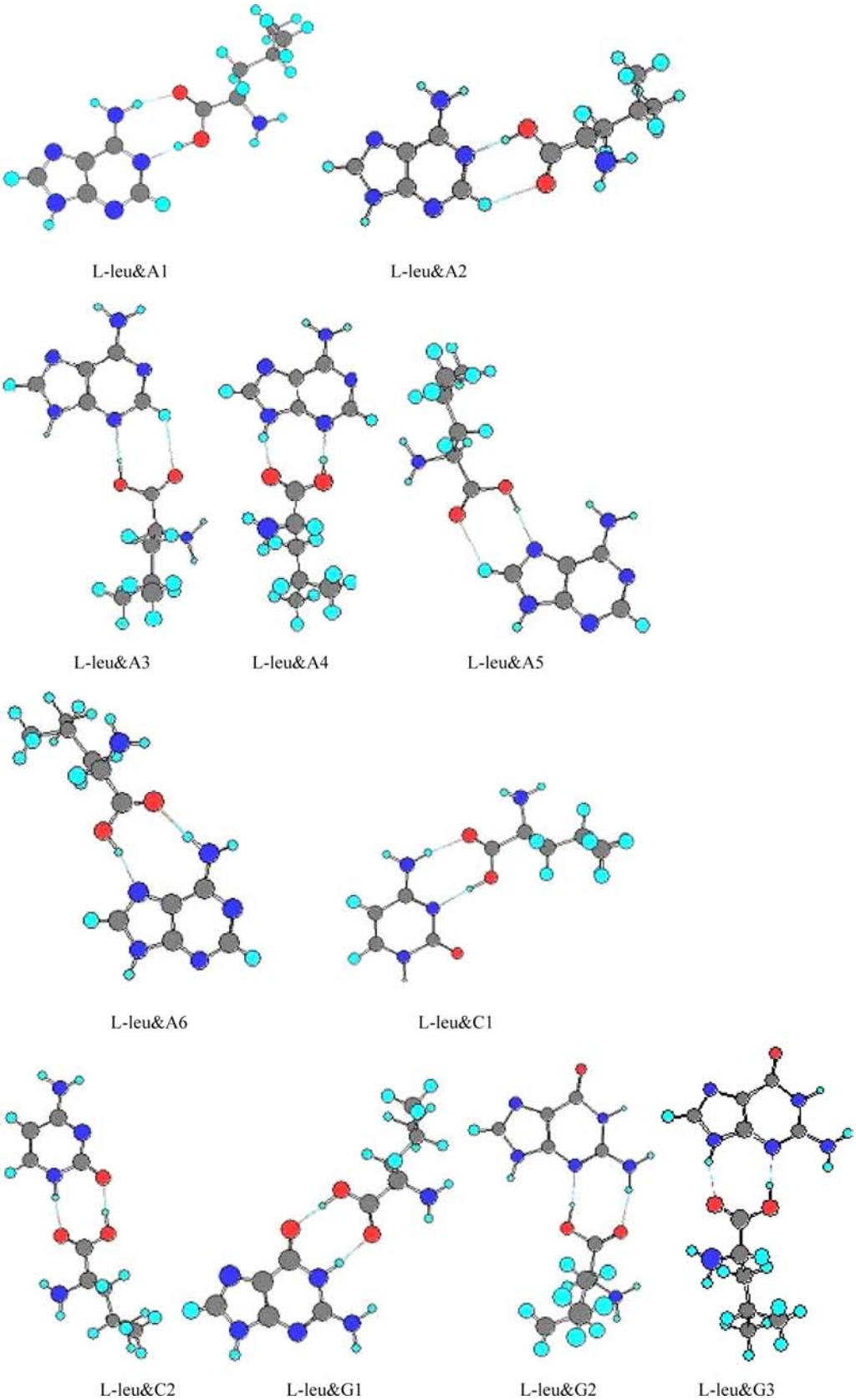

Fig. 2 B3LYP/6-311++G** optimized structures of dimers

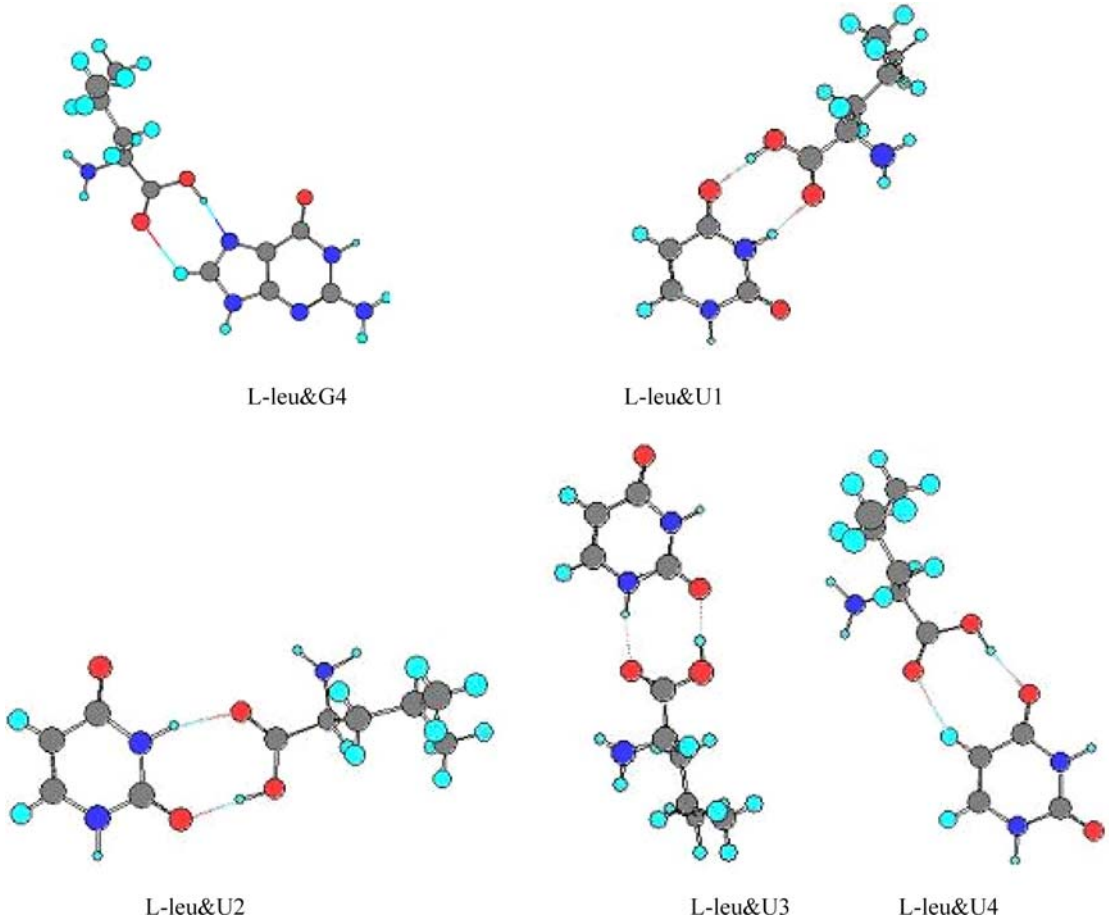

Fig. 2 (Contd.)

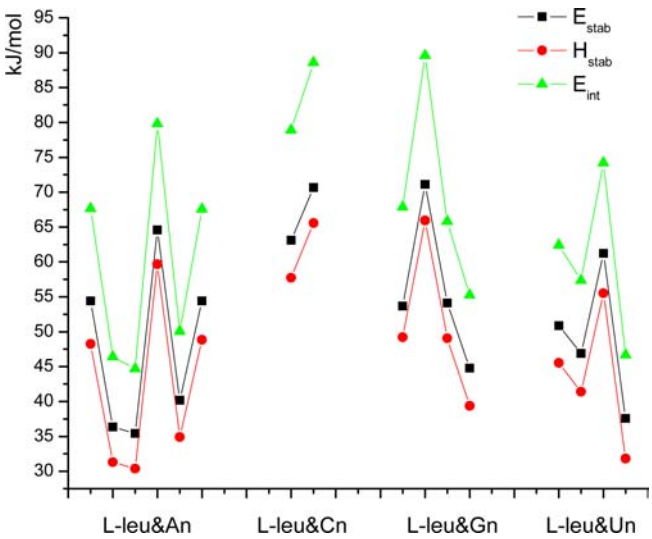

Fig. 3 Energies and enthalpies of complexes obtained at the B3LYP/6-311++G** levels of theory

interacting with uracil, [6–11] and a glycine interacting with uracil as well [16]. Moreover, it is clearly that the order to their distances of H-bonds is the just reverse (see Table 2). One thing that needs to be pointed out is that not only the values of E_{stab} but also those of H_{stab} and G_{stab} of the L-leu&Un complexes are smaller than those for the gly&Un (see Table 1 and Fig. 1 of supporting information), whereas, the values of E_{int} are almost equal, and so do their distances of H-bonds (see Table 2). This indicates that the backbone of the amino acid might influence the stability of the complexes greatly, especially, in terms of G_{stab} for the reason of entropy, but their geometrical parameters might be maintained or changed slightly.

It seems that the most stable structure within the L-leu&Gn, L-leu&Cn, L-leu&Un families is always those structure with N**n**H and O**n** of bases acting as the proton donor and acceptor. In RNA, however, this region is not operational as the base of cytosine and uracil is covalently attached to a sugar through the N1 atom, then in these families, the next most stable structures are the most important for hydrogen bond formation by bases bonded to the sugar-phosphate RNA backbone. As for the L-leu&An families, whose most stable structure is the structure with N9H and N3 site acting as the proton donor and acceptor, for the same reason that the N9 is the attached site between the bases and the sugar, then in fact the next most stable structure becomes the most important. Besides, usually the unstable structures are those complexes including one hydrogen bond formed by the C**n**H site of bases and O7 proton acceptor of L-leu.

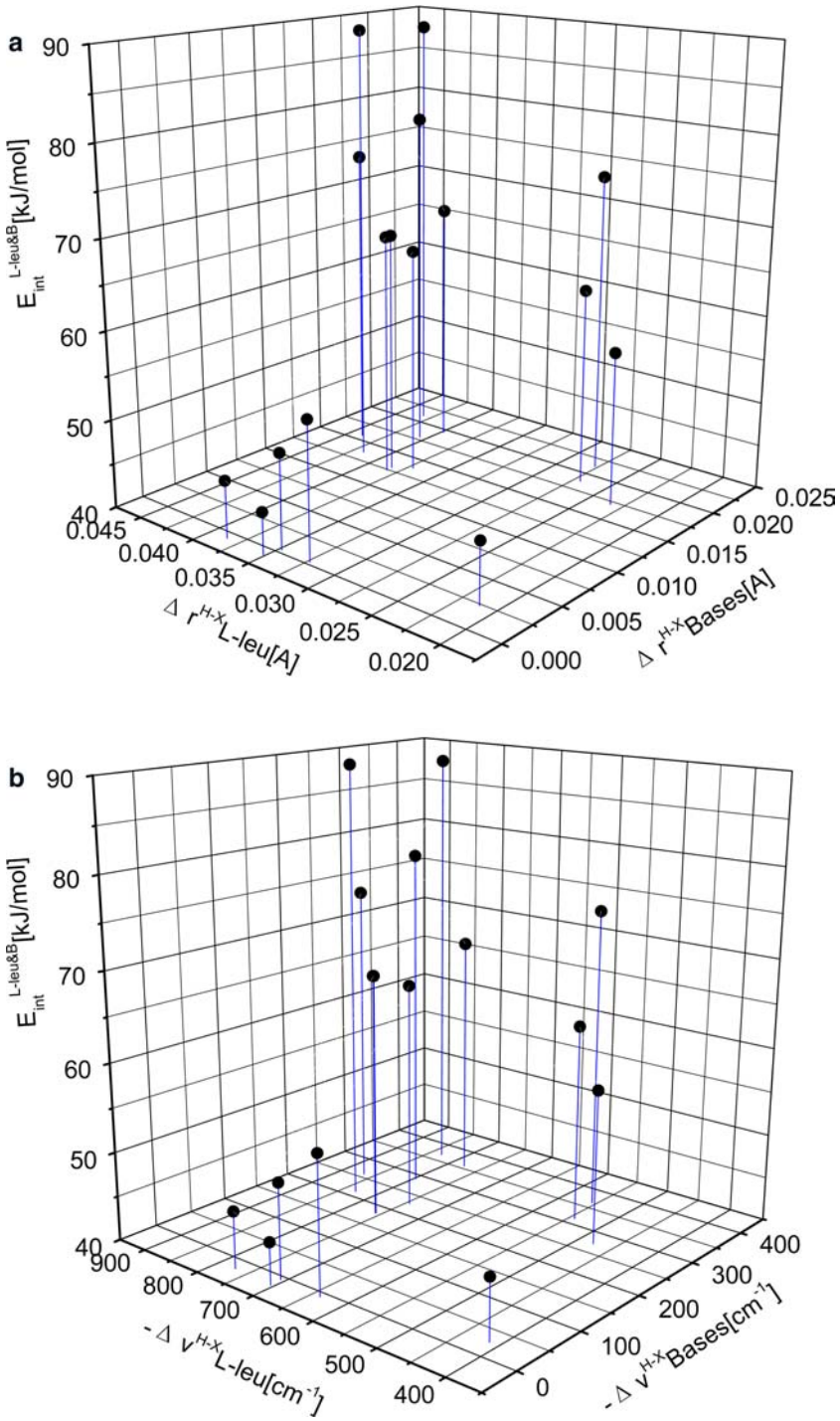

Fig. 4 $E_{int}^{L-leu\&B}$ as a function of elongation of the proton donor H-X bonds (a) and vibrational red shifts for the stretching H-X modes (b)

To summarize, the second stable structure contributes greatly to the interaction of L-leu and RNA backbone with the hydrogen bond. Considering this reason, the order of the complexes stability is L-leu&Gn> L-leu&Cn> L-leu&An≈ L-leu&Un in the values of G_{stab}(see Fig. 2 of supporting information), so the selectivity between the bases of RNA and L-leucine is G>C>A≈U.

3.3 Geometries and selected vibrational frequencies

From the optimized structural parameters of intermolecular hydrogen bonds at different complexes, it can be clearly seen that the strength of a H-bond is determined by the distance between H and Y, and the Y···HX angle, obviously, which will directly influence the monomer distortion terms

E_{dist}^{X} – it quantify strains acquired by the monomers when formed a dimer. It is reasonable that every most-stable dimer of each family has the shortest H-bonds and the most linear, as the favorable geometries provide a larger value of $E_{\text{int}}^{\text{L-leu\&}B}$, although sometimes accompanied by the relative bigger monomer distortion terms.

Logically, formation of a hydrogen bond Y$\cdots$HX is always accompanied by an elongation of the H–X bond, $\Delta r^{\text{H-X}}$, and a red shift of the frequency for the H–X stretching mode, $\Delta\nu^{\text{H-X}}$. The values of $\Delta r^{\text{H-X}}$ are mostly around 0.040 Å for the OH bond of L-leu and 0.018 Å for the NH bond of bases. The values of $\Delta\nu^{\text{H-X}}$ reach the maximum $-866\,\text{cm}^{-1}$ for the OH bond of L-leu and $-370\,\text{cm}^{-1}$ for the NH bond in bases. However, the values of $\Delta r^{\text{H-X}}$ are almost equal to 0.000 Å for the CH bond of bases, and even accompanied with a positive value of $\Delta\nu^{\text{H-X}}$, for instance, the L-leu&A2, the L-leu&A3, and the L-leu&A5. This indicates that the CH is not a favored proton donor site of bases further.

Both 3-D plots from Fig. 4 show that the largest values of $E_{\text{int}}^{\text{L-leu\&}B}$ for each family arise only when both hydrogen bonds involve significant elongation accompanied by notable vibrational red shifts. In contrast, the small values of $E_{\text{int}}^{\text{L-leu\&}B}$ for those structures are corresponding with small perturbation of H-X bonds of either L-leu, or bases of RNA, or both of them. One thing should be noted that the values of $E_{\text{int}}^{\text{L-leu\&}B}$ might be more sensitive to the elongation of the H–X bonds of the bases than to that of L-leu, and so do vibrational red shifts for the stretching H–X modes, since the proton donors are different from base to base, and even in the same base, especially the unfavorable proton donor of bases –CH.

By comparing the similarity between parts a and b of Fig. 4, we can see that there is a strong correlation between the values of $\Delta r^{\text{H-X}}$ and $\Delta\nu^{\text{H-X}}$, which has been shown in Fig. 5 parts a and b, for L-leu and bases of RNA, respectively. In the parts a, the plot $\Delta\nu^{\text{H-X}}$ versus $\Delta r^{\text{H-X}}$ is observed with the parabolic and linear fits for L-leu with a square correlation coefficient r^2 of 0.995, and 0.993, as to bases of RNA, providing 0.996, and 0.991, respectively, which is consistent with the previous study. [9,18,35].

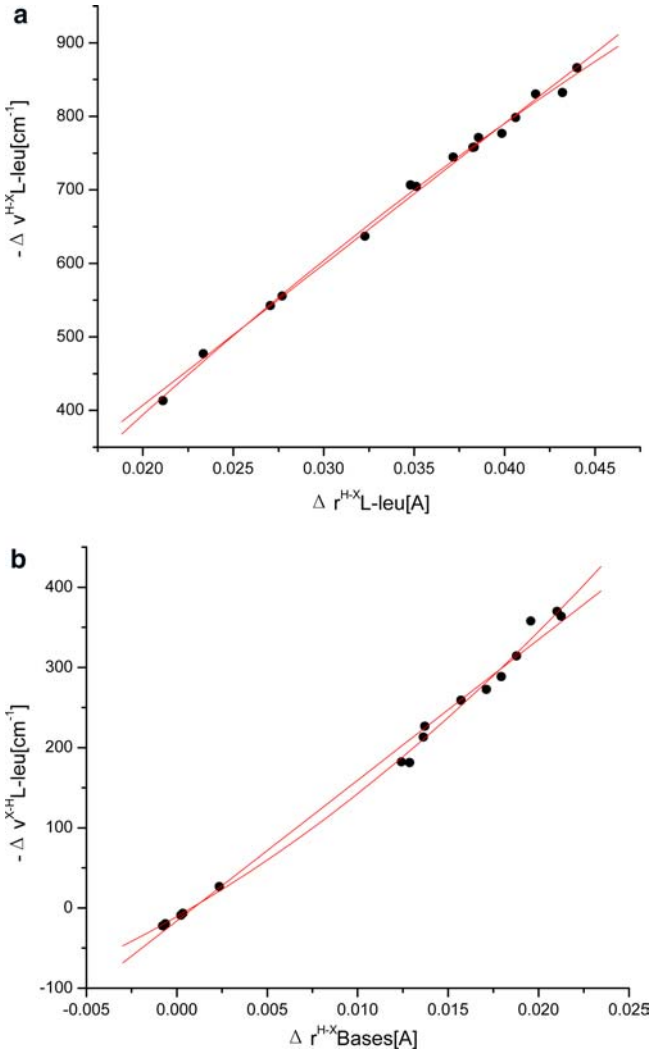

Fig. 5 Correlation between elongations of the proton donor H-X bonds and vibrational red shifts for the H-X stretching modes of L-leu (a) and bases of RNA (b)

4 Summary

We demonstrated that the most stable complexes between L-leucine and bases of RNA are formed when the carboxylic group of L-leu is bound through two hydrogen bonds to different bases. The largest stabilization energy of 71.1 kJ/mol was determined at the B3LYP/6-311++G** level for the L-leu&G2 structure. The largest themodynamic stabilization free energy of 17.9 kJ/mol was determined at the same level for the L-leu&C2 structure; however, this structure involves the N1 atom of cytosine, which in RNA is covalently bonded to the sugar-phosphate backbone, then the L-leu&A4, L-leu&U3 structure should be ignored as well for the same reason. To sum up, the conclusion of our calculation might in fair agreement with experimental result — the order of the affinity of L-leu and bases of RNA is G>C>A≈U, which might be meaningful for the research on the life-origin. On the other hand, the stability of complexes might be influenced by the backbone of amino acid greatly as well for the reason of entropy, although their structural parameters of H-bonds might not be changed markedly, and the chelation of the base ring might be a special representation. In addition to, the significant elongation of the proton donor bond (H–X) accompanied by notable vibrational red shifts of the frequency for the H–X stretching mode offer the largest values of $E_{\text{int}}^{\text{L-leu\&}B}$ for the dimers of each family, and there are a better linear correlation between the elongation of H–X bond and its vibrational red shift.

Finally, we note that CnH is not a good proton donor sites of bases, for all of those complexes involved CnH$\cdots$O7 have an negative G_{stab}, and possess a much small red shifts of the frequency for the H–C stretching mode, or even a blue shifts sometimes.

References

1. Desfrancois C, Carles S, Schermann JP (2000) Chem Rev 100:3943
2. Senaratne N, Hobbish MK, Ponnamperuma C (1990) Direct in between amino acids and nucleotides as a possible explanation for the oregin of the genetic code. In: Ponamperuma C, Einch F (eds) Prebiological Self Organization of Matter A. Deepak Publication, Hampton Virginia, pp 148
3. Galetich I, Kosevich M, Shelkovsky V, Stepanian SG, Blagoi YP, Adamowicz L (1999) J Mol Struct 478:155
4. Smolyaninova TI, Bruskov VI, Kashparova Ye V (1985) Mol Biol (Russian) 19:992
5. Hobza P, Sponer J (1999) Chem Rev 99:3247
6. Rybak S, Szalewicz K, Jeziorski B, Corongiu G (1992) Chem Phys Lett 199:567
7. Smets J, McCarthy WJ, Adamowicz L (1996) J Phys Chem 100:14655
8. Chandra AK, Nguyen MT, Zeegers-Huyskens T (1998) J Phys Chem A 102:6010
9. van Mourik T, Price SL, Clary DC (1999) J Phys Chem A 103:1611
10. Dolgounitcheva O, Zakrzewski VG, Ortiz JV (1999) J Phys Chem A 103:7912
11. van Mourik T, Benoit DM, Price SL, Clary DC (2000) Phys Chem Chem Phys 2:1281
12. Sarai A, Saito M (1984) Int J Quantum Chem 25:527
13. Sarai A, Saito M (1985) Int J Quantum Chem 28:399
14. Otto P, Clementi E, Ladik J, Martino F (1984) J Chem Phys 80:5294
15. Pichierri F, Aida A, Gromiha MM, Sarai A (1999) J Am Chem Soc 121:6152
16. Becke AD (1993) J Chem Phys 98:5648
17. Lee C, Yang W, Paar RG (1988) Phys Rev B 37:785
18. Dabkowska I, Rak J, Gutowski M (2002) J Phys Chem A 106:7423
19. Kwiatkowski JS, Pullman B (1975) Adv Heterocycl Chem 18:199
20. Schuerman GS, van Meervelt L, Loakes D, Brown DM, Lin PKT, Moore MH, Salisbury S (1998) J Mol Biol 282:1005
21. Kryachko ES, Nguyen MT, Zeegers-Huyskens T (2001) J Phys Chem A 105:1934
22. Chandra AK, Nguyen MT, Uchimaru T, Zeegers-Huyskens T (1999) J Phys Chem A 103:8853
23. Dkhissi A, Adamowicz L, Maes G (2000) J Phys Chem A 104:2112
24. Stewart JJP (1989) J Compt Chem 10:221
25. McLean AD, Chandler GS (1980) J Chem Phys 72:5639
26. Krishnan R, Binkley JS, Seeger R, Pople JA (1980) J Chem Phys 72:650
27. Kendall RA, Simons J, Gutowski M, Chalasiński G (1989) J Phys Chem 93:621
28. Szalewicz K, Jeziorski B (1998) J Chem Phys 109:1198
29. Boys SF, Bernardi F (1970) Mol Phys 19:553
30. Gutowski M, Chalasiński G (1993) J Chem Phys 98:5540
31. Chalasiński G, Gutowski M (1988) Chem Rev 88:943
32. Frisch MJ, Trucks GW, Schlegel HB, et al (1998) GAUSSIAN 98, Revision A 6, Gaussian, Pittsburgh
33. Pecul M, Leszeczynski J, Sadlej J (2000) J Chem Phys 112:7930
34. Vargas R, Garza J, Friesner RA, Stern H, Hay BP, Dixon DA (2001) J Phys Chem A 105:4963
35. Kryachko ES, Zeegers-Huyskens T (2001) J Phys Chem A 105:7118