

# SCIENTIFIC REPORTS



OPEN

## Computational study on single molecular spectroscopy of tyrosin-glycine, tryptophane-glycine and glycine-tryptophane

Bing Yang<sup>1</sup>, Shixue Liu<sup>2</sup> & Zijin Lin<sup>1</sup> 

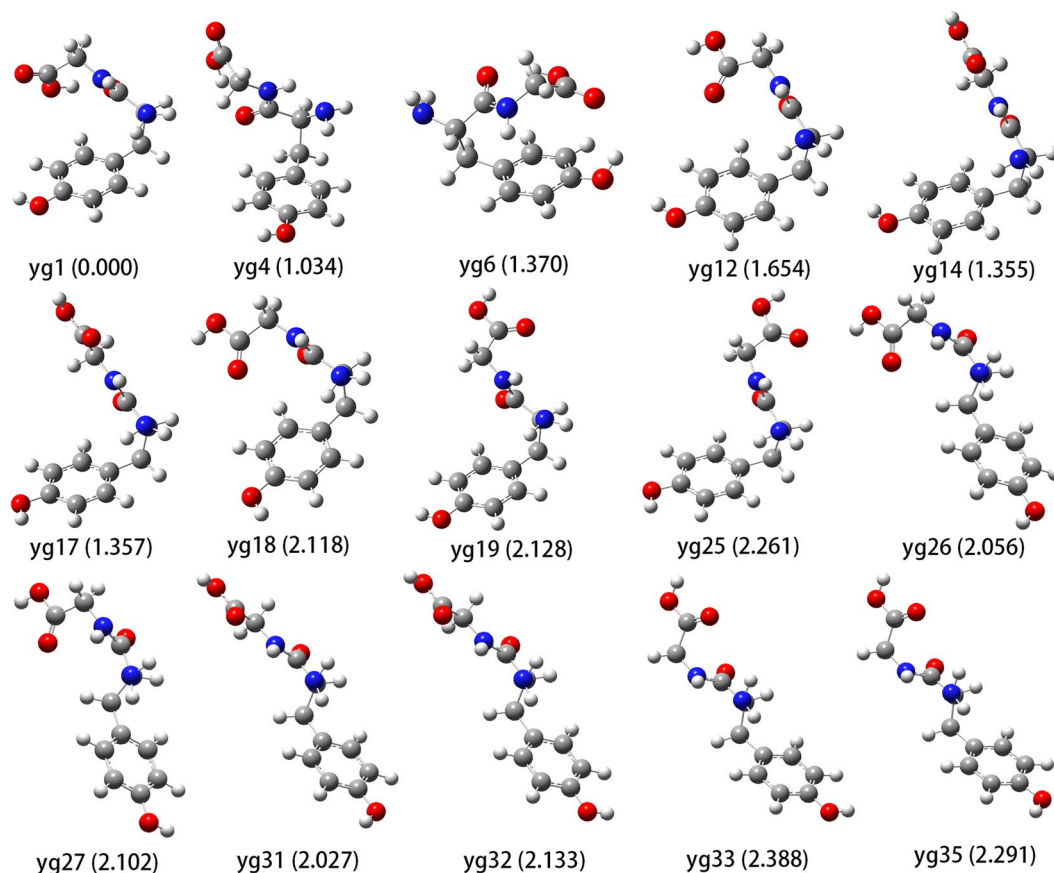
Quantum chemistry calculations play a fundamental role in revealing the molecular structures observed in gas-phase spectroscopic measurements. The supersonic jet cooling widely used in single molecular spectroscopy experiment is a non-equilibrium process and often causes confusion on the theoretical and experimental comparison. A computational approach is proposed here to account for the effect of the non-equilibrium cooling on the experimental spectra and applied to the cases of tyrosin-glycine (YG), tryptophane-glycine (WG) and glycine-tryptophane (GW). The low energy conformers of YG, WG and GW are obtained through thorough conformational searches. The structural features and equilibrium distributions of conformations and the energy barriers for conformer conversions are then determined. Three classes of transition energy barriers, high, medium and low, are found for the conversions among conformers with distinctly different, similar and the same structural types, respectively. The final conformation populations are determined by assuming an initial temperature of about 450 K and allowing for only the conformation conversion with a low energy barrier to occur during the rapid cooling process. The results provide a natural explanation for the numbers of YG, WG and GW conformations observed experimentally. The theoretical conformation assignments are also in good agreement with the experimental IR data.

The three-dimensional structure of biomolecule is a basic factor determining its biological functions and properties such as molecular recognition and ligand bindings. Much experimental and theoretical effort has been devoted to elucidate the biomolecular structure and the interactions therein. Single molecular spectroscopy techniques have been frequently used to probe the structures of biomolecules ranging in size from simple amino acids to entire protein assemblies<sup>1-5</sup>. However, the spectroscopic experiment yields only indirect structural information. The molecular structures can be determined only by comparing the experimental spectra with the results of quantum chemistry calculations<sup>6-9</sup>. Unfortunately, though the spectroscopic technique and high accuracy quantum chemistry computations have reached some level of maturity, the conformation assignment is still encountered with ambiguity<sup>7,8,10</sup>.

The practice of comparing the theoretical and experimental spectra often suffers some of the shortcomings: (1) Theoretical IR spectra are computed based on a set of candidate structures determined by some crude conformational search<sup>7,10</sup>. As the set of candidate structures may miss important low energy conformations of the molecule, the quality of structural assignment thus determined is rather questionable. (2) The structural assignment is made based on the best match between the theoretical and experimental spectra. However, the matched structures may be unfavorable based on the consideration of conformational energies<sup>1-3</sup>. (3) It is quite common to see the absence of the theoretically predicted low energy structures in the experimental spectra<sup>9</sup>. To explain the experimental results, some non-equilibrium dynamic mechanism is required. For example, a mechanism of non-radiative deactivation of excited states of folded peptide conformers has been proposed<sup>11,12</sup>. The non-radiative deactivation mechanism is capable of explaining a number of experimental spectra and is gaining an increased acceptance<sup>11-19</sup>. However, the facts about the absence of folded WG conformer and the presence of

<sup>1</sup>Hefei National Laboratory for Physical Sciences at Microscale & CAS Key Laboratory of Strongly-Coupled Quantum Matter Physics, Department of Physics, University of Science and Technology of China, Hefei, 230026, China.

<sup>2</sup>Department of Nanotechnology for Sustainable Energy, Kwansai Gakuin University, Gakuen 2-1, Sanda, Hyogo, 669-1337, Japan. Correspondence and requests for materials should be addressed to Z.L. (email: [zjlin@ustc.edu.cn](mailto:zjlin@ustc.edu.cn))



**Figure 1.** Representative YG conformations. Relative total energies (the sum of the electronic energy and the zero-point vibrational energy, in kcal/mol) of the conformers are shown in the parentheses.

folded GW conformer in the experiments cast serious doubt on the non-radiative deactivation mechanism as the mechanism should be equally applicable to both cases of WG and GW (W = tryptophan, G = glycine)<sup>8,10,18</sup>. A more natural explanation is desirable. (4) The effects of the supersonic expansion cooling process on the experimental spectra are considered only vaguely. As the global free energy minima of WG and WGG at 300 K match the experimental results<sup>9</sup>, it is suggested that the high population structures at about 300 K might have approximately been kept during the supersonic cooling process<sup>20</sup>. Similar result is also found recently on the IR spectra of YG conformers (Y = tyrosine)<sup>21</sup>. However, the number of conformations predicted by the free energy consideration is substantially more than that observed experimentally. It is then natural to conclude that the experimentally observed structures are the combined effect of conformational distribution at some high formation temperature and the dynamics of non-equilibrium cooling process<sup>20–22</sup>. Unfortunately, no serious effort is made to establish the correspondence between the computationally determined and experimentally observed structures.

This paper reports a detailed theoretical study on the conformations of YG, WG and GW, with the aim of providing an improved explanation of the experimental results<sup>7,8,10</sup>. The low energy structures are determined through extensive conformational searches and characterized based on their structural features. The conformational transition energy barriers within each structural class and between different structural classes are computed. The initial conformation distributions and the transition energy barriers are used to predict the detectable structures after the rapid supersonic expansion cooling process. The analysis provides a clear and coherent mechanism that agrees well with the experimental observations. The computational approach described here is a reliable way for determining the structures observed by supersonic jet cooling based molecular spectroscopy.

## Results and Discussion

**Tyrosin-glycine.** *Conformational search results.* Through extensive conformational searches, many local minima were found on the potential energy surface. Table 1 shows the relative energies, hydrogen bonds (H-bonds), structural types and equilibrium distributions of YG conformers of interest. The structures of representative conformers are displayed in Fig. 1. Here, a YG conformer is denoted as *yg**n* and the numeral suffix *n* refers to its position in the conformational sequence ordered by ascending electronic energies. An H-bond is defined by a cutoff distance of 2.8 Å.

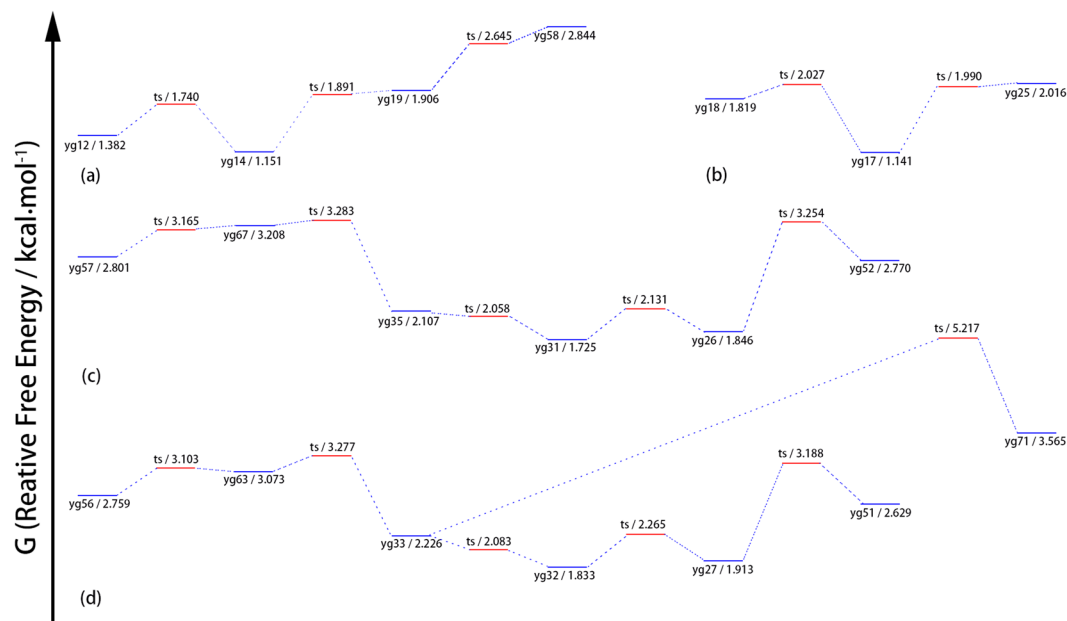
As shown in Table 1 as well as Fig. 1, most of the low energy conformers adopt folded configurations ( $\gamma$  type backbone), associated with the formations of strong H-bonds. However, the conformational distribution shows a strong temperature dependence or entropic effect. With the increase of temperature, the extended structures, typically with a  $\beta$  or  $\epsilon$  backbone and a *g*-side chain, become increasingly more important in the conformation

Conf.	Energy	H-bonds <sup>1*</sup>		Type <sup>2*</sup>	Distributions		
		Backbone	Main/Side-Chain		98K	298K	450K
yg1	0.000	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A1-γ <sub>D</sub> (F)-g+/+	100	11.42	3.05
yg2	0.523	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	O <sub>S</sub> H...O = C <sub>T</sub>	A2-γ <sub>D</sub> (F)-g+/+	—	—	—
yg3	0.786	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A1-γ <sub>D</sub> (F)-g+/-	—	4.25	1.72
yg4	0.973	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A1-γ <sub>L</sub> (F)-g+/+	—	3.18	1.56
yg5	1.121	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A1-γ <sub>L</sub> (F)-g+/-	—	2.98	1.57
yg6	1.230	N <sub>i</sub> H...OC <sub>PB</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	O <sub>S</sub> H...O = C <sub>T</sub>	B-γ <sub>L</sub> (F)-a/-	—	—	—
yg7	1.779	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A2-γ <sub>D</sub> (F)-g-/-	—	3.17	2.34
yg8	1.783	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A2-γ <sub>D</sub> (F)-g-/+	—	3.54	2.53
yg9	1.819	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A2-γ <sub>L</sub> (F)-g-/-	—	1.59	1.24
yg10	1.843	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A2-γ <sub>L</sub> (F)-g-/+	—	2.01	1.55
yg11	1.858	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	—	A2-γ <sub>L</sub> (F)-g+/+	—	2.03	1.45
yg12	1.880	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-ε <sub>D</sub> (E)-g+/+	—	2.25	1.88
yg13	2.001	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	—	A2-γ <sub>L</sub> (F)-g+/-	—	1.35	1.07
yg14	2.011	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-β(E)-g+/+	—	11.62	8.65
yg17	2.113	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-β(E)-g+/-	—	10.57	8.10
yg18	2.360	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-ε <sub>D</sub> (E)-g+/-	—	1.02	1.10
yg19	2.375	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-ε <sub>L</sub> (E)-g+/+	—	1.49	1.58
yg25	2.552	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-ε <sub>L</sub> (E)-g+/-	—	1.33	1.55
yg26	2.613	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A2-ε <sub>D</sub> (E)-g-/+	—	3.37	3.72
yg27	2.619	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A2-ε <sub>D</sub> (E)-g-/-	—	3.91	4.44
yg29	2.710	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A1-γ <sub>D</sub> (F)-g-/+	—	—	1.00
yg30	2.721	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A1-γ <sub>D</sub> (F)-g-/-	—	—	1.00
yg31	2.723	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A2-β(E)-g-/+	—	4.35	4.99
yg32	2.725	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A2-β(E)-g-/-	—	2.75	3.23
yg33	2.744	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A2-ε <sub>L</sub> (E)-g-/-	—	2.18	2.87
yg35	2.777	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A2-ε <sub>L</sub> (E)-g-/+	—	3.07	3.93
yg38	2.905	N <sub>PB</sub> H...N <sub>i</sub> ; O <sub>T</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	A1-γ <sub>L</sub> (F)-g-/+	—	—	1.15
yg51	3.489	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-ε <sub>D</sub> (E)-g-/-	—	1.17	2.02
yg52	3.493	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-ε <sub>D</sub> (E)-g-/+	—	1.17	2.07
yg56	3.607	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-β(E)-g-/-	—	—	1.56
yg57	3.638	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-β(E)-g-/+	—	—	1.51
yg58	3.650	N <sub>PB</sub> H...N <sub>i</sub>	—	A2-ε <sub>L</sub> (E)-g+/+	—	1.00	1.90
yg63	3.900	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-ε <sub>L</sub> (E)-g-/-	—	—	1.53
yg67	3.958	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A1-ε <sub>L</sub> (E)-g-/+	—	—	1.66
yg71	4.183	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A2-α <sub>L</sub> (E)-g-/-	—	—	1.29
yg72	4.214	N <sub>PB</sub> H...N <sub>i</sub>	N <sub>i</sub> H...π	A2-α <sub>L</sub> (E)-g-/+	—	—	1.04
yg100	5.416	N <sub>i</sub> H...OC <sub>PB</sub>	N <sub>i</sub> H...π	B-β(E)-g-/-	—	—	1.15

**Table 1.** Relative electronic energies (Energy, in kcal/mol), H-bond networks and structural types (Type) for all YG conformers (Conf.) of interest. The equilibrium distributions (%) at three representative temperatures are also shown but an equilibrium content below 1% is denoted as “—”. 1\*: N<sub>i</sub> refers to the N-terminal nitrogen atom, N<sub>PB</sub> is the nitrogen atom in the peptide-bond, and O<sub>T</sub> stands for the oxygen atom of the C-terminal hydroxyl. 2\*: The label of structural type consists of four parts. The first part, A1, A2 or B, is used to indicate the swing direction and H-bond type of the amino terminus. See yg14, yg31 and yg6 in Fig. 2 for the configurations of A1, A2 and B, respectively. The second part of the label refers to the dihedral angle (φ<sub>2</sub>, ψ<sub>2</sub>) in the Ramachandran plot. The third part of the label indicates the swing direction of the side chain, with g+, g- and a correspond to the dihedral angle of N-C<sub>α</sub>-C<sub>β</sub>-C<sub>γ</sub> close to +180°, +60°, and -60°, respectively. The last part, + or - after the slash, denotes the side chain hydroxyl orientation.

ensemble. Only one H-bond concerning the peptide bond is formed in the extended α/β/ε backbone. Moreover, the g- orientation is helpful to keep the side chain away from the C-terminal backbone and allow for a more liberal swing of the side chain to increase the entropic effect. As a result of the entropic effect, the equilibrium ensemble, though dominated by the low-energy folded structures at low temperature, consists of mainly extended structures at the room temperature or above.

As seen in Table 1, the energy of a conformer of interest can be more than 5 kcal/mol above the global minimum. At T~450 K, the relative electronic energies for the most populous structures fall into the interval of 2 to 3 kcal/mol. Notice that T~450 K is often expected for the relevant experimental conditions such as laser heating<sup>23,24</sup>. Therefore, conformers within a sufficiently large energy range of the global minimum should be thoroughly searched in order to reliably determine the most important conformers under the experimental condition.



**Figure 2.** The free energy profile for four groups of YG conformers with extended backbone structures: **(a)** conformers with g+/+ side chains, **(b)** conformers with g+/- side chains, **(c)** conformers with g-/+ side chains, **(d)** conformers with g-- side chains. “ts” in the graph stands for “transition state”.

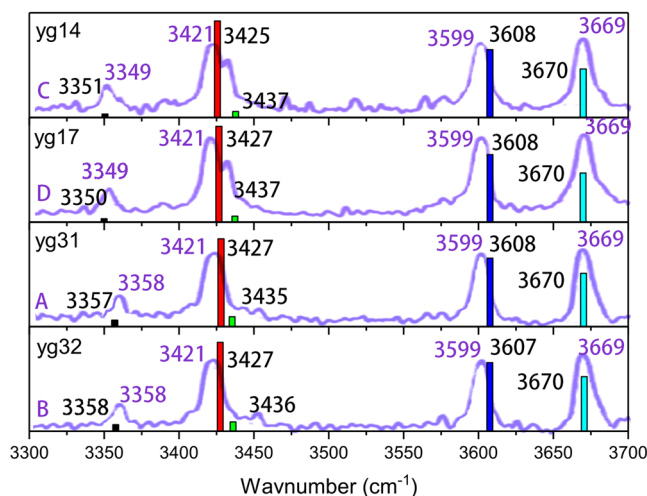
*Interpretation of the experimental results.* Four different YG conformers were observed by double-resonance spectroscopic technique<sup>7</sup>. The experimental results were obtained after rapid cooling by supersonic expansion that resulted in a very low temperature (about 10 K). As shown in Table 1, there is only one single structure in the equilibrium conformational ensemble at 10 K. The discrepancy between the experimental results and the theoretical equilibrium conformational distribution is not surprising, however. The discrepancy is caused by the fact that the conformational ensemble does not have the time to reach its equilibrium state in the rapid cooling process. Instead, a conformer existed before the supersonic expansion may only relax to a local minimum connected by a low energy barrier in the supersonic expansion cooling process<sup>6</sup>. Therefore, a proper explanation of the experimental results requires a detailed examination of the dynamics of the rapid cooling process. The theoretical and experimental comparison is further complicated by uncertainties regarding the conformational ensemble before the supersonic expansion cooling. The initial conformational ensemble was obtained by laser ablation and may not reach a thermodynamic equilibrium before the cooling. Consequently, a rigorous account of the experimental results may require a complete modeling of the laser heating and jet cooling processes that is rather difficult and not attempted here. Instead, it is illustrated below that the experiment can be accounted for quantitatively by using an effective laser heating temperature and considering the effect of transition energy barriers on the conformational conversions in the rapid cooling process.

The free energy barriers for transitions among conformers of similar and different structural characteristics were examined systematically. The free energy barriers are found to be weakly temperature dependent and generally decrease with the decrease of temperature. However, the free energy barriers at low temperature (10 K) are found to be representative and are used in the following discussion.

The obtained energy barriers can be classified into three categories: high, medium and low. The high energy barrier ( $\geq 8$  kcal/mol) corresponds to the transformation between conformers with very different backbone configurations. For example, the transition of yg4 $\leftrightarrow$ yg1 corresponds to the transition of  $\gamma_L \leftrightarrow \gamma_D$  in the folded backbone and the transition of yg12/yg14 $\leftrightarrow$ yg1 corresponds to the transition of  $\epsilon/\beta$  in the extend backbone  $\leftrightarrow \gamma$  in the folded backbone. The energy barrier for such a transformation is high as the transition process involves the breakage and reformation of strong H-bond. This kind of transformation requires a very long relaxation time to occur and may be safely ignored under the experimental condition.

The medium energy barrier (2–8 kcal/mol) is associated with the transformation between conformers with different side chain swing directions. Examples include the transition of yg12 $\leftrightarrow$ yg52 for the side chain transformation of g+  $\leftrightarrow$  g- and yg14  $\leftrightarrow$  yg17 for the transformation of the side chain hydroxyl orientation of +  $\leftrightarrow$  -. Though no H-bond breaking is involved, the energy barrier is moderately high as the transformation is encountered with some steric hindrance. The transformation with a medium energy barrier is also expected to be negligible during the rapid cooling process of supersonic expansion<sup>6</sup>.

A low energy barrier ( $\leq 2$  kcal/mol) is found for the transition involving only the rotation of terminal groups, e.g., yg58 $\leftrightarrow$ yg36 (the swing of amidogen direction), or conformers with similar extended backbone configurations, e.g., yg28 $\leftrightarrow$ yg36 ( $\beta \leftrightarrow \epsilon$  of the extend backbone). The transformation between conformers separated by the



**Figure 3.** Comparison of the experimental (curves and numbers in purple) and theoretical (colored bars and numbers in black) IR spectra of YG conformations. The experimental results are taken from ref.<sup>7</sup>.

low energy barrier is expected to be sufficiently fast and can easily take place during the supersonic expansion cooling process.

Information about the effective temperature of the conformational ensemble obtained by laser-induced desorption is required to understand the conformational distribution after the supersonic expansion cooling. Based on thermal equilibrium desorption kinetics, the temperature is found to be in the range of 380–450 K or 550–670 K, depending on the heating condition<sup>23</sup>. For convenience, an effective temperature of 450 K is temporarily assumed for the conformational ensemble before the supersonic expansion. The use of 450 K is arguable due to a lack of firm experimental support. However, the choice is found to produce results in good agreement with the experiment on YG as well as WG and GW, as to be seen below. Moreover, it is noted that the following discussion is not qualitatively affected by allowing for an uncertainty of 50 K in the assumed temperature.

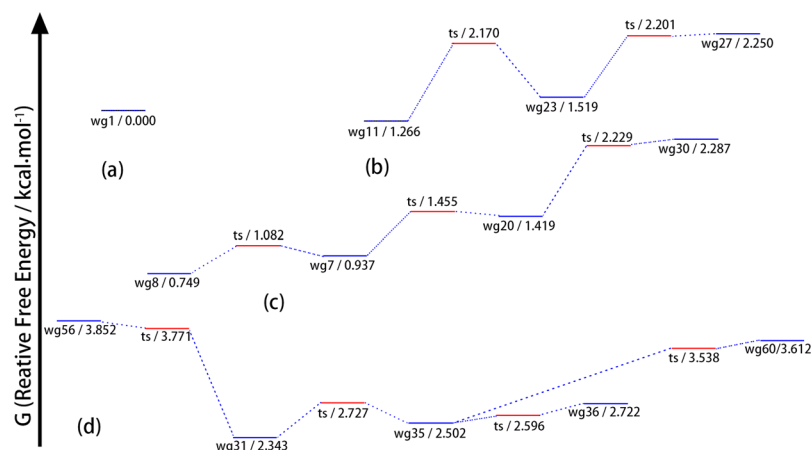
With the assumed temperature of 450 K, Table 1 shows that many conformers, including the folded conformers of yg1~yg11, yg13, yg29, yg30, yg38 and yg100, are present in the equilibrium ensemble. However, as discussed above on the energy barriers, no conversion between a folded conformer and an extended conformer may occur. Similarly, there is no conversion between differently folded conformers. As a result, no folded conformer, including yg1, may accumulate through the conversion process to have a population above 5%.

The conversion between extended conformers with a similar side chain configuration can take place easily. There are four types of side chain configurations,  $g+/+$ ,  $g+/-$ ,  $g-/+$  and  $g-/-$ . Among the conformers of interest, yg12, yg14, yg19 and yg58 belong to  $g+/+$ , yg17, yg18 and yg25 belong to  $g+/-$ , yg6, yg31, yg35, yg52, yg57 and yg67 belong to  $g-/+$ , while yg27, yg32, yg33, yg51, yg56, yg63 and yg71 belong to  $g-/-$ . The four groups of conformers are separated by some medium energy barriers and the inter-group conversion is negligible. However, the energy barrier for intra-group conformational conversion is quite low. Figure 2 shows the free energy profile for the four groups of conformers. As shown in Fig. 2a, the energy barriers for the conversions of  $g+/+$  conformers are all less than 0.4 kcal/mol. The four conformers would converge to their local free energy minimum, yg14, in the cooling process. As a result, yg14 has a 14% concentration in the final conformational ensemble. Similarly, there are about 11% of yg17, 18% of yg31 and 17% of yg32 in the final conformational ensemble. The final concentrations of yg14, yg17, yg31 and yg32 are all above 10% and much higher than that of any other conformers. The four conformers should therefore be observable, while others may be indistinguishable from the background noise. The theoretical analysis is in good accord with the experimental findings.

To further validate the above conformational analysis, IR spectra of yg14, yg17, yg31 and yg32 are computed and compared with the experiment. For the comparison, the theoretical harmonic frequencies should be scaled to account for the anharmonic effect. As the scaling factor for M062X/6-311++G\*\* is unclear while that for B3LYP/6-31G\*\* is well documented, the four conformers are re-optimized at the B3LYP/6-31G\*\* level and the resulting frequencies are scaled by a factor of 0.9602 as recommended in literature<sup>7</sup>. Figure 3 shows the comparison of theoretical and experimental IR spectra. Clearly, the theoretical characteristic IR frequencies are in excellent agreement with the experiment, providing a strong support of our conformational assignment.

It is therefore concluded that the experimentally observed structures are resulted from the dynamic process of relaxing the conformations obtained with laser heating to local minima connected with low energy barriers. It is possible that the observed structures do not correspond to the global minimum on the FES of any temperature. In particular, the folded low energy conformers, including the final global minimum, are not detected due to unfavorable entropic effect and the energy barrier prohibited structural conversion. There is no need to invoke the non-radiative deactivation mechanism to explain the missing of the global minimum in the experimental spectra. Besides, the puzzle of finding only four instead of many more conformers is solved by the conversions of conformers connected with low energy barriers.





**Figure 4.** The free energy profile for four groups of WG conformers: (a) folded backbone conformer, (b) extended backbone conformers with  $g+/+$  side chain, (c) extended backbone conformers with  $g+/-$  side chains, (d) extended backbone conformers with  $g-/+$  side chains.

**Tryptophane-glycine.** Similar to YG, the relative electronic energies, H-bonds, structural types and equilibrium distributions of WG conformers of interest are determined. Details are referred to Table 1S of the Supplementary Information (SI). The structures of representative conformers are displayed in Fig. 1S of SI. Like the case for YG, the global energy minimum of WG, wg1, adopts a folded backbone with a structural type of  $A1-\gamma_D(F)-g+/+$ . Similarly, the equilibrium population of WG conformers also shows a strong entropic effect and the extended structures increase their importance with the increase of temperature. Very different from the case of YG, however, wg1 remains to be the most populous conformer at 450 K, with a concentration of 9.9%. Naturally, the chance of detecting the folded conformation of wg1 is rather high.

The potential energy surface of WG is characteristically the same as that of YG regarding the energy barriers for conformer transformations. Figure 4 shows the free energy profiles for four sets of WG conformers that are likely to be observed. After the supersonic jet cooling, the expected concentrations of wg1 (folded backbone), wg8 (extended backbone with  $g+/-$  side chain), wg11 ( $g+/+$ ) and wg31 ( $g-/+$ ) are 9.9%, 10%, 15% and 16%, respectively. The theoretical results of four high population conformers agree well with the double resonance experiment that yields four observed structures<sup>8,10</sup>. The agreement strongly suggests that the folded structure of wg1 is long lived in the conformational ensemble and detected experimentally. That is, even though wg1 exhibits a folded backbone and a backbone-side chain  $NH-\pi$  dispersive interaction, wg1 is not subjected to the non-radiative deactivation mechanism.

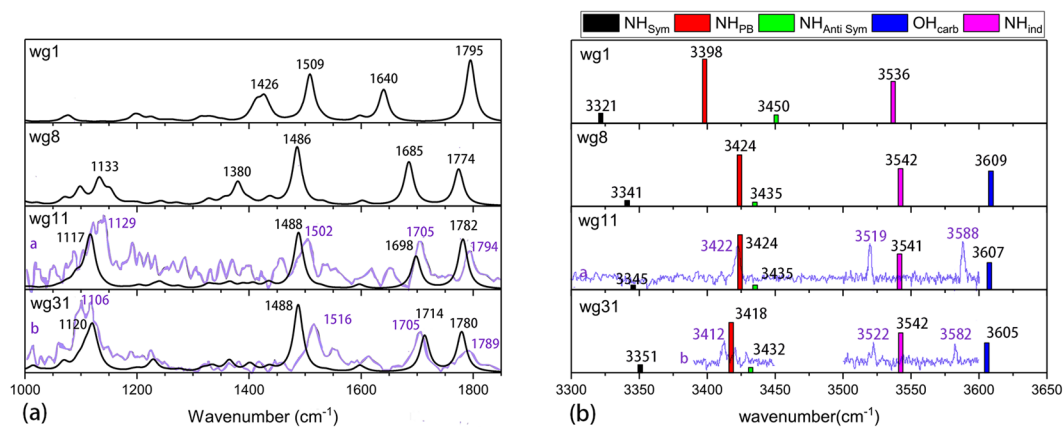
It should be pointed out that the  $g-/-$  conformation, wg43, wg46 and wg48, has a total population of only 5% and may not be observable due to low signal intensity. In comparison, the concentration of the  $g-/-$  conformation of YG, yg32, is as high as 17%. That is, although the overall structural types of YG and WG and the associated transition energy barriers are characteristically the same, the observable structures are different. It is necessary to conduct a detailed computational study dedicated to the interested molecule in order to provide a reliable explanation of the measurement results.

Figure 5 shows the comparison of the experimental and theoretically observable IR spectra of WG conformations. Considering that four WG conformers are identified in the R2PI and UV-UV hole-burning experiments<sup>10</sup>, the IR spectra of wg1 and wg8 missed in the experiment are also shown as they may be useful for comparison with the future experiments. Notice, however, the IR spectrum of wg8 in the high frequency region has in fact been observed and the  $NH_{pb}$ ,  $NH_{ind}$  and OH frequencies are 3407, 3522 and 3598  $cm^{-1}$ , respectively<sup>8,9</sup>. As seen in Fig. 5, the agreement between the theoretical results and the available experimental data is quite satisfactory. The average and maximal deviations for seven characteristic vibrations are 14 and 28  $cm^{-1}$ , respectively. The comparison is based on a universal scaling factor, while the errors are smaller than the literature ones using vibration mode adjusted scaling factors<sup>8,9</sup>.

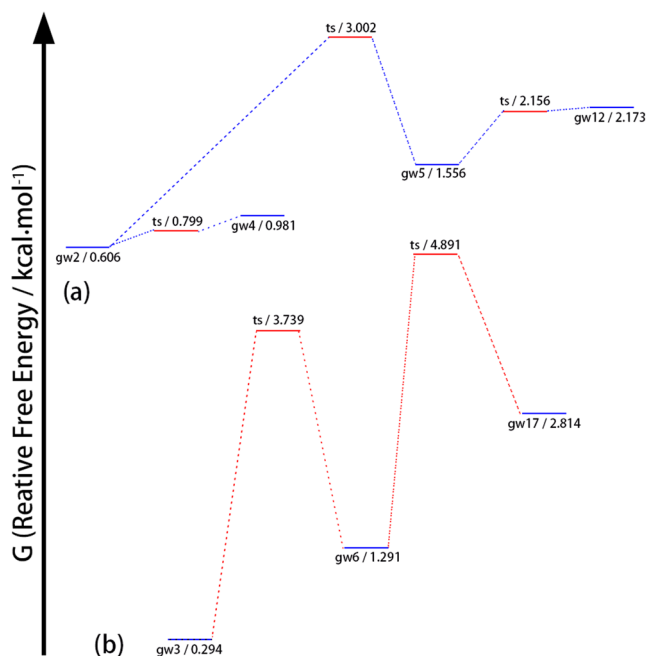
Based on the comparison shown in Fig. 5, the experimentally observed WG conformers *a*, *b*, *c* and *d* can be assigned respectively to wg11, wg31, wg1 and wg8 here<sup>8,10</sup>.

**Glycine-tryptophane.** Table 2S of SI shows the relative electronic energies, H-bonds, structural types and equilibrium distributions of GW conformers of interest. The structures of representative conformers are displayed in Fig. 2S of SI. Once again, the global energy minimum of GW, gw1, adopts a folded backbone configuration. However, gw1 has a structural type of  $A2-\gamma_D(F)-g-/-$ , instead of  $A1-\gamma_D(F)-g+/+$  for both yg1 and wg1. Besides, the extended conformer of gw3 has a total energy very close to that of gw1. Except for very low temperature, gw3 is the global free energy minimum due to favorable entropic effect.

Detailed analysis of the GW free energy profile has been carried out. It is noticed that the characterization about the energy barriers for conformer conversion discussed above, though remains overall correct, requires some minor revision. The free energy profiles for two sets of WG conformers of interest are illustrated in Fig. 6. On one hand, the energy barrier of converting gw5 with a  $g-/+$  side chain to gw2 with a  $g+/+$  side chain is only around 1.5 kcal/mol, as shown in Fig. 2S. The reduced energy barrier may be attributed to the presence



**Figure 5.** Comparison of the experimental (curves and numbers in purple) and theoretical (black curves and colored bars) observable IR spectra of WG conformations: (a) Frequency in the range of  $1000\text{ cm}^{-1}$  to  $1850\text{ cm}^{-1}$  (the theoretical curves are Lorentzians with the full width at half maximum of  $20\text{ cm}^{-1}$ ); (b) Frequency in the range of  $3300\text{ cm}^{-1}$  to  $3700\text{ cm}^{-1}$ . The experimental results are taken from refs<sup>8,10</sup>.

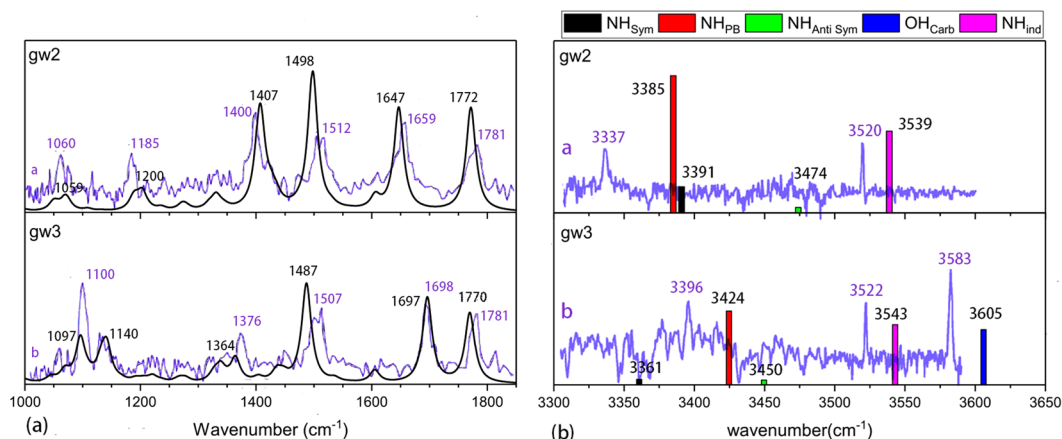


**Figure 6.** The free energy profile for two groups of GW conformers: (a) folded backbone conformers, (b) extended backbone conformers with  $g+/+$  side chain.

of the  $N_{PB}H \cdots \pi$  H-bond. As a result of the reduced energy barrier, gw5 and gw12 may convert to gw2 in the jet cooling process. Based on the data shown in Table 2S, the final concentration of gw2 may reach above 20%. On the other hand, the energy barrier of converting gw6 to gw3, both with a  $g+/+$  side chain, is increased to about 2.5 kcal/mol by the steric hindrance of the side chain. The conversion of gw6 and gw17 to gw3 may be limited. Nevertheless, gw3 has a high population at 18% by itself.

Considering the data shown in Table 2S, the conformational population for any one structural type other than that of gw2 and gw3 is less than 6%. Conformations with such low concentrations may be difficult to detect experimentally. That is, only gw2 and gw3, each with a concentration above 18%, can be easily observed. The conclusion is in good accord with the experimental findings<sup>8,10</sup>. Here again, the missing of gw1 is attributed to its low concentration in the conformational ensemble, instead of resulting from the non-radiative deactivation process. Besides, gw2 is predicted here to be observable even though its structural feature appears to be susceptible to the non-radiative deactivation mechanism<sup>11,20</sup>.

Figure 7 shows the comparison of the experimental and theoretically IR spectra of GW conformations. The agreement between the theory and experiment is quite good for the mid-frequency region. The difference is less than  $10\text{ cm}^{-1}$  on average, while the biggest deviation is only  $20\text{ cm}^{-1}$ . The difference between the theory and



**Figure 7.** Comparison of the experimental and theoretical IR spectra of GW conformations: (a) mid-frequency region, (b) high-frequency region. The experimental results are taken from refs<sup>8,10</sup>. The theoretical results are obtained with B3LYP/6-31G\*\* and scaled by a factor of 0.9602.

experiment for the high-frequency region is much larger, however. On average, the difference is  $28\text{ cm}^{-1}$ . The largest deviation,  $48\text{ cm}^{-1}$ , is found for the N-H<sub>PB</sub> vibration of gw2.

The deviation of  $48\text{ cm}^{-1}$  appears somewhat large and causes some concern on the conformational assignment. The IR spectra of other probable conformers are computed. However, none of them is capable of producing an overall improved agreement with the experiment for the whole frequency range. A possible explanation is that, as known in the literatures<sup>25,26</sup>, the anharmonic effects for different vibrational modes are different and may be accounted for only through different scaling factors<sup>25,26</sup>. To explore other possible explanation, it is noticed that the backbone-side chain interactions in gw2 and gw3 (Fig. 2S) appear to be much stronger than that in yg14, yg17, yg31 and yg32 (Fig. 2) and in wg11 and wg31 (Fig. 1S). It is known that B3LYP is inadequate for describing strong H- $\pi$  interaction, while the result by M062X is much more satisfactory<sup>27,28</sup>. Therefore, the IR spectra of gw2 and gw3 are recomputed with M062X/6-31G\*\*. By requiring that M062X/6-31G\*\* reproduces the average frequency for the six mid-frequency modes of gw2 identified experimentally, a scaling factor of 0.943 is deduced. With this scaling factor, the NH<sub>PB</sub> and NH<sub>ind</sub> frequencies of gw2 are found to be  $3342$  and  $3504\text{ cm}^{-1}$ , in comparison with the experimental values of  $3337$  and  $3520\text{ cm}^{-1}$  (Fig. 7b), respectively. The theoretical NH<sub>PB</sub>, NH<sub>ind</sub> and OH<sub>carb</sub> frequencies of gw3 are  $3382$ ,  $3515$  and  $3610\text{ cm}^{-1}$ , while the experimental values are  $3396$ ,  $3522$  and  $3583\text{ cm}^{-1}$ , respectively. The average and largest differences between the theory and experiment for the high-frequency region are then reduced to  $14$  and  $27\text{ cm}^{-1}$ , respectively. Clearly, the agreement between the experiment and the theory is much improved by describing the H- $\pi$  dispersion interaction with the M062X functional. Assigning the experimentally observed GW conformers *a* and *b* to gw2 and gw3<sup>8,10</sup>, respectively, is therefore validated.

## Conclusions

A computational framework for interpreting the results of supersonic jet cooling based single molecular spectroscopy is proposed. The method consists of the following steps: 1) a thorough conformational search to locate all low energy structures, 2) evaluating the equilibrium distributions of conformations at about  $450\text{ K}$ , 3) classifying the structural features of low energy conformations, 4) determining the transition energy barriers for conformation conversions, 5) adding the populations of conformers with low energy barriers together to determine the contents of the relevant stable conformations after the supersonic expansion cooling. Highly populated conformations thus determined are the structures observed in the spectroscopic measurement.

As a rule of thumb, an energy barrier is deemed low if it is less than  $2\text{ kcal/mol}$ . A conformer population is high and observable if it is above  $8\%$ . Moreover, a high energy barrier is expected for the conversion of conformers with different folded backbones or with extended and folded backbones. With inconsequential exceptions, a low energy barrier is found only for conformers with the same side-chain structural type. The empirical rules about the energy barriers can be used to avoid the computation extensive transition state calculations.

Each of the proposed computational steps has been considered before, but they have not been combined together yet in the interpretation of the single molecular spectroscopy. Here, the proposed computational approach as a whole is applied to the studies of the spectra of gaseous dipeptides YG, WG and GW. The results agree well with the experiments on the numbers of observable YG, WG and GW conformations. The theoretical conformation assignments are further validated by the available experimental IR data. The previously conceived discrepancies between theory and experiment are all resolved consistently. It is concluded that the computational approach provides a natural way for the understanding of single molecular spectroscopy experiments. The recently proposed non-radiative deactivation mechanism, however, appears to be incoherent when the three cases of YG, WG and GW are considered together.



**Computational methods.** The low-energy structures of YG, WG and GW were determined by optimizing a large set of trial structures. The trial structures were generated by considering all possible combinations of the bond rotational degrees of freedom that are illustrated in Fig. 3S of SI. Such a thorough conformational search can in principle provide a complete set of all structures<sup>29–32</sup>. The trial structures were optimized at the M062X/6-31+G\*\* level of theory. Vibrational frequencies were also computed at the M062X/6-31+G\*\* level for all conformers that are within a range of 8 kcal/mol from the global minimum. All the conformers were verified to be the true local minima by the frequency analysis. The single-point energy calculations were carried out using the functional DSD-PBEP86-D3BJ with the basis set of aug-cc-pVTZ. The functional M062X is used here as it is capable of describing H-bonds well at a reasonable calculation cost<sup>20,24,33</sup>. The functional DSD-PBEP86-D3BJ is known to describe H-bond systems with accuracy close to that of CCSD(T) and is used here to provide high quality conformational energies<sup>34,35</sup>.

Based on the DSD-PBEP86-D3BJ/aug-cc-pVTZ energies and the free energy corrections obtained at the M062X/6-31+G\*\* level, the equilibrium Boltzmann distributions of conformers at different temperatures were computed. Conformers of interest were then determined. A conformer of interest here means that it is one of the ten lowest electronic energy conformers, or that its equilibrium concentration is over 1% for some temperature under 450 K.

The structural characteristics of the conformers of interest were analyzed and classified into structural types based on their hydrogen bond (H-bond) features and secondary structures according to the method of Csaszar and Perczel<sup>36</sup>. The transition states between conformers of the same structural type and between representative conformers of different structural types were determined at the DSD-PBEP86-D3BJ/aug-cc-pVTZ//M062X/6-311++G\*\* level using the Berny algorithm<sup>37</sup>.

Computations of the DSD-PBEP86-D3BJ/aug-cc-pVTZ energy were carried out using the ORCA 4.0 software<sup>38</sup>. All the other calculations were performed with the GAUSSIAN 09 suite of programs<sup>39</sup>.

## References

- Oomens, J. *et al.* Charge-state resolved mid-infrared spectroscopy of a gas-phase protein. *Phys Chem Chem Phys* **7**, 1345–1348, <https://doi.org/10.1039/b502322j> (2005).
- Hunig, I., Seefeld, K. A. & Kleineremanns, K. REMPI and UV-UV double resonance spectroscopy of tryptophan ethylester and the dipeptides tryptophan-serine, glycine-tryptophan and proline-tryptophan. *Chem Phys Lett* **369**, 173–179, [https://doi.org/10.1016/S0009-2614\(02\)01966-8](https://doi.org/10.1016/S0009-2614(02)01966-8) (2003).
- Abo-Riziq, A., Crews, B. O., Callahan, M. P., Grace, L. & de Vries, M. S. Spectroscopy of isolated gramicidin peptides. *Angew Chem Int Edit* **45**, 5166–5169, <https://doi.org/10.1002/anie.200601516> (2006).
- Alonso, J. L. & Lopez, J. C. In *Top Curr Chem Vol. 364 Topics in Current Chemistry-Series* (eds A. M. Rijs & J. Oomens) 335–401 (Springer-Verlag Berlin, 2015).
- Cabezas, C., Varela, M. & Alonso, J. L. The Structure of the Elusive Simplest Dipeptide Gly-Gly. *Angew Chem Int Edit* **56**, 6420–6425, <https://doi.org/10.1002/anie.201702425> (2017).
- Rijs, A. M. & Oomens, J. In *Top Curr Chem Vol. 364 Topics in Current Chemistry* (eds A. M. Rijs & J. Oomens) 1–42 (Springer-Verlag Berlin, 2015).
- Abo-Riziq, A. *et al.* Conformational Structure of Tyrosine, Tyrosyl-glycine, and Tyrosyl-glycyl-glycine by Double Resonance Spectroscopy. *J Phys Chem A* **115**, 6077–6087, <https://doi.org/10.1021/jp110601w> (2011).
- Bakker, J. M. *et al.* Folding structures of isolated peptides as revealed by gas-phase mid-infrared spectroscopy. *Chemphyschem* **6**, 120–128, <https://doi.org/10.1002/cphc.200400345> (2005).
- Valdes, H., Reha, D. & Hobza, P. Structure of isolated tryptophyl-glycine dipeptide and tryptophyl-glycyl-glycine tripeptide: Ab initio SCC-DFTB-D molecular dynamics simulations and high-level correlated ab initio quantum chemical calculations. *J Phys Chem B* **110**, 6385–6396, <https://doi.org/10.1021/jp057425y> (2006).
- Hunig, I. & Kleineremanns, K. Conformers of the peptides glycine-tryptophan, tryptophan-glycine and tryptophan-glycine-glycine as revealed by double resonance laser spectroscopy. *Phys Chem Chem Phys* **6**, 2650–2658, <https://doi.org/10.1039/b316295h> (2004).
- Domcke, W. & Sobolewski, A. L. Peptide Deactivation Spectroscopy meets theory. *Nat Chem* **5**, 257–258, <https://doi.org/10.1038/nchem.1601> (2013).
- Loquais, Y. *et al.* Secondary Structures in Phe-Containing Isolated Dipeptide Chains: Laser Spectroscopy vs Quantum Chemistry. *J Phys Chem A* **119**, 5932–5941, <https://doi.org/10.1021/jp509494c> (2015).
- Kistler, K. A. & Matsika, S. Cytosine in context: A theoretical study of substituent effects on the excitation energies of 2-pyrimidinone derivatives. *J Phys Chem A* **111**, 8708–8716, <https://doi.org/10.1021/jp074361d> (2007).
- Kistler, K. A. & Matsika, S. Radiationless decay mechanism of cytosine: An ab initio study with comparisons to the fluorescent analogue 5-methyl-2-pyrimidinone. *J Phys Chem A* **111**, 2650–2661, <https://doi.org/10.1021/jp066366i> (2007).
- Marian, C. M. The guanine tautomer puzzle: Quantum chemical investigation of ground and excited states. *J Phys Chem A* **111**, 1545–1553, <https://doi.org/10.1021/jp068620v> (2007).
- Perun, S., Sobolewski, A. L. & Domcke, W. Photostability of 9H-adenine: mechanisms of the radiationless deactivation of the lowest excited singlet states. *Chem Phys* **313**, 107–112, <https://doi.org/10.1016/j.chemphys.2005.01.005> (2005).
- Sobolewski, A. L., Domcke, W. & Hattig, C. Tautomeric selectivity of the excited-state lifetime of guanine/cytosine base pairs: The role of electron-driven proton-transfer processes. *P Natl Acad Sci USA* **102**, 17903–17906, <https://doi.org/10.1073/pnas.0504087102> (2005).
- Shemesh, D., Hattig, C. & Domcke, W. Photophysics of the Trp-Gly dipeptide: Role of electron and proton transfer processes for efficient excited-state deactivation. *Chem Phys Lett* **482**, 38–43, <https://doi.org/10.1016/j.cplett.2009.09.087> (2009).
- Shemesh, D., Sobolewski, A. L. & Domcke, W. Efficient Excited-State Deactivation of the Gly-Phe-Ala Tripeptide via an Electron-Driven Proton-Transfer Process. *J Am Chem Soc* **131**, 1374–+, <https://doi.org/10.1021/ja0808485b> (2009).
- Gloaguen, E. *et al.* Gas-Phase Folding of a Two-Residue Model Peptide Chain: On the Importance of an Interplay between Experiment and Theory. *J Am Chem Soc* **132**, 11860–11863, <https://doi.org/10.1021/ja103996q> (2010).
- Holroyd, L. F. & van Mourik, T. Tyrosine-glycine revisited: Resolving the discrepancy between theory and experiment. *Chem Phys Lett* **621**, 124–128, <https://doi.org/10.1016/j.cplett.2014.12.055> (2015).
- Valdes, H. *et al.* Potential-energy and free-energy surfaces of glycyl-phenylalanyl-alanine (GFA) tripeptide: Experiment and theory. *Chem-Eur J* **14**, 4886–4898, <https://doi.org/10.1002/chem.200800085> (2008).
- Handschuh, M., Nettesheim, S. & Zenobi, R. Is infrared laser-induced desorption a thermal process? The case of aniline. *J Phys Chem B* **103**, 1719–1726, <https://doi.org/10.1021/jp984126i> (1999).
- Gloaguen, E. & Mons, M. Isolated Neutral Peptides. *Gas-Phase Ir Spectroscopy and Structure of Biological Molecules* **364**, 225–270 (2015).

25. Miller, Y. *et al.* Vibrational spectroscopy of (SO<sub>4</sub>(2-))<sub>n</sub> center dot (H<sub>2</sub>O)<sub>n</sub> clusters, n = 1–5: Harmonic and anharmonic calculations and experiment. *J Chem Phys* **127**, doi:Artn 09430510.1063/1.2764074 (2007).
26. Biswal, H. S., Loquais, Y., Tardivel, B., Gloaguen, E. & Mons, M. Isolated Monohydrates of a Model Peptide Chain: Effect of a First Water Molecule on the Secondary Structure of a Capped Phenylalanine. *J Am Chem Soc* **133**, 3931–3942, <https://doi.org/10.1021/ja108643p> (2011).
27. Roy, D., Marianski, M., Maitra, N. T. & Dannenberg, J. J. Comparison of some dispersion-corrected and traditional functionals with CCSD(T) and MP2 ab initio methods: Dispersion, induction, and basis set superposition error. *J Chem Phys* **137**, doi:Artn 13410910.1063/1.4755990 (2012).
28. Rodriguez-Sanz, A. A., Cabaleiro-Lago, E. M. & Rodriguez-Otero, J. Cation center dot center dot center dot pi interaction and microhydration effects in complexes formed by pyrrolidinium cation and aromatic species in amino acid side chains. *Org Biomol Chem* **12**, 2938–2949, <https://doi.org/10.1039/c3ob42388c> (2014).
29. Ling, S. L. *et al.* Gaseous arginine conformers and their unique intramolecular interactions. *J Phys Chem A* **110**, 12282–12291, <https://doi.org/10.1021/jp0645115> (2006).
30. Yang, B. & Lin, Z. J. Systematic search of conformations of five tetrapeptides and a divide and conquer strategy for the predictions of peptide structures. *Comput Theor Chem* **1108**, 40–49, <https://doi.org/10.1016/j.comptc.2017.03.014> (2017).
31. Ru, X., Song, C. & Lin, Z. J. Structural Information-Based Method for the Efficient and Reliable Prediction of Oligopeptide Conformations. *J Phys Chem B* **121**, 2525–2533, <https://doi.org/10.1021/acs.jpcc.6b12415> (2017).
32. Chen, X. & Lin, Z. J. Performances of Five Representative Force Fields on Gaseous Amino Acids with Different Termini. *Chinese J Chem Phys* **29**, 179–186, <https://doi.org/10.1063/1674-0068/29/cjcp1507153> (2016).
33. Gloaguen, E. *et al.* Experimental and Theoretical Investigation of the Aromatic–Aromatic Interaction in Isolated Capped Dipeptides. *Journal of Physical Chemistry A* **114**, 2973–2982 (2010).
34. Grimme, S. Density functional theory with London dispersion corrections. *Wires Comput Mol Sci* **1**, 211–228, <https://doi.org/10.1002/wcms.30> (2011).
35. Kang, Y. K. & Park, H. S. Assessment of CCSD(T), MP2, DFT-D, CBS-QB3, and G4(MP2) methods for conformational study of alanine and proline dipeptides. *Chem Phys Lett* **600**, 112–117, <https://doi.org/10.1016/j.cplett.2014.03.067> (2014).
36. Csaszar, A. G. & Perczel, A. Ab initio characterization of building units in peptides and proteins. *Prog Biophys Mol Bio* **71**, 243–309, [https://doi.org/10.1016/S0079-6107\(98\)00031-5](https://doi.org/10.1016/S0079-6107(98)00031-5) (1999).
37. Simons, J., Jorgensen, P., Taylor, H. & Ozment, J. Walking on Potential-Energy Surfaces. *J Phys Chem-Us* **87**, 2745–2753, <https://doi.org/10.1021/j100238a013> (1983).
38. Neese, F. The ORCA program system. *Wires Comput Mol Sci* **2**, 73–78, <https://doi.org/10.1002/wcms.81> (2012).
39. Frisch, M. J. *et al.* Gaussian 09, Revision A. [http://www.gaussian.com/g\\_tech/g\\_ur/m\\_citation.htm](http://www.gaussian.com/g_tech/g_ur/m_citation.htm) (Gaussian, 2009).

## Acknowledgements

Z.L. thanks the financial supports of the National Natural Science Foundation of China (11374272, 11574284 & 11774324) and Collaborative Innovation Center of Suzhou Nano Science and Technology and the computing time from the Supercomputing Center of the University of Science and Technology of China.

## Author Contributions

Z.L. is responsible for the project direction and planning. B.Y. performed the first-principles calculations and analyzed the data with help from S.L. All were involved in the discussion of results. B.Y. and Z.L. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-16234-3>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017