# Core-level electronic structure of solid-phase glycine, glycyl-glycine, diglycyl-glycine, and polyglycine: X-ray photoemission analysis and Hartree–Fock calculations of their zwitterions

Avisek Chatterjee, Liyan Zhao, Lei Zhang, Debabrata Pradhan, Xiaojing Zhou,[a] and K. T. Leung[b]

*WATLab and Department of Chemistry, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

(Received 17 June 2008; accepted 4 August 2008; published online 11 September 2008)

X-ray photoelectron spectroscopy (XPS) has been used to investigate the core-level electronic structures of glycine (G) and its peptides, including glycyl-glycine (GG), diglycyl-glycine (GGG), and polyglycine (poly-G), in their powder forms. Increasing the number of G units in the peptides does not change the locations of the respective C 1*s*, N 1*s*, and O 1*s* features corresponding to different functional groups: $-COO^-$, $-NH_3^+$, $>CH_2$, and –CONH–. The electronic structures of the zwitterions of these molecules have been calculated as isolated molecules and as molecules in an aqueous environment under the periodic boundary conditions by quantum-mechanical and molecular mechanics methods. In the case of glycine zwitterion, the binding energies of the C 1*s*, N 1*s*, and O 1*s* XPS features are found to be in reasonable accord with the respective orbital energies obtained by Hartree–Fock self-consistent-field calculations, within the context of Koopmans' approximation. However, considerably worse agreement in the binding energies is found for the larger zwitterions (with the specific conformations considered in this work), indicating the need for higher-level calculations. The present work shows that optimizing the zwitterion in an aqueous environment under the periodic boundary conditions by molecular mechanics could be a very cost-effective approach for calculating the electronic structures of large, complex biomolecular systems. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2976151]

## I. INTRODUCTION

Electronic structures of amino acids are fundamentally important to our understanding of electron-transfer processes in biochemical reactions and biological interactions.[1] In addition to solution-phase chemistry, the adsorption of amino acids on soft and hard surfaces provides a crucial alternative pathway for the formation of peptides and proteins.[2] The bioorganic functionalization of substrates offers potential applications in biosensors and bioelectronics.[3,4] The evolution of individual amino acids to condensed phases (clusters, nanoparticles, powders, and films) presents a great challenge to computational chemistry for modeling hydrogen bonding interactions,[5–8] proton transfer,[9–13] and other properties of the condensed phase.[14–16]

As the simplest amino acid, glycine ($G, NH_2CH_2COOH$), is one of the most studied model biomolecules for understanding the interactions with other amino acids and with various solid surfaces.[17–20] Glycyl-glycine ($GG, NH_2CH_2CONHCH_2COOH$), diglycyl-glycine [$GGG, NH_2CH_2(CONHCH_2)_2COOH$], and polyglycine [poly-G, $NH_2CH_2(CONHCH_2)_nCOOH$] offer an excellent series of peptides for studying the formation of peptide bonds and the evolution of the electronic structure with the number of G units. Figure 1 illustrates the equilibrium geometries of plausible conformers of G and its peptide homologs (in straight chain forms), obtained by *ab initio* Hartree–Fock (HF) self-consistent-field (SCF) calculations with a 6-31G basis set. To date, only a limited number of studies of gas-phase G by x-ray photoelectron spectroscopy (XPS) have been reported,[21,22] while no XPS data on gas-phase GG, GGG, and poly-G are available in the literature. In particular, Slaughter and Banna obtained from the measured XPS data of G the binding energies (BEs) of C 1*s* for

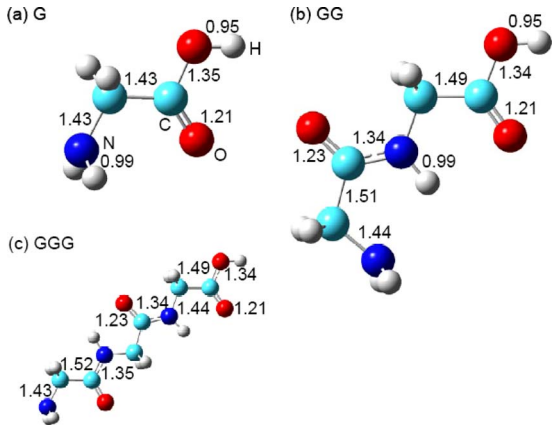


FIG. 1. (Color online) Equilibrium structures of (a) glycine (G), (b) glycyl-glycine (GG), and (c) diglycyl-glycine (GGG) obtained by HF/6-31G calculations. The scale used in (c) is twice as large as those used in (a) and (b). The bond lengths between selected atoms are indicated in unit of Å.

[a]Present address: School of Mathematics and Physical Science, University of Newcastle, Newcastle, New South Wales 2300, Australia.
[b]Author to whom correspondence should be addressed. Electronic mail: tong@uwaterloo.ca.

 

the methylene (292.25 eV) and carboxyl groups (295.15 eV), O 1*s* for the keto (538.2 eV) and hydroxyl groups (540.0 eV), and N 1*s* for the amino group (405.58 eV).[21] In the case of solid-phase G and their peptides, only one complete work has been reported. In 1976, Clark *et al.* measured the XPS spectra of the C 1*s*, N 1*s*, and O 1*s* regions of G, GG, and poly-G and concluded that the amino acids have zwitterionic structures in their solid state.[23] They also reported the corresponding calculated orbital energies of the respective zwitterions (ZIs) by SCF molecular orbital calculations. Gerin *et al.* also gave an XPS analysis of many biochemical compounds, including polyalanine, polyvaline, sorbitol, glucose, and poly-G.[24] However, due to the somewhat limited energy resolution and charge compensation capability, the quality of these earlier spectra is often found to be inadequate. In the present work, we provide a systematic XPS study of the electronic structures of solid-phase G, GG, GGG, and poly-G. We also perform *ab initio* SCF calculations of the ZIs, geometry optimized with and without molecular mechanics, and compare their orbital energies with the corresponding experimental BEs of C 1*s*, N 1*s*, and O 1*s* in the context of Koopmans' approximation.

## II. EXPERIMENTAL AND COMPUTATIONAL DETAILS

Polycrystalline powders of G (98.5% purity), GG (99.5% purity), GGG, and poly-G (both 99% purity), obtained commercially from Aldrich, were individually ground in a mortar, and a thin layer of the resulting powders was mounted by gently pressing on a double-sided carbon tape on a sample holder. The corresponding XPS spectra were measured by using a Thermo-VG Scientific ESCALab 250 Microprobe equipped with a monochromatic Al $K\alpha$ x-ray source (1486.6 eV), operated with a typical energy resolution of 0.4–0.5 eV full width at half maximum. The BE scale was calibrated by using the C 1*s* line at 284.8 eV originated from the part of the carbon tape not covered by the powders. As nonconductive materials, the amino acid powders were found to exhibit serious charging effects during XPS measurement, especially with the use of a high-power monochromated x-ray source. In the present work, superior sample charge compensation was achieved by a combination of low-energy electrons and Ar ions delivered by two flood guns. No discernible changes in the XPS spectra (caused by x-ray-induced sample degradation) have been observed in the duration of the experiment (typically less than 6 h). Curve fitting of XPS spectra was performed by using the CasaXPS VAMAS processing software, and the accuracy of the fitted peak positions was estimated to be $\pm 0.1$ eV. The atomic concentration ratio of different components was determined by the ratio of the peak areas normalized by the relative sensitivity factors.[25]

Calculations of the equilibrium geometries of the ZIs $[NH_3^+CH_2(CONHCH_2)_mCOO^-]$ of G ($m=0$), GG ($m=1$), and GGG ($m=2$) involving *ab initio* SCF and molecular mechanics methods were performed by using the HYPERCHEM software (Hypercube Inc.).[26] Due to the charge localizations associated with the $NH_3^+$ and $COO^-$ moieties in the ZI, special attention was needed for optimizing the geometries and calculating the total energies. In the present work, we obtained these quantities by two simple approaches, i.e., by treating the ZI as an isolated molecule (IM) or a molecule inside an aqueous environment under the periodic boundary conditions (PBC). Figure 2 gives an example of the geometries of a GG ZI obtained by the IM and PBC approaches. In the IM approach, the equilibrium geometry of the isolated (or free) ZI was optimized and the corresponding total energy was calculated by using the HF method. The 6-31G basis set has been used and compared with two other larger basis sets (6-31++G and 6-31+G*), because it has been found to produce better overall agreement in the molecular properties for amino acids and their peptides than the other split-valence basis sets.[27] In the PBC approach, the ZI was placed in a periodic cubic cell containing a sufficiently large number of water molecules to simulate an aqueous environment. A minimum separation between the ZI (the solute) and the water molecules (the solvent) of 2.3 Å was used to prevent the formation of covalent bonds as in a hydrated molecule and to assure a weak force-field environment. Specify-

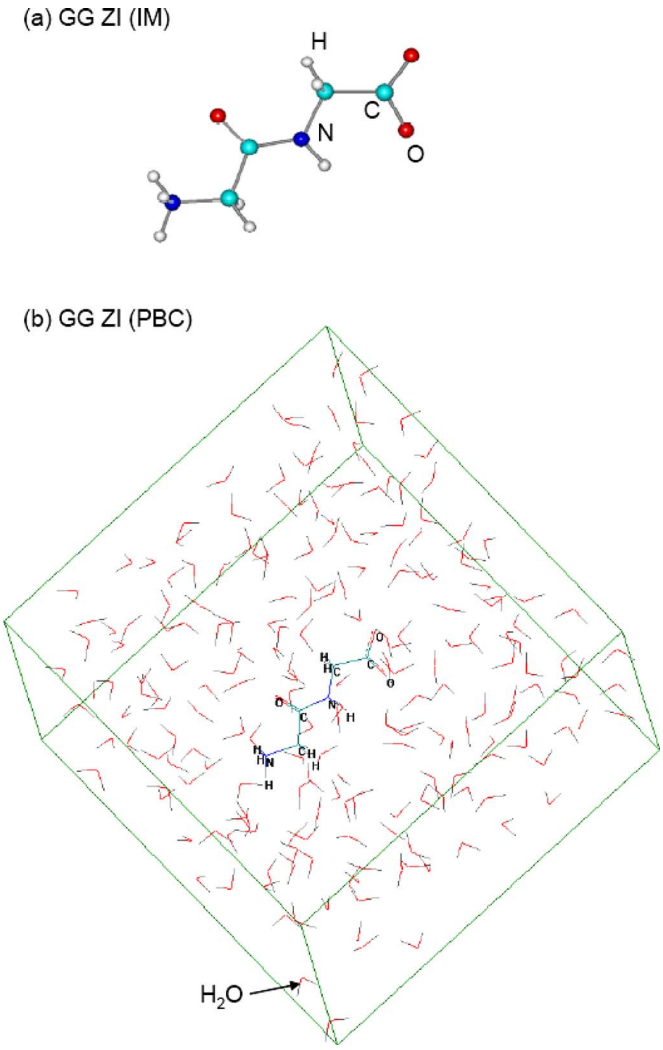


FIG. 2. (Color online) Optimized structures of glycyl-glycine (GG) zwitterion (ZI) obtained by (a) HF/631G in the isolated molecule (IM) approach and (b) MM+ molecular mechanics in the periodic boundary conditions (PBC) approach.

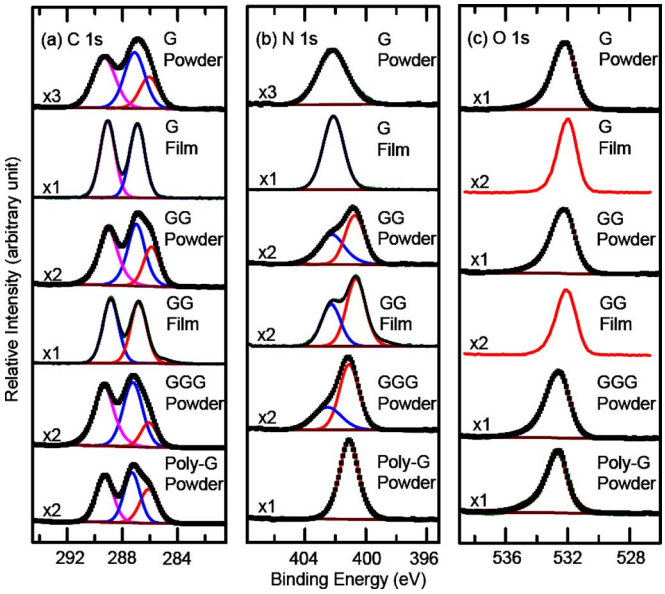


FIG. 3. (Color online) XPS spectra of (a) C 1*s*, (b) N 1*s*, and (c) O 1*s* regions for (from top to bottom) glycine (G) powder and film, glycyl-glycine (GG) powder and film, and diglycyl-glycine (GGG) and polyglycine (poly-G) powders. The experimental data (■) are fitted with Gaussian line-shapes shown as solid lines. For each of the species, the relative intensities of the respective XPS spectra are indicated by the scale factors.

ing a larger cell dimension also increased the number of equilibrated water molecules surrounding the ZI, thereby improving the force field around the ZI as approximated by molecular mechanics. The MM+ functional form[28] for the force field (with the bond lengths, bond angles, torsional angles, nonbonded interactions, and hydrogen bonds as the components) has been used in the molecular mechanics calculation. Different cell dimensions (with different numbers of water molecules) have also been tried, and it was determined that a periodic cell with a 18.7 Å edge length containing 216 randomly distributed water molecules gave a minimized total energy (that did not vary significantly with further increase in the cell dimension) upon optimization. The resulting ZI structure was then used for a single-point energy calculation by using the HF method (with the 6-31G, 6-31++G, and 6-31+G* basis sets). The input parameters for the geometries of G, GG, and GGG were taken from the standard values reported in the earlier studies.[29–31]

## III. RESULTS AND DISCUSSION

Figure 3 shows the XPS spectra of C 1*s*, N 1*s* and O 1*s* regions of G, GG, GGG, and poly-G powders. Evidently, all four species exhibit very similar C 1*s* spectra, with the peaks at 287.0–287.3 and 289.0–289.3 eV corresponding to the methylene ($>CH_2$) carbon and to the deprotonated carboxyl or carboxylate ($-COO^-$) and amide (–CONH–) carbons, respectively.[23] Furthermore, the intensity ratio between the two C 1*s* features is approximately unity, as expected from the stoichiometric ratio of the two C atoms in the respective species. It should be noted that the lowest-lying C 1*s* peak at 286.0 eV corresponds to the surface carbonaceous species arising from handling of the powder sample. Similarly, all four species also exhibit very similar O 1*s* spectra, with a single peak at 532.1–532.5 eV attributed to the carboxylate group,[23] which therefore suggests that all four species exist in their ZI forms. On the other hand, the N 1*s* spectra for all four species appear to be discernibly different from one another. In particular, a single N 1*s* feature at 402.2 eV for G corresponds to the protonated amino group ($NH_3^+$) of its ZI ($NH_3^+CH_2COO^-$),[23] thus confirming the presence of the carboxylate group as observed from the corresponding O 1*s* spectrum. Two N 1*s* features for GG are observed at 400.7 and 402.2 eV, which correspond, respectively, to the amide group and the protonated amino ($-NH_3^+$) group of the ZI ($NH_3^+CH_2CONHCH_2COO^-$). For GGG, two N 1*s* peaks are also found at 401.1 and 402.5 eV, which can be similarly assigned to the corresponding amide and protonated amino group of the ZI, $NH_3^+CH_2(CONHCH_2)_2COO^-$. Although the respective energy separations between the amide and protonated amino N 1*s* features are remarkably similar (1.5 and 1.4 eV), the intensity ratios for the two N 1*s* features for GG ZI (1:1) and GGG ZI (2:1) are notably different, as expected. For the poly-G ZI [$NH_3^+CH_2(CONHCH_2)_mCOO^-$, large $m$], a single N 1*s* feature at 401.1 eV is observed and it clearly corresponds to the predominant amide group.

Table I summarizes the peak locations for the four species. It should be noted that the atomic ratios for C:N:O obtained from the fitted spectra for G (2.2:1.0:1.8), GG (2.1:1.0:1.3), GGG (2.2:1.0:1.2), and poly-G (2.0:1.0:1.2) are in excellent accord with the respective stoichiometric ratios for G (2:1:2), GG (2:1:1.5), GGG (2:1:1.3), and poly-G (2:1:1). In Table I, we also compare the BEs for the C 1*s*, N 1*s*, and O 1*s* features for G, GG, and poly-G with those obtained by Clark *et al.*[23] In particular, the reported BE positions for C 1*s* ($>CH_2$), C 1*s* ($-COO^-/-CONH-$), N 1*s* (–CONH–), and N 1*s* ($-NH_3^+$) are found to be lower by, respectively, 0.3–0.8, 0.6–1.1, 0.3–0.5, and 0.2–0.3 eV, while that for O 1*s* is found to be 0.1–0.3 eV higher than the corresponding values obtained in the present work. These minor discrepancies could likely be due to a small sample charging effect often encountered in these powder samples and/or to a lower energy resolution employed by Clark *et al.*[23] To further assure that the observed BE positions are not due to sample interactions with the carbon tape used for sample mounting, we have conducted XPS measurements of a G and a GG films in a separate XPS instrument (manufactured by Omicron Nanotechnology).[32] These films (approximately 16 and 11 adsorption layers for G and GG, respectively) were deposited *in situ* on a clean Si(111)7×7 substrate under ultrahigh vacuum condition by thermal evaporation of their powders in effusion cells. The corresponding XPS spectra of the films for G and GG are compared with those of the powders in Fig. 3. Due to the more precise and cleaner conditions (UHV and single crystal substrate), the peaks obtained in the thick-film case appear to be discernibly sharper. Evidently, the C 1*s*, N 1*s*, and O 1*s* spectra of the thick films on the Si(111)7×7 surface are in good agreement with those of the respective powder samples. As it can be observed from the BEs of C 1*s*, N 1*s*, and O 1*s* of the films and powders, the powder spectra could be considered as the limiting case of the thick-film spectra. The fitted BE positions for these films, shown also in Table I,

TABLE I. Comparison of the experimental binding energies with, where available, an earlier experimental work of glycine, glycylglycine, diglycylglycine, and poly-glycine powders, and with the corresponding calculated orbital energies obtained by HF in the different basis sets in both isolated molecule (IM) and periodic boundary conditions (PBC) approaches. The respective calculated total energy, dipole moments and virial coefficients (−V/T) are also compared with the earlier calculations. The corresponding experimental binding energies for thin films of glycine and glycylglycine are also given (in square parentheses).

| Glycine | Expt. This Work | Expt. (Ref. 23) | HF/6-31G | | HF/6-31++G | | HF/6-31+G* | | Earlier work (Refs. 23 and 30) |
|---|---|---|---|---|---|---|---|---|---|
| | | | IM | PBC | IM | PBC | IM | PBC | |
| C 1*s* (>$CH_2$) (eV) | 287.1 [287] | 286.8 ±0.2 | 279.49 | 258.12 | 284.87 | 272.78 | 287.84 | 268.63 | 307.34 |
| C 1*s* (–$COO^-$) (eV) | 289.3 [289.2] | 288.7 ±0.2 | 284.70 | 274.31 | 291.57 | 289.0 | 288.45 | 288.96 | 306.86 |
| N 1*s* ($NH_3^+$) (eV) | 402.2 [402.2] | 402.0 ±0.1 | 395.08 | 366.68 | 399.12 | 381.44 | 397.52 | 383.33 | 428.79 |
| O 1*s* (eV) | 532.1 [532.1] | 532.4 ±0.1 | 536.15 | 507.84 | 541.72 | 514.31 | 535.40 | 512.58 | 558.05 |
| Total energy (a.u.) | | | −282.65711 | −282.58472 | −282.67143 | −282.60442 | −282.79761 | −282.71400 | −282.6199 |
| $\mu$ (Debye) | | | 10.7779 | 14.1366 | 11.243 | 14.319 | 10.8057 | 14.2287 | 13.85 |
| −V/T | | | 1.9997 | 2.0012 | 2.0000 | 2.0017 | 2.0000 | 2.0037 | |
| Glycyl-glycine | | | | | | | | | |
| C 1*s* (>$CH_2$) (eV) | 287.0 [286.8] | 286.6 ±0.3 | 298.82 | 285.32 | 305.59 | 296.06 | 262.08 | 292.03 | |
| C 1*s* (–$COO^-$) (eV) | 289.0 [288.9] | 288.4 ±0.2 | 347.30 | 307.94 | 355.18 | 319.07 | 303.93 | 315.07 | |
| C 1*s* (–CONH–) (eV) | | | | | | | | | |
| N 1*s* (–CONH–) (eV) | 400.7 [400.7] | 400.4 ±0.1 | 397.88 | 371.73 | 384.90 | 386.04 | 388.80 | 382.757 | |
| N 1*s* ($NH_3^+$) (eV) | 402.2 [402.3] | 401.9 ±0.1 | 413.65 | 414.98 | 408.35 | 422.23 | 405.89 | 418.27 | |
| O 1*s* (eV) | 532.2 [532.2] | 532.4 ±0.2 | 536.15 | 529.14 | 541.72 | 531.85 | 517.17 | 526.13 | |
| Total energy (a.u.) | | | −489.33312 | −489.25451 | −489.35650 | −489.28293 | −489.57560 | −489.48667 | −485.7173 |
| $\mu$ (Debye) | | | 22.9687 | 28.5718 | 23.5031 | 28.767 | 23.0659 | 28.3737 | 24.60 |
| −V/T | | | 1.9995 | 2.0014 | 1.9999 | 2.0018 | 2.0011 | 2.0036 | |
| Diglycyl-glycine | | | | | | | | | |
| C 1*s* (>$CH_2$) (eV) | 287.3 | | 268.02 | 276.30 | 275.48 | 271.40 | 272.12 | 285.39 | |
| C 1*s* (–$COO^-$) (eV) | 289.3 | | 314.93 | 298.36 | 323.20 | 289.95 | 319.97 | 315.11 | |
| C 1*s* (–CONH–) (eV) | | | | | | | | | |
| N 1*s* (–CONH–) (eV) | 401.1 | | 382.66 | 391.91 | 386.85 | 387.59 | 385.19 | 384.03 | |
| N 1*s* ($NH_3^+$) (eV) | 402.5 | | 408.85 | 404.69 | 388.32 | 397.04 | 387.14 | 395.47 | |
| O 1*s* (eV) | 532.5 | | 523.92 | 528.88 | 529.79 | 526.73 | 528.56 | 531.71 | |
| Total energy (a.u.) | | | −696.03394 | −695.95127 | −696.06508 | −695.98717 | −696.37632 | −696.28663 | −690.0096 |
| $\mu$ (Debye) | | | 37.0889 | 44.166 | 37.6032 | 44.166 | 37.2030 | 43.5991 | 42.06 |
| −V/T | | | 1.9995 | 2.0011 | 1.9998 | 2.0014 | 2.001 | 2.0032 | |
| Poly glycine[a] | | | | | | | | | |
| C 1*s* (>$CH_2$) (eV) | 287.3 | 286.5 ±0.2 | | | | | | | |
| C 1*s* (–CONH–) (eV) | 289.3 | 288.2 ±0.2 | | | | | | | |
| N 1*s* (–CONH–) (eV) | 401.1 | 400.6 ±0.1 | | | | | | | |
| O 1*s* (eV) | 532.5 | 532.6 ±0.1 | | | | | | | |

[a]The experimental BE values for poly-glycine are also compared with those for C 1*s* (>$CH_2$) (289.1 ± 0.05 eV), C 1*s* (–CONH–/–$COO^-$) (288.0 ± 0.06 eV), N 1*s* (–CONH–) (399.8 ± 0.04 eV), and O 1*s* (531.3 ± 0.09 eV) reported by Gerin *et al.* (Ref. 24).

are found to be in excellent agreement with the present work. As expected, the lowest-BE C 1*s* features at 286.0 eV found for the powder spectra are not observed in the film spectra [Fig. 3(a)], which is consistent with our earlier assignment of this feature to the surface carbonaceous species on the powders.

Figure 4 shows the equilibrium structures for the ZIs of G, GG, and GGG obtained by HF/6-31G using the IM and PBC approaches. The structural parameters of the ZIs obtained by the IM and PBC methods are generally found to be similar to each other (with differences less than 10%). The peptide C–N bond lengths calculated by IM and PBC ap-

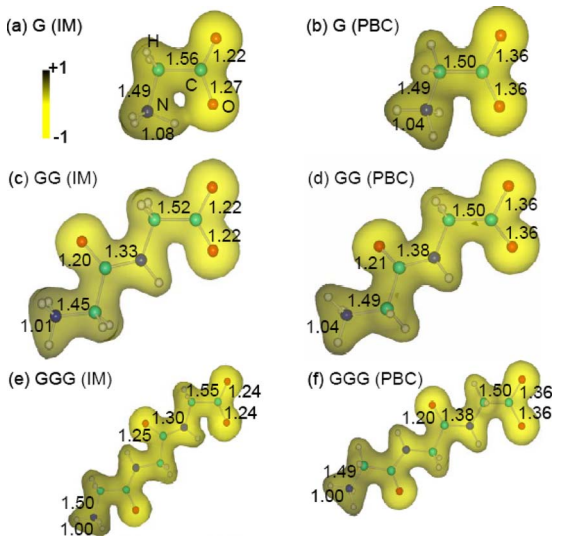


FIG. 4. (Color online) Structures and electrostatic potential distributions of glycine (G), glycyl-glycine (GG) and diglycyl-glycine (GGG) zwitterions (ZIs) calculated by using the HF/6-31G method, following the isolated molecule (IM) and periodic boundary conditions (PBC) approaches. The bond lengths between selected atoms are indicated in unit of Å.

proaches also do not appear to change noticeably (i.e., by less than 3%) from the neutral to ZI form for G, GG, and GGG. Moreover, two different C–O bond lengths in the carboxyl group are obtained for the neutral form, with 1.21 Å for the carbonyl C═O bond length and 1.34 Å for the C–OH bond length. Given that the C–O bond length for the terminal $COO^-$ group is found to be 1.21 Å for the IM case [Figs. 4(a), 4(c), and 4(e)] but 1.36 Å for the PBC case [Figs. 4(b), 4(d), and 4(f)], the C–O (terminal) bond therefore resembles more like a double bond for the IM case but a single bond for the PBC case. The latter could be due to the nature of the weak force field surrounding the ZI in the PBC unit cell. Furthermore, in the case of the G ZI, the OCO bond angle is found to increase to 132° (from 122° for neutral G) with one of the C–O bond lengths also increased to 1.27 Å in the IM case, which indicates the presence of intramolecular hydrogen bonding between the–C–O and–NH groups [Fig. 4(a)]. The distances between the H and O atoms of the neighboring –NH and the $–COO^-$ groups in the IM and PBC cases are found to be, respectively, 2.26 and 2.72 Å for GG and 1.86 and 2.52 Å for GGG. This indicates that hydrogen bonding is always favoured in the IM case compared to the PBC case, in which the presence of the water molecules serves to stabilize the negative and positive charges of the ZI. It should be noted that an increasing number of polarization functions in the basis set would allow the hydrogen bonding between the neighboring –NH and the $–COO^-$ groups to become more prominent, resulting in an increase in the N–H bond length and a decrease in the separation between the carboxylate O and the protonated amino H. For isolated ZIs, this leads to spontaneous migration of a proton from the protonated amino group ($NH_3^+$) to the carboxylate group ($–COO^-$), resulting in the neutral species. Figure 4 also shows the corresponding electrostatic potentials of the three ZIs as three-dimensional (3D) isosurfaces, where the dark regions correspond to the positive electrostatic potential while the lighter (yellow) regions show the negative electrostatic potential over the range $+1e/a_0$ to $-1e/a_0$ (where $e=+1.6\times10^{-19}$ C, $a_0=0.52918$ Å). As expected, the $–COO^-$ terminal group and the O atom in the peptide C═O all have negative potential, while the $NH_3^+$ terminal group, and the H atoms in the –CONH–group and the $>CH_2$ groups have positive potential. As electrostatic potential corresponds to the interaction energy of the molecular system with a positive point charge, these 3D isosurface plots can be useful for locating the electrophilic (e.g., $–NH_3^+$) and nucleophilic sites (e.g., $–COO^-$) in the molecule. As the number of G units increases (e.g., from GG to GGG), the center nucleophilic sites (e.g., >C═O), in addition to the terminal nucleophilic site (e.g., $–COO^-$), will become available, which opens up new bonding possibilities for the longer peptides.

From the optimized structures, the corresponding orbital energies for the C 1*s*, O 1*s*, and N 1*s* orbitals of the ZI for G, GG, and GGG obtained for the IM and PBC cases were also calculated. The calculated total energies, dipole moments and virial coefficients for G, GG, and GGG ZIs obtained by the HF method with different basis sets following the IM and PBC approaches are compared in Table I. Not surprisingly, the larger basis set tends to give a lower total energy for both IM and PBC cases, with the 6-31+G* basis set providing the lowest total energy. Furthermore, the replacement of diffuse functions by polarization functions (e.g., 6-31+G* versus 6-31++G) appears to be slightly more effective in lowering the total energy than the diffuse functions (e.g., 6-31++G versus 6-31G). The total energies obtained by the IM approach (e.g., −282.657 11 a.u. for HF/6-31G for G ZI) are consistently lower than the corresponding values obtained by the PBC approach (e.g., −282.584 72 a.u. for HF/6-31G for G ZI), which is expected because geometry optimization is performed by HF method in the IM case but only by classical molecular mechanics in the PBC case. The differences in the total energies obtained for the IM and PBC cases with the same basis set are essentially the same for G ZI (e.g., −0.072 39 a.u. or −2.0 eV for HF/6-31G), GG ZI (e.g., −0.078 61 a.u. or −2.1 eV for HF/6-31G) and GGG ZI (e.g., −0.082 67 a.u. or −2.2 eV for HF/6-31G), which shows that the PBC approach is equally effective in stabilizing the larger ZIs as the smaller G ZI. This difference also appears to be nearly the same obtained for the three basis sets for the same ZI. Furthermore, it is not surprising that the present calculations obtained with the larger basis sets also give generally lower total energies than the calculation performed with the STO-NG basis set reported by Wright and Borkman.[30] Despite the minor differences in the calculated total energies, the calculated dipole moments obtained by the IM approach are generally lower than those by the PBC approach, which could be attributed to internal charge reallocation caused by intramolecular hydrogen bonding. These IM values also appear to be in better agreement with the calculated dipole moments reported by Wright and Borkman.[30] It is of interest to note the dramatic increase in the calculated dipole moment with increasing G units in the peptides, which is consistent with the increase in the chain length in the straight-chain conformations of GG and GGG that have been used in the present calculation. Finally, the calculated virial coefficients

for all cases are essentially two, suggesting that it is not a sensitive parameter to gauge the quality of the calculation.

The comparison of these calculated orbital energies with the respective experimental BEs in the context of Koopmans' theorem[33] is also given in Table I. Evidently, the calculated orbital energies are found to be generally within 66 eV from the respective BEs, with the larger basis sets not necessarily giving better agreement. For instance, the discrepancy between calculation and experiment for G is particularly worse for O 1*s* for the PBC case, with 20 eV difference for HF/6-31+G* and 25 eV for O 1*s* for HF/6-31G. For GG, the discrepancy is worse for C 1*s* (–$COO^-$) for HF/6-31++G for the IM case (66 eV) and N 1*s* (–CONH–) for HF/6-31G for the PBC case (29 eV). For GGG, the discrepancy is greater for C 1*s* (–$COO^-$) for HF/6-31++G for the IM case (34 eV) and N 1*s* (–CONH–) for HF/6-31G for the PBC case (19 eV). Despite these discrepancies, the present calculation shows that the PBC approach gives a better overall agreement than the IM approach for the larger ZI such as GGG for the majority of the XPS core-level features. More importantly, the PBC method gives considerably better computational efficiency over the IM approach, with the least amount of cpu time required for the calculation. This improvement in computational efficiency will become especially important for modeling larger and more complex biologically relevant systems such as DNA.

## IV. CONCLUDING REMARKS

The present work provides an improved XPS measurement for the C 1*s*, N 1*s*, and O 1*s* regions on glycine and its peptides, including glycyl-glycine, diglycyl-glycine, and polyglycine, in their solid (powder) forms. Excellent agreement between the XPS spectra of the powders and those of the condensed films on Si(111)7×7 has been observed. The electronic structures and the bonding characters, as well as the equilibrium geometries of the zwitterions of these important biomolecules are obtained by *ab initio* Hartree–Fock calculations with large basis sets. Despite the reasonable agreement between the experimental BEs and the corresponding orbital energies observed for glycine (within the limit of Koopmans' approximation), considerable discrepancies remain for the zwitterions of its peptides, suggesting the need for higher-level calculations for these larger biomolecular systems. While the PBC approach tends to give slightly higher total energies and dipole moments than the IM approach, it appears to provide a somewhat better agreement with the experiment for the orbital energies of the larger peptides. It should also be noted that the ZIs are not stable in the gaseous form or as isolated species as demonstrated by a higher-level calculation (MP2/6-31G*) on G ZI (Ref. 34) and the other experiments.[35–37] In contrast, ZIs are found to exist in the solution and in the condensed or solid state,[34] making the PBC model discussed here a more realistic approach to handle these special ZI species. The PBC method gives considerably better computational efficiency over the IM approach, which will become especially important for modeling larger and more complex biologically relevant systems such as DNA.

## ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

[1] T. A. Romanova and P. V. Avramov, Dokl. Biochem. Biophys. **383**, 71 (2002).
[2] R. B. Merrifield, J. Am. Chem. Soc. **85**, 2149 (1963).
[3] B. Kasemo, Curr. Opin. Solid State Mater. Sci. **3**, 451 (1998).
[4] Y. Cui, Q. Wei, H. Park, and C. M. Lieber, Science **293**, 1289 (2001).
[5] M. A. Spackman, J. Chem. Phys. **85**, 6587 (1986).
[6] C. Pannayiotou and I. C. Sanchez, J. Phys. Chem. **95**, 10090 (1991).
[7] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras, J. Chem. Phys. **114**, 5149 (2001).
[8] S. Shan, S. Loh, and D. Herschlag, Science **272**, 97 (1996).
[9] D. Borgis and J. T. Hynes, J. Chem. Phys. **94**, 3619 (1991).
[10] S. Hammers-Schiffer and J. C. Tully, J. Chem. Phys. **101**, 4657 (1994).
[11] S. Scheiner, J. Phys. Chem. A **104**, 5898 (2000).
[12] T. Fiebig, M. Chachisvilis, M. Manger, A. H. Zewail, A. Douhal, I. Garcia-Ochoa, and A. de La Hoz Ayuso, J. Phys. Chem. A **103**, 7419 (1999).
[13] P. F. Barbara, G. C. Walker, and T. P. Smith, Science **256**, 975 (1992).
[14] A. Nilsson and L. G. M. Pettersson, Surf. Sci. Rep. **55**, 49 (2004).
[15] M. Preuss, W. G. Schmidt, and F. Bechstedt, Phys. Rev. Lett. **94**, 236102 (2005).
[16] Q. Chen and N. V. Richardson, Nat. Mater. **2**, 324 (2003).
[17] G. Tzvetkov, G. Koller, Y. Zubavichus, O. Fuchs, M. B. Casu, C. Heske, E. Umbach, M. Grunze, M. G. Ramsey, and F. P. Netzer, Langmuir **20**, 10551 (2004).
[18] V. Efstathiou and D. P. Woodruff, Surf. Sci. **531**, 304 (2003).
[19] P. Löfgren, A. Krozer, J. Lausmaa, and B. Kasemo, Surf. Sci. **370**, 277 (1997).
[20] G. Tzvetkov, M. G. Ramsey, and F. P. Netzer, Surf. Sci. **526**, 383 (2003).
[21] A. R. Slaughter and M. S. Banna, J. Phys. Chem. **92**, 2165 (1988).
[22] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, and V. Carravetta, Chem. Phys. Lett. **442**, 429 (2007).
[23] D. T. Clark, J. Peeling, and L. Colling, Biochim. Biophys. Acta **453**, 533 (1976).
[24] P. A. Gerin, P. B. Dengis, and P. G. Rouxhet, J. Chim. Phys. Phys.-Chim. Biol. **92**, 1043 (1995).
[25] J. F. Moulder, W. F. Stickle, P. E. Sobol, and K. D. Bomben, *Handbook of X-ray Photoelectron Spectroscopy*, 2nd ed., edited by J. Chastain (Perkin-Elmer, Waltham, MA, 1992).
[26] HYPERCHEM 7.02, Hypercube, Inc., Gainesville, Florida.
[27] L. R. Wright and R. F. Borkman, J. Am. Chem. Soc. **102**, 6207 (1980).
[28] N. L. Allinger, J. Am. Chem. Soc. **99**, 8127 (1977).
[29] S. G. Stepanian, I. D. Reva, E. D. Radchenko, M. T. S. Rosado, M. L. T. S. Duarte, R. Fausto, and L. Adamowicz, J. Phys. Chem. A **102**, 1041 (1998).
[30] L. R. Wright and R. F. Borkman, J. Phys. Chem. **86**, 3956 (1982).
[31] Y. C. Tse, M. D. Newton, S. Visheshwara, and J. A. Pople, J. Am. Chem. Soc. **100**, 4329 (1978).
[32] L. Zhang, A. Chatterjee, and K. T. Leung (unpublished).
[33] T. C. Koopmans, Physica (Amsterdam) **1**, 104 (1933).
[34] D. Yu, D. A. Armstrong, and A. Rauk, Can. J. Chem. **70**, 1762 (1992).
[35] L. Schafer, H. L. Sellers, F. J. Lovas, and R. D. Suenram, J. Am. Chem. Soc. **102**, 6566 (1980).
[36] R. D. Suenram and F. J. Lovas, J. Am. Chem. Soc. **102**, 7180 (1980).
[37] R. D. Suenram and F. J. Lovas, J. Mol. Spectrosc. **72**, 372 (1978).