ARTICLE

Gas Phase Conformations of Tetrapeptide Glycine-Phenylalanine-Glycine-Glycine

Hui-bin Chen, Yao Wang, Xin Chen, Zi-jing Lin*

Hefei National Laboratory for Physical Sciences at Microscale and Department of Physics, University of Science and Technology of China, Hefei 230026, China

(Dated: Received on November 24, 2011; Accepted on November 25, 2011)

Systematic search of the potential energy surface of tetrapeptide glycine-phenylalanine-glycine-glycine (GFGG) in gas phase is conducted by a combination of PM3, HF and BHandHLYP methods. The conformational search method is described in detail. The relative electronic energies, zero point vibrational energies, dipole moments, rotational constants, vertical ionization energies and the temperature dependent conformational distributions for a number of important conformers are obtained. The structural characteristics of these conformers are analyzed and it is found that the entropic effect is a dominating factor in determining the relative stabilities of the conformers. The measurements of dipole moments and some characteristic IR mode are shown to be effective approaches to verify the theoretical prediction. The structures of the low energy GFGG conformers are also analyzed in their connection with the secondary structures of proteins. Similarity between the local structures of low energy GFGG conformers and the α-helix is discussed and many β- and γ-turn local structures in GFGG conformers are found.

Key words: Potential energy surface, Conformational stability, Hydrogen bond, Entropy effect, Secondary structure

I. INTRODUCTION

Amino acids and peptides are extremely important biological molecules as they are not only the building blocks of proteins, but also directly involved in the activities of life. As the functions of biological molecules are intimately dependent on the conformations they may adopt, the theoretical conformational studies on these molecular systems are of high scientific significance. On one hand, the theoretical results may help to explain the experimental measurements carried out on these systems, such as dipole moments [1], rotational constants [2], IR [3] and UV spectra [4], ionization potentials [5], two-photon circular dichroism [6], *etc.* On the other hand, the computational results may provide insight information about the molecules that are difficult to probe by the experiment, but are critically important for the understanding of the structural basis of the molecular functions, *e.g.*, the atomic scale resolution of the biomolecular structures and the specific binding nature of a hydrogen bond. Moreover, the conformations of small peptides are critically important for the understanding of the protein structures [7, 8].

L-phenylalanine is one of the twenty amino acids that are the building blocks of all proteins of any living species. Phenylalanine is also one of the eight essential amino acids for human. It is used in the manufacture of food and drink products and sold as a nutritional supplement for its reputed analgesic and antidepressant effects [9−11]. A rare metabolic disorder called phenylketonuria (PKU) may occur in people for whom an enzyme that the body needs to use phenylalanine is missing [12]. Peptides containing phenylalanine also play an important medical role. For example, tetrapeptide glycine-phenylalanine-glycine-glycine (GFGG) has been widely used in the synthesis of anti-cancer drugs [13, 14]. Due to their fundamental significance in biology and important role in medicine, phenylalanine and its peptide complexes have received a lot of attention, both experimentally and theoretically [15−20]. However, hydrogen bonds are abundant in biomolecules and their accurate treatment requires high level quantum chemistry (QC) methods. Due to the complexity of the problem, the existing QC studies are limited to F [15] and its small peptides FF [18], FFG [19], and GGF [20].

In this work, we report an extensive computational QC search of the gas phase conformations of the tetrapeptide GFGG. The goal of this study is to locate all low energy gas phase GFGG conformers with full geometry optimizations, to obtain precise knowledge about the relative stabilities of different conformers on the energy surfaces, and to provide theoretical results such as rotational constants, vibrational frequencies, dipole moments of conformers and conformer dis-

*Author to whom correspondence should be addressed. E-mail: zjlin@ustc.edu.cn, Tel.:+86-551-3606345, FAX: +86-551-3606348

tributions at various temperatures that may be helpful to future experimentalists.

II. COMPUTATIONAL METHOD

The properties of a biomolecule are often determined by a few stable conformations that the molecule may adopt in practical conditions. However, the task of finding the low energy conformations is far from trivial [21, 22]. In order to reliably locate the stable conformations, a thorough search of the potential energy surface (PES) may be required [23]. The thorough PES search is carried out by optimizing all trial structures generated by allowing for combinations of all rotational degrees of internal bonds. The unique structures obtained by the geometry optimizations are subjected to the frequency calculations and single point energy calculations with a high level QC method in order to identify the true important conformations. The hydrogen bonding features and other structural characteristics of the important conformers are further analyzed. The detailed computational procedures are described below.

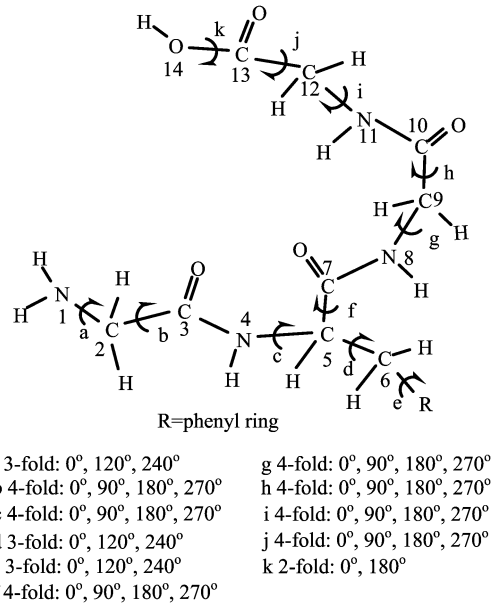

FIG. 1 Illustration of the 11 bond rotational degrees of freedom and the number of rotations for each degree of freedom in the generation of trial structures of GFGG.

A. Generation of trial structures

The planar structures of GFGG are shown in Fig.1. In principle, the full conformational space of gaseous GFGG may be explored through a systematic variation of all rotational degrees of freedom. As shown in Fig.1, there are a total of 14 bond rotational degrees of freedom in GFGG. As a dihedral angle can change from 0° to 360°, typically increments of 60°−90° for asymmetrical dihedral angles and 120° for symmetrical dihedral angles are required to ensure a complete scan of the potential energy surface [24−26]. However, as the *cis*-peptide bonds are energetically unfavorable and rarely found in the experiments, only the *trans*-peptide bonds need to be considered. That is, the three peptide bonds in GFGG are fixed at the *trans* configuration. Consequently, a total of 11 rotational bonds are considered in the trial structure generation process. Moreover, it suffices to consider *syn*- or *anti*- periplanar arrangements corresponding to 0° and 180° torsions for the C−OH groups [15, 25]. Overall, the number of bond rotations for each dihedral angle is as indicated in Fig.1, resulting in a total of 884736 (=3×4×4×3×3×4×4×4×4×4×2) possible trial structures.

B. Process of geometry optimizations

All trial structures were first optimized by the semi-empirical method of PM3. A total of 37859 unique structures were thus obtained, representing approximately the possible number of local minima in the potential energy surface of GFGG at the PM3 level of theory. As only a limited number of local minima around the global minimum of the free energy surface are of high interest, it is important to locate these low free energy conformers accurately and reliably. Unfortunately, tests show that the conformational energy orderings by the PM3 method and the reputed DFT methods such as BHandHLYP and B3LYP are quite different. In fact, the PM3 energy ordering is practically useless for the reference purpose. Fortunately, tests have shown that the single point energy ordering by the HF/3-21G(d) method correlates reasonably well with the results by the higher levels of theory [26]. Therefore, the single point energy calculations by the HF/3-21G(d) method were applied to all the 37859 structures and the energy ordering is used to screen the plausible low energy structures.

The 6000 lowest energy structures determined by the HF/3-21G(d) single point energies are then optimized at the HF/3-21G(d) level of theory. To ensure that no important structures were missed, additional sets of 200 structures each were succeeding optimized until the new set of structural optimizations did not produce any new structure that was in the range of 41.8 kJ/mol from the global minimum. As a result of the HF/3-21G(d) optimization process, a total of 7200 PM3 geometries were used and 1618 unique structures were obtained. Among them, 230 structures are in the range of 37.66 kJ/mol from the global minimum. The 230 structures were subjected to further geometry optimizations at the BHandHLYP/6-31G(d) level of theory and 206 unique structures were obtained. The BHandHLYP/6-31G(d) optimization process should be sufficient for locating the low energy conformers as the optimiza-

DOI:10.1088/1674-0068/25/01/77-85

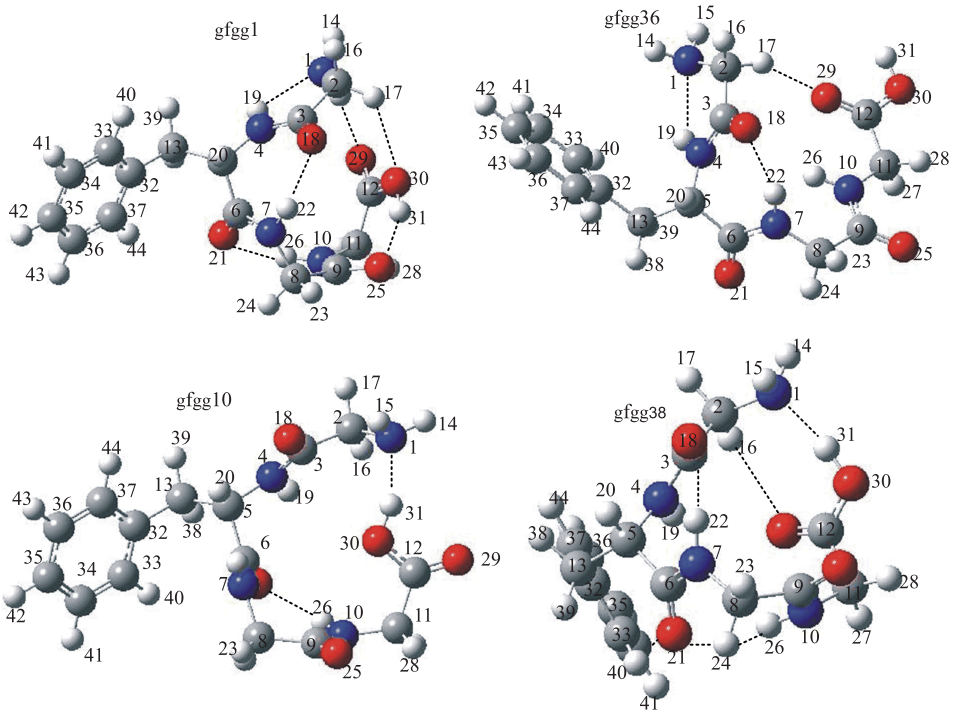

FIG. 2 Structures and hydrogen bonds of four representative conformers.

tion of the last 30 structures of the 230 HF/3-21G(d) optimized structures did not produce any new structure from the global minimum to 20.9 kJ/mol . Single point energy calculations at the level of BHandHLYP/6-311++G(d,p) were applied to the 206 structures and 124 of them were found to be in the range from the global minimum to 20.9 kJ/mol . The frequency calculations at the level of BHandHLYP/6-31G(d) were performed for the 124 and 122 structures were verified to be true local minima.

C. Post geometry optimization analysis

The structures and the frequencies of the low energy conformations were calculated at the BHandHLYP/6-31G(d) level as the 6-31G(d) basis set was known to produce accurate structures [15]. The electronic energies and dipole moments of the conformers were determined at the BHandHLYP/6-311++G(d,p) level of theory as the method is shown to be a good choice for accurately describing the amino acid systems that are rich in hydrogen bonds [27]. The frequencies were used in determining the zero point energies and thermochemical corrections based on the standard harmonic approximation and rigid-rotor model. The frequencies were scaled by a factor of 0.926 as suggested in Refs.[26, 28]. The equilibrium conformational distributions were determined by the relative conformational free energies at the interested temperatures. The hydrogen bond formations were determined by both the geometric criteria of the bond length and the bond angle and the AIM theory [29, 30]. The vertical ionization energy (VIE) is defined as the energy difference between the neutral and ionized species at the geometry of the neutral species and the VIEs of all low energy conformations were calculated at the BHandHLYP/6-311++G(d,p) level.

The trial structures were generated by our in-house developed software written in C++. All other the calculations were performed using the Gaussian 03 suit of programs [31] on our PC clusters.

III. RESULTS AND DISCUSSION

A. Stable conformations of GFGG

There are a total of 122 conformers whose electronic energies lie within the 20.9 kJ/mol range from the global minimum. For convenience, a conformer is denoted with a numeral suffix indicating its relative stability ordered in sequence with ascending electronic energies. For example, the electronic energy of gfgg1, with the structure shown in Fig.2, is the lowest among all GFGG conformers. Some important information such as the relative electronic energies ($E_{\rm el}$), zero point vibrational energies (ZPVE), relative total energies ($E_{\rm t}$=$E_{\rm el}$+ZPVE), dipole moments, rotational constants, VIE and the temperature dependent conformational distributions about 30 representative conformers are listed in Table I. The criteria for selecting the representative conformers in Table I include mainly the following: (i) 10 conformers

DOI:10.1088/1674-0068/25/01/77-85

TABLE I Relative electronic energies (E_{el}), relative zero point vibrational energies (ZPVE), relative total energies (E_t=E_{el}+ZPVE), vertical ionization energies (VIE), rotational constants, dipole moments (D) and the temperature dependent conformational distributions (content) of 30 representative conformers of GFGG[a].

n	Relative energies			VIE	Rotational constant			D	Content[b]/%				
	E_{el}	ZPVE	E_t		X	Y	Z		98 K	198 K	298 K	398 K	498 K
1	0.00	0.00	0.00	8.74	0.407	0.145	0.126	1.795					
2	1.05	−2.85	−1.80	8.61	0.369	0.139	0.120	4.366	27	11	6	3	2
3	2.26	−0.71	1.55	8.82	0.420	0.148	0.132	1.741					
4	2.80	−3.85	−1.05	8.73	0.407	0.134	0.122	4.458	25	19	12	8	6
5	3.14	0.42	3.56	8.73	0.410	0.149	0.125	4.649					
6	5.48	−6.57	−1.05	8.77	0.392	0.132	0.111	6.917	32	33	27	21	16
7	5.61	−6.07	−0.42	8.56	0.387	0.129	0.108	6.455	15	22	20	16	13
8	5.65	−3.81	1.80	8.66	0.423	0.131	0.110	5.906		2	2	2	1
9	6.32	−2.01	4.31	8.85	0.408	0.142	0.118	6.862					
10	6.40	−2.05	4.35	8.91	0.367	0.152	0.120	5.682				1	
14	7.74	−5.31	2.43	8.66	0.279	0.166	0.153	6.832		2	3	3	3
16	8.41	−4.44	3.97	8.82	0.421	0.134	0.112	6.504			1	2	2
17	8.45	−5.23	3.26	8.98	0.437	0.129	0.109	8.423		2	3	3	3
19	9.12	−5.44	3.68	8.70	0.400	0.127	0.105	6.694		1	2	3	3
20	9.50	0.92	10.42	8.77	0.440	0.148	0.130	1.398					
22	10.00	0.13	10.13	8.66	0.323	0.186	0.180	1.763					
23	10.04	−4.77	5.27	8.86	0.363	0.151	0.130	7.935					1
26	10.42	−4.90	5.52	8.80	0.411	0.136	0.122	5.158			1	2	2
27	10.59	−3.93	6.61	9.08	0.294	0.140	0.106	10.098					1
28	10.67	−4.81	5.82	8.71	0.409	0.133	0.111	7.257				1	1
29	10.71	−4.64	6.07	8.44	0.354	0.115	0.094	4.524				1	2
36	11.38	−7.28	4.10	8.86	0.396	0.130	0.108	7.416		3	8	10	11
39	11.80	−6.11	5.69	8.90	0.425	0.126	0.106	7.390			2	3	4
44	13.14	−5.40	7.74	8.53	0.527	0.096	0.092	4.348			1	2	2
46	13.26	0.46	13.72	8.74	0.416	0.150	0.130	1.639					
63	15.90	−5.65	10.25	8.81	0.409	0.131	0.108	7.811				1	2
65	16.19	−5.27	10.92	8.75	0.364	0.139	0.135	3.536					1
72	16.78	−6.61	10.17	8.43	0.333	0.113	0.092	3.923				2	3
85	17.78	−5.19	12.59	8.68	0.273	0.134	0.103	6.634					1
122	20.92	−2.47	18.49	8.72	0.313	0.176	0.138	5.110					

[a] E_{el}, ZPVE and E_t are in kJ/mol, VIE in eV, rotational constants in GHz and D in Debye. For reference, E_{el} of gfgg1 is −1178.5112649 a.u. at the computational level of BHandHLYP/6-311++G(d,p).
[b] Conformational contents below 1% are not listed in the table.

with the lowest electronic energies, (ii) all conformers with an equilibrium content over 1% at any temperature below 498 K, (iii) conformers with dipole moments similar to that of the lowest electronic energy conformer, (iv) the 122nd conformer. As may be seen from Table I, the stability orderings based on the electronic energy and the total energy are quite different due to the influence of the conformation dependent ZPVE. gfgg1 is the global minimum in terms of E_{el}, but is only the 5th most stable in terms of E_t. gfgg2 with the structure shown in Fig.3 is the global minimum in terms of E_t and has a high equilibrium content at a low temperature. However, the conformational stability should be in principle ordered according to the Gibbs free energy that is temperature dependent. According to the free energy scale, gfgg6 is the most stable conformer in the examined temperature range of 98−498 K. gfgg6 is the most stable due to it very small ZPVE that is 6.57 kJ/mol smaller than gfgg1, while its E_{el} is only 5.48 kJ/mol higher than gfgg1. In terms of E_t, gfgg6 is 0.75 kJ/mol less stable than gfgg2, *i.e.*, gfgg2 is dominant when the temperature is close to 0 K. However, gfgg6 (Fig.3) has a lower ZPVE than gfgg2 and the vibrational entropic effect quickly increases with the temperature. gfgg6 becomes more stable than gfgg2 for T>98 K due solely to the entropic effect.

DOI:10.1088/1674-0068/25/01/77-85

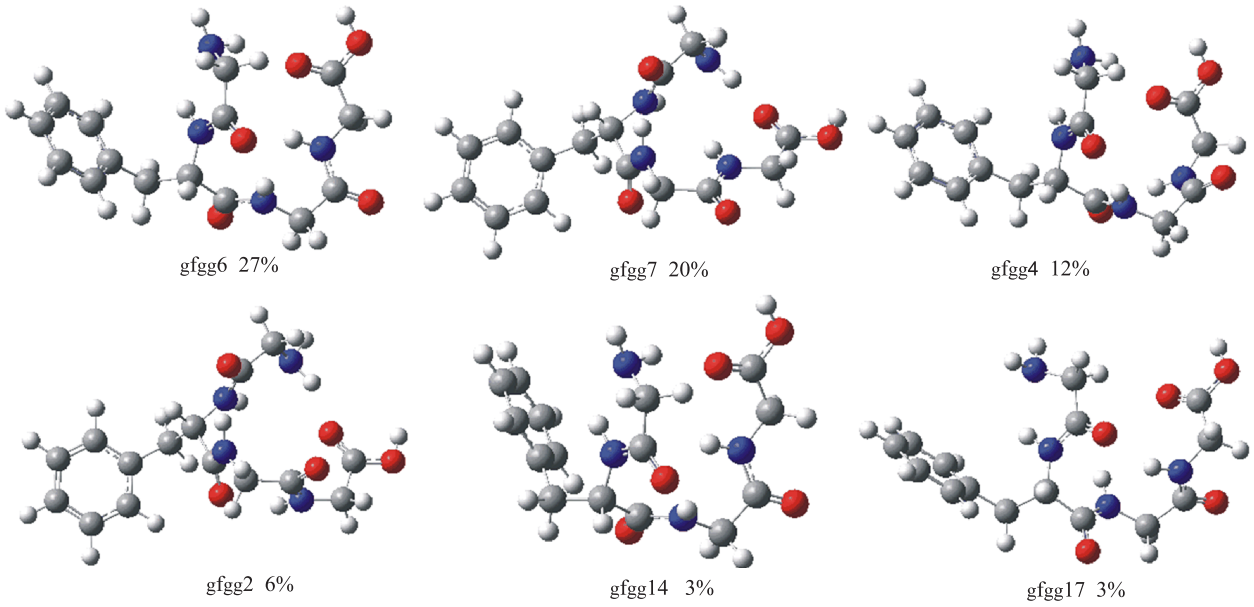

FIG. 3 Structures of six conformers of GFGG with some secondary structural features of proteins. The percentage numbers shown in the figure are the equilibrium concentrations of the conformations at the standard state.

The entropy effect is hugely important in determining the stability and equilibrium content of the GFGG conformers. As shown in Table I, the content of gfgg1, the lowest electronic energy conformer, is below 1% at any temperature. In contrast, gfgg36 has the smallest ZPVE among all low energy conformers and, though with an $E_{\rm el}$ of 11.38 kJ/mol and $E_{\rm t}$ of 4.10 kJ/mol larger than gfgg1, has a significant presence (8%−11%) in the equilibrium ensemble for the room temperature or above. Clearly, gfgg36 (Fig.2) is more important than gfgg1. The entropic effect on the conformational stability points out clearly that the methods of conformational searches based on the static structures alone are not reliable for biomolecules. A static structure based method may locate a conformation such as gfgg1 that is completely unimportant in practice, regardless of the accuracy of the energy calculation method. Truly important conformations can only be found reliably by considering the vibrational energy and its contribution to the free energy. As the energies for the static structures have to be calculated and used for screening the large amount of possible conformations, it is important that the screening should choose a range of candidate structures instead of the lowest energy conformation alone. This point should be kept in mind no matter which conformational search method is used, whether it be Monte Carlo, simulated annealing or genetic algorithm. This is unfortunate as it surely increases the computational cost substantially on top of an already CPU demanding task. The entropic effect is expected to increase with the size of the molecule as the number of vibrational modes increases with the number of atoms in the system. In fact, the entropic effect may become dominate on the conformational stability for biomolecules with a few hundred atoms or more such as large peptides and proteins. In other words, the accuracy requirement on determining the energy of a static structure may only be modest as long as the accuracy of determining the vibrational spectrum is sufficiently high. This is fortunate as an accurate determination of conformational energy requires high level electronic structure methods that are computationally very expensive for large molecules. Instead, the conformational energy may be determined relatively easily by some properly parameterized force field models. This also points to a direction for improving the force field model, *i.e.*, it is important to improve the accuracy of a force field model on accounting the vibrational modes, especially the low energy vibrational modes that are highly influential on the conformational entropy.

It is interesting to note in Table I that the dipole moments of all significant conformations are distinctly different from that of the global minimum of the static electronic structure energy, gfgg1. The dipole moments for gfgg1 and gfgg3 are less than 2.0 Debye, while the dipole moments for all observable conformations are larger than 4.0 Debye. As conformational dipole moments may be accurately determined experimentally [1], such experiments may be used to verify the correctness of our theoretical conformational search results and confirm the entropic effect on the conformational stability discussed here. Further confirmation of the theoretical results may be made if the conformational dipole moments are measured at different temperatures. At the temperature substantially below 98 K, most conformers have the dipole moments of about 4.4 Debye, while conformers with dipole moments larger than 6.0 Debye are minority species. The content of conformers with dipole

DOI:10.1088/1674-0068/25/01/77-85

moments larger than 6.0 Debye increases with the temperature and becomes a clear majority for T>198 K. As some representative conformers, the structures of gfgg1, gfgg10, gfgg36 and gfgg38 are shown in Fig.2 with the hydrogen bonds indicated, while the structures of gfgg2, gfgg4, gfgg6, gfgg7, gfgg14, and gfgg17 are shown in Fig.3 with their concentrations at the room temperature stated. As shown in Fig.2, there are seven hydrogen bonds in gfgg1: N7−H22$\cdots$O18, N10−H26$\cdots$O21, O30−H31$\cdots$O25, N4−H19$\cdots$N1, N1−H15$\cdots$O29, C2−H17$\cdots$O30 and C37−H44$\cdots$O21. The N-terminus and the C-terminus of gfgg1 are connected by two hydrogen bonds N1−H15$\cdots$O29 and C2−H17$\cdots$O30, forming a very compact configuration for the peptide backbone. As only the backbone contains the polar atoms and these polar atoms are roughly evenly distributed inside a small space, it is understandable that the dipole moment of gfgg1 is rather small. Moreover, the compactness of the backbone structure is also the reason for gfgg1 to have a relatively high ZPVE shown in Table I. Except the hydrogen bond of C37−H44$\cdots$O21 that is unique to gfgg1, the other six hydrogen bonds and the connection of the N- and C-terminus by two hydrogen bonds are two common structural features shared by the other five conformers, gfgg1, gfgg3, gfgg20, gfgg22 and gfgg46. They also share the common features of large ZPVE and small dipole moment, as shown in Table I. Incidentally, the conformer with the smallest dipole moment, gfgg20, has the largest ZPVE shown in Table I.

The hydrogen bonds are often helpful for lowering the electronic energy of a conformation and the number of hydrogen bonds has been used as a criterion for screening the trial structures. However, the rule of thumb should be applied with caution. Indeed, every low energy conformer examined, except gfgg10, have at least three hydrogen bonds. However, the electronic energy of gfgg10 is 5.19 kJ/mol below gfgg38, even though gfgg10 has only two hydrogen bonds, N10−H26$\cdots$O21 and O30−H31$\cdots$N1, while there are five hydrogen bonds, N7−H22$\cdots$O18, N10−H26$\cdots$O21, C37−H40$\cdots$O21, O30−H31$\cdots$N1 and C2−H16$\cdots$O29, in gfgg38. Moreover, in these two conformers, the bond lengths of O30−H31$\cdots$N1 are about the same and the bond lengths of N10−H26$\cdots$O21 only differ by 0.1 Å. Therefore, it is not safe to select the trial structures merely by the numbers of hydrogen bonds and the lengths of the hydrogen bonds. The situation is further complicated by the entropic effect, as discussed above. For example, the backbone of gfgg36 is more expanded in space, with loosely connected hydrogen bonds. As a result, gfgg36 has the lowest ZPVE among all the low energy conformers, making it one of the most populous conformations at the room temperature or above.

The IR spectra are informative on revealing the structural characteristics of the conformers. As may be seen from Table I, GFGG is a genuine multi-conformation ensemble for a broad range of temperature. To be helpful for explaining the future IR measurement results and be concise, the IR spectrum of the most populous conformer, gfgg6, is shown in Fig.4 [32]. The vibrational mode of peak A marked in Fig.4 is characteristic of the planar structure formed by two amino acid residues in GFGG, a common feature shared by several important conformers such as gfgg7 and gfgg36 as well as gfgg39. Peak A is very weak for the main low total energy conformers such as gfgg2 and gfgg4 and absent for the other low electronic energy conformers such as gfgg1 and gfgg3. As indicated in Table I, the ensemble averaged intensity of peak A should increase with the temperature in a temperature dependent IR spectra measurement. Such experiments may be used to further verify the theoretical search results of GFGG conformers and the entropic effects on the conformational stability discussed here.

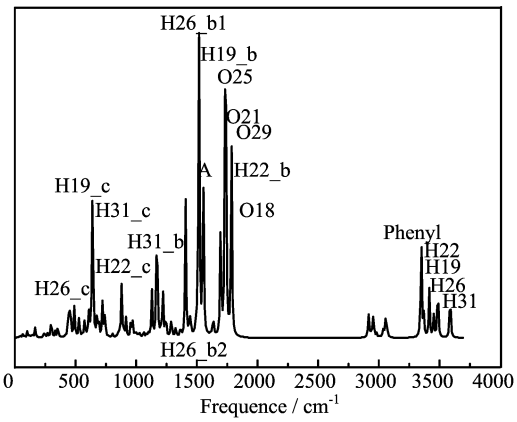

FIG. 4 Vibrational spectrum of the most populous conformer of GFGG, gfgg6.

B. Connection with the secondary structure of proteins

There is a local structure named α-helix in the secondary structure of proteins. The α-helix is a common motif in the secondary structure of proteins. It is a right-handed coiled or spiral conformation, in which the hydrogen atom in the first peptide bond forms hydrogen bond with the oxygen atom in the forth peptide bond [33, 34]. Among different types of local structures in proteins, the α-helix is the most regular and the most predictable from sequence, as well as the most prevalent. As there are only three peptide bonds in GFGG, finding a local structure with a literate α-helix feature in GFGG conformations is impossible. However, it is interesting to note that some GFGG conformations bear the fingerprint of the α-helix structure. In the conformers of gfgg1, gfgg3, gfgg20, gfgg22 and gfgg46, the configuration of three hydrogen bonds, N7−H22$\cdots$O18, N10−H26$\cdots$O21 and O30−H31$\cdots$O25, follows a rule that the oxygen atoms of the first, second and third pep-

DOI:10.1088/1674-0068/25/01/77-85

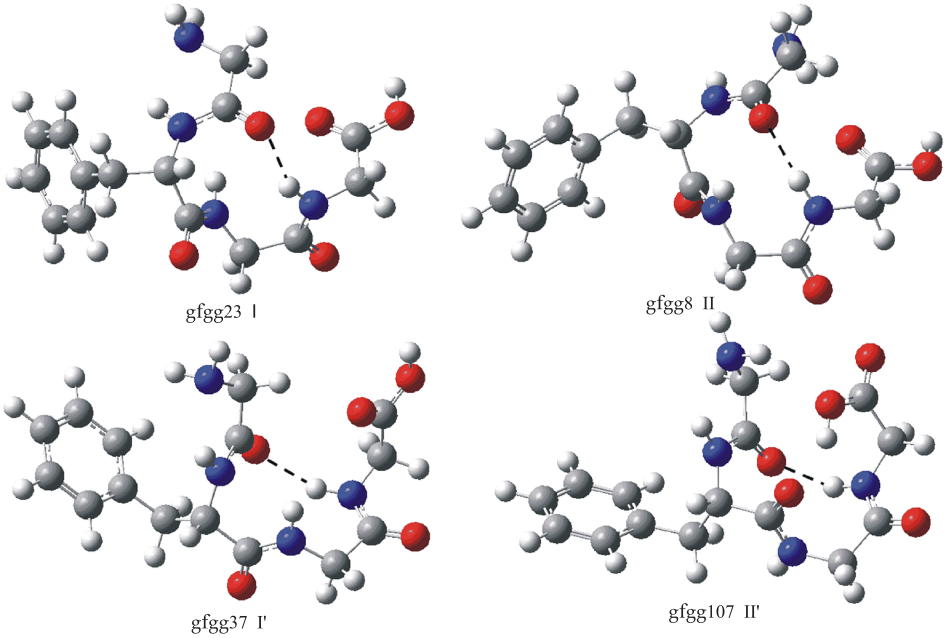

FIG. 5 Illustration of the four β-turn types found in the GFGG conformers.

tide bonds form the hydrogen bonds with the hydrogen atoms of the second and third peptide bonds and the C-terminus *trans*-carboxyl group, respectively. The similarity between the hydrogen bonds in these conformers and that in the α-helix is that the H-bonds are formed between neighboring peptide bonds of equal spacing. Another similarity is that the side chains of these conformers and the α-helix all extend outwards. It should be noted that the possibility for a regular spacing pattern in a short peptide like GFGG is very limited. If the number of peptide bonds is sufficiently large in a large peptide or protein, there will be multiple possibilities for forming a regularly patterned H-bond network, in which the α-helix is only a special case. Though the benefit of the H-bonds in lowering the electronic energy of gfgg1 or gfgg3 is offset by the increased ZPVE due to the accompanying over compacted local structure, the α-helix in a large biomolecule may have the best compromise for forming the maximum number of H-bonds in a spatially reasonably expanded configuration, making it a prominent feature in the system. The β-turn structure is known to play an important role in the folding process of polypeptide chain [35]. A β-turn may be defined for four consecutive residues (denoted by i, i+1, i+2 and i+3) if the distance between the Cα atom of residue i and the Cα atom of residue i+3 is less than 7 Å and if the central two residues are not helical [35−39]. The β-turns are assigned to one of 9 classes on the basis of φ and ψ angles of residues i+1 and i+2. For the convenience of referencing, these angles as defined in Ref.[37] are reproduced in Table II. Based on the common definition, the φ and ψ angles are allowed to vary by $\pm 30°$ from these ideal values with the added flexibility of one angle being allowed to deviate by as much as 40°. According to these definitions, β-turn structures of type I, II, I′ and II′ are found in GFGG conformers: gfgg23, gfgg43, gfgg78 and gfgg89 belong to type I, gfgg8, gfgg11, gfgg48, gfgg110 and gfgg117 belong to type II, gfgg37, gfgg57, gfgg97 and gfgg102 belong to type I′, gfgg107 belongs to type II′. The structures of representative conformers with the β-turns are shown in Fig.5. Compared to the α-helix like structures, the β-turn conformations in GFGG are less energetically favorable. Notice, however, the observation on the relative energy ordering of the α-helix and the β-turn may not be applicable to large peptides and proteins due to the effects of other interactions.

TABLE II The ideal angles for each of the β-turn types[a]. Types VIa1, VIa2 and VIb turns are subject to the additional condition that residue i must be a *cis*-proline.

Type	φ(i+1)	ψ(i+1)	φ(i+2)	ψ(i+2)	Addition
I	−60	−30	−90	0	
II	−60	120	80	0	
VIII	−60	−30	−120	120	
I′	60	30	90	0	
II′	60	−120	−80	0	
VIa1	−60	120	−90	0	*cis*-proline(i+2)
VIa2	−120	120	−60	0	*cis*-proline(i+2)
VIb	−135	135	−75	160	*cis*-proline(i+2)

[a] Turns which do not fit any of the above criteria are classified as type IV [37].

A γ-turn is defined for three residues i, i+1, i+2

DOI:10.1088/1674-0068/25/01/77-85

if a hydrogen bond exists between residues i and i+2 and the φ and ψ angles of residue i+1 fall within 40° of one of the following two classes: classic type, φ(i+1)=75.0, ψ(i+1)=−64.0; inverse type, φ(i+1)=−79.0, ψ(i+1)=69.0 [40, 41]. As there are four residues in a tetrapeptide, it is relatively easy for the GFGG conformations to meet the definition of the γ-turns. Nevertheless, it is quite surprising to find so many γ-turns in the 122 low energy conformations of GFGG: there are 16 conformations of classic type γ-turn and 71 conformations of inverse type γ-turn. It appears that the γ-turns are the favorite configurations for small peptides.

IV. CONCLUSION

We have performed an extensive computational search of conformations of tetrapeptide GFGG in gas phase. A large set of trial structures generated by full combinations of all internal single-bond rotamers were optimized by a hierarchical methods of PM3, HF/3-21(d) and BHandHLYP/6-31(d,p), followed by single-point energy calculations at the BHandHLYP/6-311++(d,p) level. A total of 122 conformers are found within a range of 20.9 kJ/mol from the global the electronic energy minimum. The temperature dependent conformational distributions are calculated. The structures of important conformers are analyzed in detail. It is found that, while the hydrogen bonds are helpful for lowering the electronic energies, the zero point vibrational energy and the vibrational entropy have strong effects on the stabilities of conformers. The most important conformers are those with low zero point vibrational energies and high vibrational entropies, while the global minimum conformer in the electronic energy scale is undetectable in the equilibrium ensemble of any temperature. It is pointed out that the experimental verification of this dramatic theoretical prediction may be carried out by measuring the dipole moments and certain characteristic IR modes of the conformers.

The structures of the low energy GFGG conformers are also analyzed in the context of the secondary structures of proteins. Similarity between the local structures of low energy GFGG conformers and the α-helix is revealed and many β- and γ-turn local structures in GFGG conformers are found. The richness of the secondary structures of proteins in GFGG conformers indicates that valuable information about the protein structures may be obtained by carefully analyzing the structures of peptides consisting of only a limited number of amino acid residues.

V. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No.11074233) and the State Key Development Program for Basic Research of China (No.2012CB215405).

[1] I. Compagnon, F. Hagemeister, R. Antoine, D. Rayane, M. Broyer, P. Dugourd, R. Hudgins, and M. Jarrold, J. Am. Chem. Soc. **123**, 8440 (2001).
[2] E. J. Cocinero, A. Lesarri, J. Grabow, J. C. López, and J. L. Alonso, ChemPhysChem. **8**, 599 (2007).
[3] R. Linder, K. Seefeld, A. Vavra, and K. Kleinermanns, Chem. Phys. Lett. **453**, 1 (2008).
[4] A. Gunther, G. Geipel, and G. Bernhard, J. Poly. **26**, 59 (2006).
[5] M. Kainosho and K. Ajisaka, J. Am. Chem. Soc. **97**, 5630 (1975).
[6] B. Jansik, A. Rizzo, and H. Agren, J. Chem. Phys. **111**, 446 (2007).
[7] S. Dayalan, N. D. Gooneratne, S. Bevinakoppa, and H. Schroder, Bioinformation **1**, 78 (2006).
[8] S. Anishetty, G. Pennathur, and R. Anishetty, BMC Struct. Biol. **2**, 9 (2002).
[9] W. Birkmayer, P. Riederer, W. Linauer, and J. Knoll, J. Neural Transmission **59**, 81(1984).
[10] F. Camacho and J. Mazuecos, J Drugs. Dermatol. **1**, 127 (2002).
[11] National Research Council, Food and Nutrition Board, *Recommended Dietary Allowances*. 11th Edn. Washington DC: National Academy Press, (1998).
[12] J. Pietz, Curr. Opin. Neurol. **11**, 679 (1998).
[13] M. Nichifor, E. H. Schacht, L. W. Seymour, D. Anderson, and M. Shoaibi, J. Bioact. Comp. Polym. **12**, 265 (1997)
[14] M. Nichifor, E. H. Schachta, and L. W. Seymou, J. Controlled Release **48**, 165 (1997).
[15] Z. Huang, W. Yu, and Z. Lin, J. Mol. Struct. **758**, 193 (2005).
[16] L. Snoek, E. G. Robertson, R. T. Kroemer, and J. P. Simons, Chem. Phys. Lett. **321**, 49 (2000).
[17] G. von Helden, I. Compagnon, M. Blorn, M. Frankowski, U. Erlekam, J. Oomens, B. Brauer, R. B. Gerber, and G. Meijer, Phys. Chem. Chem. Phys. **10**, 1248 (2008).
[18] A. G. Abo-Riziq, J. E. Bushnell, B. Crews, M. P. Callahan, L. Grace, and M. S. De Vries, Int. J. Quantum Chem. **105**, 437 (2005).
[19] D. Řeha, H. Valdés, J. Vondráek, P. Hobza, A. Abu-Riziq, B. Crews, and M. S. de Vries, Chem. Eur. J. **11**, 6803 (2005).
[20] H. Valdés, K. Pluháckova, M. Pitonák, J. Rezac, and P. Hobza, Phys. Chem. Chem. Phys. **10**, 2747 (2008).
[21] M. Zhang, Z. Huang, and Z. Lin, J. Chem. Phys. **122**, 134313 (2005).
[22] Z. Huang and Z. Lin, J. Phys. Chem. A **109**, 2556 (2005).
[23] S. Ling, W. Yu, Z. Huang, Z. Lin, M. Haranczyk, and M. Gutowski, J. Phys. Chem. A **110**, 12282 (2006).
[24] A. Perczel, J. G. Angyan, M. Kajtar, W. Viviani, J. Rivail, J. Marcoccia, and I. G. Csizmadia, J. Am. Chem. Soc. **113**, 6256 (1991).
[25] A. Perczel, P. Hudaky, and I. G. Csizmadia, J. Mol. Struct. **500**, 59 (2000).

DOI:10.1088/1674-0068/25/01/77-85

[26] W. Yu, X. Xu, H. Li, R. Pang, K. Fang, and Z. Lin, J. Comput. Chem. **30**, 2105 (2009)
[27] W. Yu, L. Liang, Z. Lin, S. Ling, M. Haranczyk, and M. Gutowski, J. Comput. Chem. **30**, 589 (2009).
[28] A. P. Scott and L. Radom, J. Phys. Chem. **100**, 16502 (1996).
[29] R. F. M. Bader, *Atoms in Molecules A Quantum Theory*, Oxford: Clarendon Press, (1990).
[30] P. L. A. Popelier, J. Phys. Chem. A **102**, 1873 (1998).
[31] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M.Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J.L. Andres, A. C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, *Gaussin 98, Revision A.9*, Pittsburgh PA: Gaussian, Inc., (1998).
[32] The structural data and IR spectra of the other important conformers are available by contacting the authors.
[33] D. Eisenberg, Proc. Natl. Acad. Sci. **100**, 11207 (2003).
[34] C. Branden and J. Tooze, I*ntroduction to Protein Structure,* 2nd Edn. New York: Garland Publishing, (1999).
[35] P. N. Lewis, F. A. Momany, and H. A. Scheraga, Proc. Nat. Acad. Sci. **68**, 2293 (1971).
[36] E. G. Hutchinson and J. M. Thornton, Protein Science **3**, 2207 (1994).
[37] W. Kabsch and C. Sander, Biopolymers **22**, 2577 (1983).
[38] J. S. Richardson, Adv. Protein Chem. **34**, 167 (1981).
[39] C. M. Venkatachalam, Biopolymers **6**, 1425 (1968).
[40] E. J. Milner-White, B. M. Ross, R. Ismail, K. Belhadj-Mastefa, and R. Poet, J. Mol. Biol. **204**, 777 (1988).
[41] G. D. Rose, L. M. Gierasch, and J. A. Smith, Adv. Prot. Chem. **37**, 1 (1985).

DOI:10.1088/1674-0068/25/01/77-85