PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

PCCP

ARTICLE

Theoretical spectroscopic studies on chemical and electronic structures of arginylglycine

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Hongbao Li,[a] Leilei Li,[b] Jun Jiang,[b] Zijing Lin[b] and Yi Luo*[a, b]

The energy differences between canonical and zwitterionic isomers of arginylglycine (ArgGly) at CCSD/aug-cc-pVDZ level are too small (less than 1 kcal/mol) to determine the dominant form in gas phase from the energetic point of view. First-principles simulations have been performed for near-edge X-ray absorption fine-structure (NEXAFS) spectra and X-ray photoelectron spectra (XPS) at C, N and O K-edges, as well as for infrared (IR) spectra of neutral ArgGly. Noticeable spectral differences were found which enables unambiguous identifications for different neutral groups. We thus demonstrate the X-ray spectroscopy as a powerful technique to study the conformation dependent chemical and electronic properties of neutral ArgGly.

1 Introduction

Zwitterions with two oppositely charged centres are believed to play important roles in determining the structure and function of crucial biological systems including peptides and proteins.[1,2] However, such charge-separated compounds can only stably exist in solid state or in solution but within a certain range of pH, while is normally hard to be stabilized in gas phase environment. It is found that these optimized structures in gas phase can occasionally be a reasonable alternative to mimic the structures in the continuum medium.[3] Therefore, whether zwitterions can be predominantly preferred in gas phase has been extensively discussed.[4-9] For most natural amino acids, there are many experimental and theoretical studies suggesting that the canonical form is always dominant in the gas phase.[10-12] The zwitterions, as local minima on the potential energy surface (PES), only theoretically exist for arginine and histidine.[13-19] These charged structures do not have effective conformational distributions due to the relatively high Gibbs free energies. However, it might be reasonable to image that the zwitterions may exist as the most stable conformations in the oligopeptide due to the strong interactions between the charged and polar groups in the molecules.

Recently, a theoretical study on the PES of dipeptide arginylglycine (ArgGly) in the gas phase predicted that this molecule is perhaps the smallest peptide with a zwitterion as the global minimum.[20,21] Based on the molecular mechanics force-field study, Prell et al.[20] found a zwitterionic structure of ArgGly, which is energetically very close to the global minimum of the canonical form. Wang et al.[21] performed a more systematic theoretical work. They identified that the global minimum of ArgGly is actually in the zwitterionic form, which is more stable than the most stable canonical conformer. However, those conclusions are not so compelling because of the low-accuracy calculations on the PES, and should be verified by more accurate simulations using high-level computational methods. On the other hand, it is well-known that the chemical and electronic structures of biomolecules can be well identified by various spectroscopic features, which are available for both theoretical computing and experimental measuring. In this context, it would be a good idea to distinguish the neutral isomers of ArgGly by their spectroscopic differences.

The widely used infrared (IR) spectroscopy provides important information on the chemical structure of the molecule due to the sensitivity of vibrational frequencies to the electronic density in chemical bonds, especially the bonds that involves N and O atoms.[22] In recent years, with the great development in the X-ray instrumentation, soft X-ray techniques, such as X-ray photoelectron spectra (XPS) and near-edge X-ray absorption fine-structure (NEXAFS) spectra have become powerful tools to probe the electronic structures of molecules with fundamental biological importance owing to their element-selection and sensitivity to the local chemical environment.[23-32] These effective techniques have frequently been used to resolve questions that had been under debate for years. For molecules like arginine and arginylglycine with several different neutral forms, one could expect to distinguish these isomers by applying such spectroscopic methods.

In this work, based on the structure information obtained from previous study,[21] we calculated the electronic energy differences of isomers at advanced CCSD level, so as to identify

[a] Guizhou Provincial Key Laboratory of Computational Nano-Material Science and Guizhou Synergetic Innovation Center of Scientific Big Data for Advanced Manufacturing Technology, Guizhou Normal College, Guiyang, Guizhou, 550018, China.
[b] Hefei National Laboratory for Physical Sciences at the Microscale, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, Anhui, 230026, China.

the dominant form of ArgGly. It is found that the global minimum of ArgGly is still in its canonical form, rather than the previous predicted zwitterionic form. The energy difference between them is too small to be used to determine the dominant form. Based on the computed spectra (XPS and NEXAFS) of the most populated conformers of ArgGly, we found that the canonical and zwitterionic forms of this molecule can be unambiguously distinguished, which provides useful information for future experimental identification.

This article is organized as follows: in Section 2, we describe the computational methods for the high level calculations and for the IR, NEXAFS and XPS spectra of ArgGly in the gas phase. Section 3 presents the structural and spectral results and gives a discussion on the structure-property relationships. Finally, concluding remarks are given in Section 4.

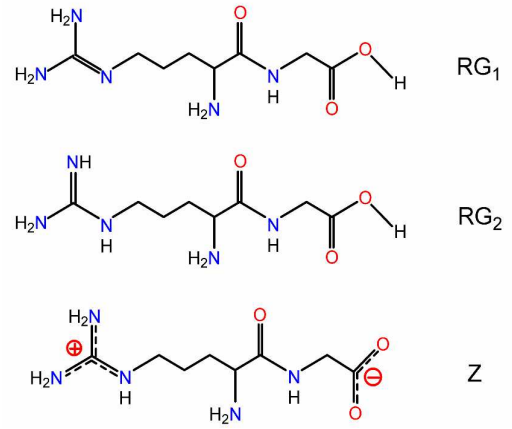

Fig. 1 Schematic structures of the dipeptide arginylglycine in two canonical forms (RG_1 and RG_2) and one zwitterionic form (Z) with heavy atoms labelled.

Table 1 Conformers with low Gibbs free energy in the three neutral arginylglycine groups, together with their corresponding structural names in Ref. 21 and their respective percent shares at 298K.

RG_1	Ref.[a]	Percent[b] (%)	RG_2	Ref.[a]	Percent[b] (%)	Z	Ref.[a]	Percent[b] (%)
RG_{1a}	c11	81.95	RG_{2a}	c7	80.31	Z_1	z1	51.81
RG_{1b}	c44	15.46	RG_{2b}	c24	7.55	Z_2	z2	13.03
RG_{1c}	c39	2.59	RG_{2c}	c16	3.61	Z_3	z13	7.42
Sum		100	RG_{2d}	c15	3.46	Z_4	z20	5.73
			RG_{2e}	c22	2.88	Z_5	z10	3.78
			RG_{2f}	c6	2.19	Z_6	z5	3.71
			Sum		100	Sum		85.49
RG_1	Ref.[a]	Percent[c] (%)	RG_2	Ref.[a]	Percent[c] (%)	Z	Ref.[a]	Percent[c] (%)
RG_{1a}'	c11	94.29	RG_{2a}'	c7	81.22	Z_1'	z1	58.44
RG_{1b}'	c44	3.41	RG_{2b}'	c22	5.45	Z_2'	z2	24.37
RG_{1c}'	c39	2.30	RG_{2c}'	c15	4.74	Z_3'	z13	3.35
Sum		100	RG_{2d}'	c24	3.99	Z_4'	z3	2.94
			RG_{2e}'	c6	2.57	Z_5'	z5	2.76
			RG_{2f}'	c16	2.03	Sum		91.85
			Sum		100			

[a] The corresponding names were taken from Ref. 21.

[b, c] The percent shares based on the CCSD/cc-pVDZ electronic energies and Gibbs free energy corrections at (b) BHandHLYP/6-311++G(d, p) and (c) M062X/6-311++G(d, p) levels.

2 Computational methods

As illustrated in Figure 1, the molecule ArgGly has three different neutral forms: two canonical groups (RG_1 and RG_2) and one zwitterionic group (Z) in the gas phase. Due to the presence of the strongly basic guanidine group, the proton is shifted from the C terminal to the special guanidine group, rather than to the amino group of arginine residue.[13, 16-19, 33] Although the PES of ArgGly was previously explored by the systematic search method,[21] their electronic energies have only been determined at the relatively simple DFT-B3LYP and MP2 levels. Previous studies on the relative energies of the most stable arginine conformers[13] indicated that, comparing with the more accurate CCSD results, the energy ordering at the B3LYP level was misleading, while the relative energies at the MP2 level were found to be overestimated for the canonical conformers and underestimated for the zwitterionic ones. Most likely, those computations are not accurate enough to support the conclusion that the zwitterion is the global minimum of ArgGly. In this study, the low-energy conformers were classified into three groups and optimized at the DFT/BHandHLYP level[34-36] with the basis set of 6-311++G (d, p). The electronic energies were finally calculated at the CCSD/cc-pVDZ level (the four most important conformers were calculated at even more expensive CCSD/aug-cc-pVDZ level, with 521 basis functions considered), which is advanced in offering high accuracy.[17,37,38]

As shown in Table 1, all the low energy conformers were taken from Ref. 21. The conformational distributions at 298K are calculated based on the Boltzman distribution form in the same way we did previously [39,40] by considering the Gibbs free energy correction. The transition states for the transformation from the most populated zwitterion (Z_1) to the two canonical groups (RG_{1a} and RG_{2a}), as well as between the two canonical groups were identified at the BHandHLYP/6-311++G (d, p) level. The nature of the stationary and saddle points were verified by Hessian calculations. The electronic energies of the transition states were finally determined at the CCSD/cc-pVDZ level while the Gibbs free energy correction at 298 K for the reaction barriers has been taken into account. The low-energy conformers and their frequencies were also calculated at the M062X/6-311++G (d, p) levels to better describe the hydrogen bonding and dispersion interactions.[41,42] For the M062X calculations, an "untrafine" numerical integration grid has been used to ensure the reliable results for systems with noncovalent interactions. The calculated vibrational frequencies of the conformers are all scaled by uniform factors of 0.926 and 0.954 for the BHandHLYP and M062X methods. All the above calculations were carried out using the Gaussian 09 software package.[43]

All the most populated ArgGly conformers at 498 K (RG_{1a}, RG_{2a}-RG_{2e} and Z_1) were selected for the X-ray spectra calculations since the related experiments are always measured around this temperature.[24,28] The C1s, N1s and O1s XPS and NEXAFS spectra of these molecules are calculated by using the StoBe program[44] at the DFT level with the gradient-correlation Becke (BE88) exchange[45] and Perdew (PD86)

correlation functionals.[46] The IGLO-III basis set[47] is set for the excited atom, triple-ζ plus valence polarization (TZVP) basis set is employed for the others, and miscellaneous auxiliary basis sets are also set for all atoms. To facilitate the convergence of the core-hole state, for those unexcited atoms that are of the same element as the excited one, their 1s electrons are modeled by model core potentials (MCP). The XPS spectra are obtained by a Lorentzian convolution of the ionization potentials (IPs) with a FWHM (full width at half maximum) of 0.1 eV. The NEXAFS spectra are calculated with the full core hole (FCH) approximation. All calculated spectra are calibrated by aligning the 1s→LUMO (lowest unoccupied MO) transition to the one obtained from a ΔKohn-Sham (ΔKS) approach.[19] Relativistic effects of +0.2, +0.3 and +0.4 eV for the C, N and O edges, respectively, are used to produce the overall shifts of the spectra.[48] It is noticed that the C1s→$\pi^*_{C=O}$, N1s→π^*_{NC} and O1s→$\pi^*_{C=O}$ transitions of neutral amino acids are normally found at 288.4, 402.3, and 532.2 eV, respectively, in many gas-phase NEXAFS experiments.[49-51] In order to be more relevant to the future experiments, we have further shifted our calculated spectra to align with the first experimental π^* peaks, which correspond to +0.85, +0.19 and +0.16 eV for C, N and O edges, respectively. Such a shift is applied purely for practical purpose with no particular physical meanings. Stick NEXAFS spectra are convoluted with the Lorentzian function with FWHMs of 0.1 and 0.7 eV (below and above the IPs). The total spectra of the neutral ArgGly (the "SUM") are obtained as a summation weighted by their relative abundance.

3 Results and discussion

3.1 Conformations: Structure and Energy

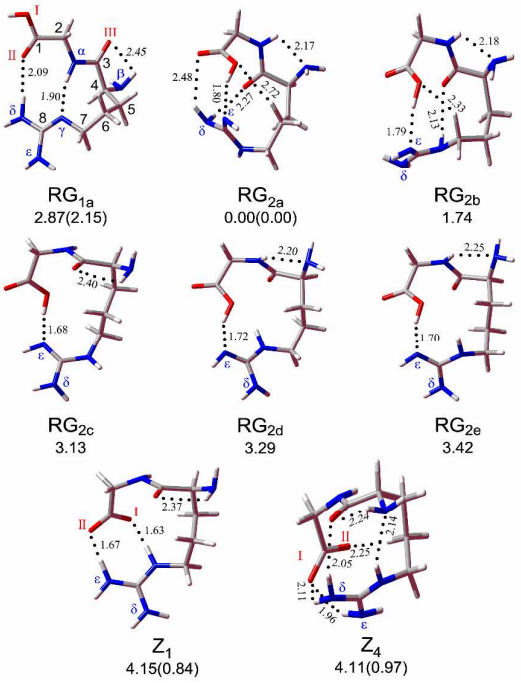

Fig. 2 Structures of the most populated conformers of the three neutral groups of ArgGly and their relative energies based on temperature-independent electronic energies (in kcal/mol) calculated at CCSD/cc-pVDZ level with zero-energy correction from BHandHLYP/6-311++G (d, p). The relative energies in parentheses are determined by further CCSD/aug-cc-pVDZ calculation results.

The geometric structures and electronic energies of various conformers of ArgGly in the three neutral groups were depicted in Figure 2, where different intramolecular hydrogen bonds (HBs) are illustrated by dotted lines. Similar to arginine, three or four HBs exist in most of these low-energy conformers. The interactions between the carboxylate group and the side chain guanidine group make most structures folded spherically. It should be noted that the canonical form RG_2 with a NH group in the guanidine group is more stable than RG_1 with two NH_2 groups, which is opposite to the canonical arginine structures.[13]

Previous calculations[21] at the BHandHLYP (or MP2) level with the 6-311++G (2df, 2p) basis set have found that the most stable zwitterionic conformer is 1.70 (or 1.53) kcal/mol lower in energy than that of the canonical one. However, from our CCSD/cc-pVDZ calculation with higher accuracy, the canonical conformer RG_{2a} is actually the global minimum on the PES. RG_{2a} is more stable than the two low-energy zwitterionic structures (Z_1 and Z_4) by about 4.10 kcal/mol, casting doubts on the previous proposal of ArgGly being the smallest peptide with a zwitterion as the global minimum.[21]

To collecting more convincing evidence, we performed a further CCSD/aug-cc-pVDZ calculation for the four conformers (RG_{1a}, RG_{2a}, Z_1 and Z_4). It is found that the global minimum is still the canonical conformer RG_{2a}, although the energy differences between the canonical and zwitterionic form become smaller (less than 1 kcal/mol). It seems that with the improving of computation accuracy, the identification of the dominant form from the energy point of view becomes even more difficult.

A higher level of theory and bigger basis sets were not attempted for this large molecule due to the expensive computational costs (up to 1127 basis functions at the CCSD(T)/aug-cc-pVTZ level). However, the previous CCSD(T) calculations with very large basis sets (up to 1380 functions included) for the five canonical and three zwitterionic arginine conformers provided almost the same relative stabilities as that at a cheaper CCSD/6-31++G(d, p) approach (only 312 basis functions included).[13,52]

Both the measurement of energy in several kcal/mol and probe of molecular geometry of the neutral isomers are extremely difficult in experiment. One can therefore apply spectroscopic techniques. The main difference between the two neutral forms is the charge state distribution in the carboxylate group and the guanidine group. Such information helps to understand the structure-property relationship and especially reveal the spectral differences between the canonical and zwitterionic structures.

3.2 Chemical Structure: Infrared Spectra

Since the three neutral groups of ArgGly coexist in the experiment, the IR spectra for all of the populated conformers listed in Table 1 were calculated. The theoretical IR spectra are illustrated in Figure 3. More than 85% of the total population in each group were included in the averaged theoretical spectra (SUM). The three groups hold several unique IR features, providing a way to distinguish them.

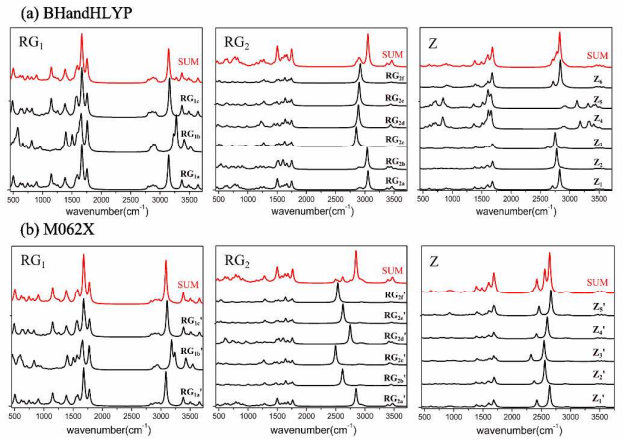

Fig. 3 Simulated IR spectra of dominant conformers in the three neutral groups of arginylglycine RG_1, RG_2 and Z calculated with (a) BHandHLYP and (b) M062X functionals, as well as their summation (SUM) calculated using the percentage listed in Table 1. A Lorentzian profile with the half width at half maximum, 20 cm^{-1}, is used to convolute the calculated spectra.

In the 490-1300 cm^{-1} region, no peaks were observed for the zwitterions, but two strong peaks appear in group RG_1. The peak at 498 cm^{-1} can be assigned to the NH_2 out-of-plane bending mode in the guanidine group, while the peak at 1147 cm^{-1} is mainly contributed from the OH in-the-plane bending mode in the carboxylate group. In fact, the latter one is in good agreement with the IR experimental observation (at ca. 1140 cm^{-1}) of its protonated form ArgGly·H^+.[20] In group RG_2, these two bending modes were suppressed and blue-shifted to ca. 800 and 1250 cm^{-1} due to the formation of HBs.

In the 1300-1800 cm^{-1} region, IR features for different groups are also identified. Both RG_1 and Z show a peak at 1383 cm^{-1}, but with very different origins. For the canonical RG_1, it is mainly from the NCN asymmetric stretching mode in the guanidine group, while it is resulted from the OCO symmetric stretching mode in the deprotonated carboxylate group of zwitterionic Z. The C=O stretching mode in the peptide bond has a much stronger peak at 1668 cm^{-1} for RG_1 group, but it was blue-shifted to 1681 cm^{-1} for group Z because of the increased interactions in the zwitterionic structures. It is interesting to notice that an unique peak at around 1755 cm^{-1} is observed only for both of the two canonical groups, which is contributed from the C=O stretching mode in the carboxylate group. This peak was also clearly detected at ca. 1760 cm^{-1} in the IR experimental spectrum of ArgGly·H^+.[20] Such peak can be used to ambiguously distinguish the canonical forms from the zwitterionic one.

In the 2700-3500 cm^{-1} region, a very strong peak with different energy for each neutral group can be observed. For group RG_1, the peak is at 3148 cm^{-1} and from the NH stretching mode in the peptide bond. But for group RG_2, it appears at 3053 cm^{-1} and contributed from the OH stretching mode in the carboxylate group. For group Z, it is observed at 2829 cm^{-1} and from the NH stretching mode in the guanidine group. These three unique peaks can be used to clearly distinguish the three neutral ArgGly groups.

Generally, the theoretical IR spectra calculated at the M062X level show the same trend as those at the BHandHLYP level. However, due to the different Gibbs free energy corrections, the most populated conformer in each group possesses more percentage at the M062X level, as listed in Table 1.This leads to some changes in the theoretical total spectrum (SUM) (for example, the three peaks at around 2829 cm^{-1} in group Z). Similar to arginine,[53] although the dispersion interactions are included in the M062X functional, it does not introduce effective changes in spectra.

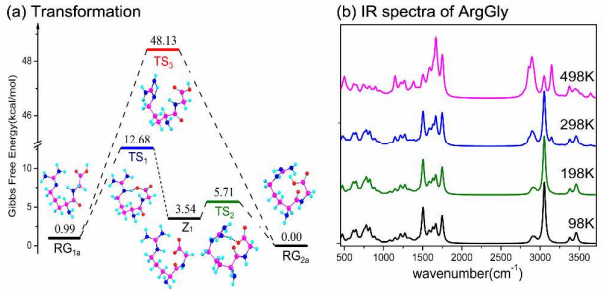

Fig. 4 (a) Transition states for the transformation between conformers with the lowest energy in the three groups at 298 K. TS_1: Z_1 to RG_{1a}; TS_2: Z_1 to RG_{2a}; TS_3: RG_{1a} to RG_{2a}. (b) Simulated IR spectra of the most populated conformers of ArgGly at four different temperatures. A Lorentzian profile with the half width at half maximum, 20 cm^{-1}, is used to convolute the calculated spectra.

The transformation paths among the three different groups are also investigated. The most populated conformer (RG_{1a}, RG_{2a} and Z_1) that was corrected by the BHandHLYP functional from each group (RG_1, RG_2 and Z) is selected. The transformation between Z_1 and RG_{1a}, Z_1 and RG_{2a}, as well as RG_{1a} and RG_{2a} has been carefully examined. Three transition states, labeled as TS_1 TS_2 and TS_3, are identified. The relative Gibbs free energies and the corresponding geometry structures are depicted in Figure 4(a). It can be seen that the calculated energy barrier from Z_1 to the canonical form RG_{1a} is 9.14 kcal/mol at the CCSD/cc-pVDZ level while the energy barrier between Z_1 and RG_{2a} is only 1.17 kcal/mol. It means that the proton in the guanidine group can easily be transferred to the deprotonated carboxylate group (OCO) and the transformation between the canonical and zwitterionic forms is easy. On the other hand, the transformation between the two canonical forms is not feasible due to a large energy barrier (47.14 kcal/mol). Nevertheless, such transformation

can be achieved by using the zwitterionic form as an intermediate state. Therefore, the conformers in the three neutral groups can reach their equilibrium eventually.

The total IR spectra of the neutral ArgGly conformers at four different temperatures are shown in Figure 4(b). Since the zwitterions have negligible distributions, the total spectra are mainly contributed from the two canonical forms. Due to the large population of the canonical group RG_2, the spectra at low temperatures are almost the same as that in Figure 3 (the SUM of RG_2). At high temperatures, especially at 498 K, one can notice great changes in the spectra. In this case, another canonical form RG_1 has almost the same population as that of RG_2. The strong, wild peak around 2900 cm^{-1} is mainly contributed by the CH stretching mode in the backbond part of RG_1 and RG_2, and the NH stretching mode in the guanidine group of Z. It is interesting to note that the peak at 3051 cm^{-1} is actually from the OH stretching mode in the carboxylate group of RG_2, while the one at 3147 cm^{-1} is from the NH stretching mode in the peptide bond part of RG_1. These also enable people to identify the two canonical forms.

3.3 Electronic Structure: XPS and NEXAFS Spectra

The C1s, N1s and O1s XPS spectra of the six canonical and one zwitterionic ArgGly isomers are plotted in Figure 5a-5c, together with the averaged spectrum (SUM) at 498 K. Due to its high energy, the zwitterion Z_1 has almost no contribution to the averaged spectrum, so the SUM can be just viewed as those from the two canonical forms. One can find the notable differences in the XPS spectra of the three neutral groups at all K-edges.

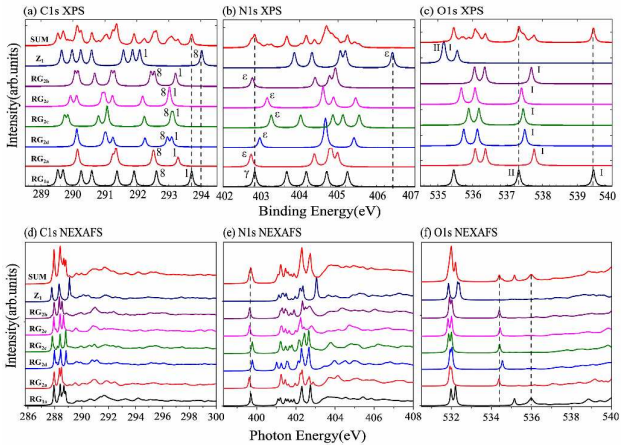

Fig. 5 Calculated (a) C1s, (b) N1s, and (c) O1s XPS spectra and (d) C1s, (e) N1s, and (f) O1s NEXAFS spectra of the lowest-energy conformers of neutral ArgGly as well as the averaged spectra (SUM) at 498 K according to their equilibrium distributions.

In the C edge, the discrepancy in the local chemical environment of C_1 and C_8 caused a wider splitting (1.94 eV) of the core binding energies (BEs) for the zwitterionic isomer than the canonical ones (ca. 0.03-1.10 eV). In the N edge, the double bond C=N involved N_γ (in group RG_1) or N_ε (in group RG_2) leads to a much lower binding energy at around 402.83 eV, which was red-shifted for ca. 3.16-3.69 eV in comparing with the protonated guanidine group in zwitterion Z_1. Besides, the peak at 406.42 eV can only be observed in the zwitterionic isomer. In the O edge, the difference in the carboxylate group (from COOH to COO^-) has made the BEs of O_{II} and O_I of the zwitterionic isomer red-shift to 535.17 eV (ca. 2.21-4.27 eV lower than the canonical isomers). It is interesting to find that the peak at 539.46 eV (from group RG_1) can be clearly separated from other canonical isomers in group RG_2.

Figure 5d-5f presents the individual NEXAFS spectra of the most populated neutral ArgGly isomers and their averaged spectra (SUM) at 498 K. Visible absorption spectral differences can be found between the canonical and zwitterionic forms at all K-edges. In the C1s edge, the first π resonance at about 287.95 eV mainly comes from the transition C_3 1s→$\pi^*_{C=O}$, which is involved in the peptide bond. This peak is red-shifted 0.23 eV comparing with the experimental feature observed for glycylglycine,[28] mainly due to the strong hydrogen bonds involved in the C=O group. The following double peaks come from transitions C_1 1s→$\pi^*_{C=O}$ and C_8 1s→$\pi^*_{C=N}$, respectively. The zwitterionic isomer Z_1 exhibits a slightly wider splitting of 0.77 eV (288.32 and 289.09 eV) than the 0.4 eV (288.40 and 288.80 eV) of the canonical conformers, which is in accordance with the situation of arginine.[19] This is mainly because that the charge states of the two atoms C_1 and C_8 are very different for the two isomers during the proton transformation process from the carboxylate group to the guanidine group. In the N spectra, due to the different charge distribution in atom N_γ in RG_1 (or N_ε in RG_2), only the canonical isomers display a strong π resonance at 399.70 eV, which comes from the transition N_γ (or N_ε) 1s→$\pi^*_{C=N}$. In the O1s edge, because of the environmental similarity for the oxygen atoms, both forms show two strong resonances at ca. 531.80-532.40 eV. The canonical form RG_{1a} shows two resonances at 535.15 and 535.99 eV and all other RG_2 conformers give a resonance at ca.534.39 eV, contributed by transitions O_I 1s→$\pi^*_{C=O}$. All the above spectral features thus provide a promising way to unambiguously distinguish the three different neutral isomers of ArgGly.

4 Conclusions

To summarize, we have conducted a first-principles study on electronic energy states of various conformers of neutral ArgGly, as well as their chemical structure (IR spectra) and electronic structure (XPS and NEXAFS spectra), to resolve the debate on the dominant form in the gas phase. The global minimum of dipeptide ArgGly is found to be in the canonical form, rather than the zwitterionic form. The unique features appeared in the vibrational spectra, especially the strong resonance ones in the 2700-3500 cm^{-1} region, can be used to distinguish three different neutral groups of ArgGly. In the XPS spectra, the band at 402.83, 537.50 and 539.46 eV are unique for the canonical form, while the band at 406.42 and 535.17 eV can only be observed in the zwitterion. In the NEXAFS spectra, the π resonance at 399.70 eV is due to the transition

N_V 1s→$\pi^*_{C=N}$ and the resonance at 534.39 eV corresponds to O_I 1s→$\pi^*_{C=O}$, all of which are only from the canonical contributions. The spectral differences will help people to unambiguously determine the dominant form of ArgGly in both theory and experiment.

Acknowledgements

The work is supported by Ministry of Science and Technology of China (Grant No. 2012CB215405), Natural Science Foundation of China (Grant No. 11074233 and 11374272), open project of Beijing National Laboratory for Molecular Sciences (2013002), and Program for Innovative Research Team of Guizhou Province of China (QKTD-[2012]4009). The Swedish National Infrastructure for Computing (SNIC) is acknowledged for computer time.

Notes and references

1 C. R. Cantor, P. R. Schimmel, *Biophysical Chemistry, Part I: The Conformation of Biological Macromolecules*. Freeman, W. H., San Francisco, 1980.
2 W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag: New York, 1984.
3 S. F. Sousa, P. A. Fernandes, M. J. Ramos, *J. Phys. Chem. A* 2009, **113**, 14231-14236.
4 Y. Grenie, J. Lassegues, C. Garrigou-Lagrange, *J. Chem. Phys.* 1970, **53**, 2980-2982.
5 R. D. Suenram, F. J. Lovas, *J. Mol. Spectrosc.* 1980, **72**, 372-382.
6 M. J. Locke, R. T. Jr. McIver, *J. Am. Chem. Soc.* 1983, **105**, 4226-4232.
7 T. R. Rizzo, Y. D. Park, D. H. Levy, *J. Chem. Phys.* 1986, **85**, 6945-6951.
8 C. H. Hu, M. Shen, H. F. Schaefer III, *J. Am. Chem. Soc.* 1993, **115**, 2923-2929.
9 M. S. Gordon, J. H. Jensen, *Acc. Chem. Res.* 1996, **29**, 536-543.
10 G. Bouchoux, *Mass Spectrom. Rev.* 2012, **31**, 391-435.
11 Y. Leng, M. Zhang, C. Song, M. Chen, Z. Lin, *Journal of Molecular Structure: THEOCHEM.* 2008, **858**, 52-65.
12 I. Peña, M. E. Sanz, J. C. López, J. L. Alonso, *J. Am. Chem. Soc.* 2012, **134**, 2305-2312.
13 S. Ling, W. Yu, Z. Huang, Z. Lin, M. Harañczyk, M. Gutowski, *J. Phys. Chem. A* 2006, **110**, 12282-12291.
14 W. Fei, A. K. Rai, Z. Lu, Z. Lin, *J. Mol. Struct (THEOCHEM).* 2009, **895**, 65-71.
15 W. D. Price, R. A. Jockusch, E. R. Williams, *J. Am. Chem. Soc.* 1997, **119**, 11988-11989.
16 C. J. Chapo, J. B. Paul, R. A. Provencal, K. Roth, R. J. Saykally, *J. Am. Chem. Soc.* 1998, **120**, 12956-12957.
17 J. Rak, P. Skurski, J. Simons, M. Gutowski, *J. Am. Chem. Soc.* 2001, **123**, 11695-11707.
18 S. Schlund, R. Muller, C. Grabmann, B. Engels, *J. Comput. Chem.* 2008, **29**, 407-415.
19 H. Li, W. Hua, Z. Lin, Y. Luo, *J. Phys. Chem. B* 2012, **116**, 12641-12650.
20 J. S. Prell, M. Demireva, J. Oomens, E. R. Williams, *J. Am. Chem. Soc.* 2009, **131**, 1232-1242.
21 C. Wang, Z. Lin, R. Zhang, *Comput. Theor. Chem.* 2013, **1008**, 96-102.
22 J. Oomens, J. D. Steill, B. Redlich, *J. Am. Chem. Soc.* 2009, **131**, 4310-4319.
23 J. Stöhr, *NEXAFS Spectroscopy*. Springerlag: Berlin, Heidelberg, NY, 1992.
24 O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, V. Carravetta, *J. Phys. Chem. A* 2007, **111**, 10998-11005.
25 O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, V. Carravetta, *J. Electron Spectrosc. Relat. Phenom.* 2007, **155**, 47-53.
26 O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, V. Carravetta, *Chem. Phys. Lett.* 2007, **442**, 429-433.
27 V. Feyer, O. Plekan, R. Richter, M. Coreno, K. C. Prince, V. Carravetta, *J. Phys. Chem. A* 2008, **112**, 7806-7815.
28 V. Feyer, O. Plekan, R. Richter, M. Coreno, K. C. Prince, V. Carravetta, *J. Phys. Chem. A* 2009, **113**, 10726-10733.
29 W. Zhang, V. Carravetta, O. Plekan, V. Feyer, R. Richter, M. Coreno, K. C. Prince, *J. Chem. Phys.* 2009, **131**, 035103.
30 B. M. Messer, C. D. Cappa, J. D. Smith, W. S. Drisdell, C. P. Schwartz, R. C. Cohen, R. J. Saykally, *J. Phys. Chem. B* 2005, **109**, 21640-21646.
31 B. M. Messer, C. D. Cappa, J. D. Smith, K. R. Wilson, M. K. Gilles, R. C. Cohen, R. J. Saykally, *J. Phys. Chem. B* 2005, **109**, 5375-5382.
32 N. Ottosson, K. J. Børve, D. Spångberg, H. Bergersen, L. D. Sæthre, M. Faubel, W. Pokapanich, G. Öhrwall, O. Björneholm, B. Winter, *J. Am. Chem. Soc.* 2011, **133**, 3120-3130.
33 Z. B. Maksic, B. Kovacevic, *J. Chem. Soc., Perkin Trans.* 1999, **2**, 2623–2629.
34 C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* 1988, **37**, 785-789.
35 J. Csontos, N. Y. Palermo, R. F. Murphy, S. Lovas, *J. Comput. Chem.* 2008, **29**, 1344-1352.
36 J. L. Durant, *Chem. Phys. Lett.* 1996, **256**, 595.
37 D. T. Nguyen, A. C. Scheiner, J. W. Andzelm, S. Sirois, D. R. Salahub and A. T. Hagler, *J. Comp. Chem.* 1997, **18**, 1609-1631.
38 G. Fogarasi, *J. Mol. Struct.* 1997, **413-414**, 271-278.
39 H. Li, Z. Lin, Y. Luo, *Chem. Phys. Lett.* 2014, **598**, 86-90.
40 H. Li, Z. Lin, Y. Luo, *Chem. Phys. Lett.* 2014, **610-611**, 303-309.
41 Y. Zhao, D. G. Thuhlar, *Acc. Chem. Rsc.* 2008, **41**, 157-167.
42 E. G. Hphenstein, S. T. Chill, C. D. Sherrill, *J. Chem. Theor. Comput.* 2008, **4**, 1996-2000.
43 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng,J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V.

N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 revision A.02*, Gaussian Inc., Wallingford CT, 2009.

44 K. Hermann, L. G. M. Pettersson, M. E. Casida, C. Daul, A. Goursot, A. Koester, E. Proynov, A. St-Amant, D. R. Salahub, *StoBe-deMon, version 3.0*; StoBe-deMon Software: Stockholm-Berlin, 2007.

45 A. D. Becke, *Phys. Rev. A* 1988, **38**, 3098-3100.

46 J. P. Perdew, *Phys. Rev. B* 1986, **33**, 8822-8824.

47 W. Kutzelnigg, U. Fleischer, M. Schindler, *NMR-Basic Principles and Progress*; Springer-Verlag: Heidelberg, 1990, **23**, 165-262.

48 L. Triguero, L. G. M. Petersson, H. Ågren, *Phys. Rev. B* 1998, **58**, 8097-8110.

49 A. R. Slaughter, M. S. Banna, *J. Phys. Chem.* 1988, **92**, 2165-2167.

50 E. E. Aziz, N. Ottosson, S. Eisebitt, W. Eberhardt, B. Jagoda-Cwiklik, R. Vacha, P. Jungwirth, B. Winter, *J. Phys. Chem. B* 2008, **112**, 12567-12570.

51 J. Grasjo, E. Andersson, J. Forsberg, L. Duda, E. Henke, W. Pokapanich, O. Bjorneholm, J. Andersson, A. Pietzsch, F. Hennies, J. E. Rubensson, *J. Phys. Chem. B* 2009, **113**, 16002-16006.

52 R. J. Gdanitz, W. Cardoen, T. L. Windus, J. Simons, *J. Phys. Chem. A* 2004, **108**, 515-518.

53 H. Li, Z. Lin, Y, Luo, *Chem. Phys. Lett.* 2014, **608**, 398-403.