Correlation of electronic structures of three cyclic dipeptides with their photoemission spectra

Anoja P. Wickrama Arachchilage,[1] Feng Wang,[1] Vitaliy Feyer,[2] Oksana Plekan,[2,a)] and Kevin C. Prince[2,3,b)]
[1]*Centre for Molecular Simulation, Swinburne University of Technology, Hawthorn, Melbourne, Victoria 3122, Australia*
[2]*Sincrotrone Trieste, in Area Science Park, I-34012 Basovizza, Trieste, Italy*
[3]*Istituto Officina dei Materiali, Consiglio Nazionale delle Ricerche, Area Science Park, I-34149 Trieste, Italy*

(Received 19 May 2010; accepted 20 September 2010; published online 5 November 2010)

We have investigated the electronic structure of three cyclic dipeptides: cyclo(Glycyl-Glycyl) (cGG), cyclo(Leucyl-Prolyl) (cLP), and cyclo(Phenylalanyl-Prolyl) (cPP). These compounds are biologically active and cLP and cPP are derived from cGG (also known as diketopiperazine), by the addition of the respective functional groups of the amino acids, namely, phenyl, alkyl or a fused pyrrolidine ring (proline). Experimental valence and core level spectra have been interpreted in the light of theoretical calculations to identify the basic chemical properties associated with the central ring, and with the additional functional groups in cLP and cPP. The theoretically simulated spectra of all three cyclic dipeptides in both valence and core spaces agreed reasonably well with the experimental spectra. The three molecules displayed similarities in their core spectra, suggesting that the diketopiperazine structure plays an important role in determining the inner shell spectrum. The experimental C 1s spectra of cLP and cPP are analogous but differ from cGG due to the side chains attached to the diketopiperazine structure. Single spectral peaks in the N 1s (and O 1s) spectra of the dipeptides indicate that the chemical environment of the nitrogen atoms (and oxygen atoms) are very similar, although they show a small splitting in the simulated spectra of cPL and cPP, due to the reduction of their point group symmetry. Valence band spectra of the three dipeptides in the frontier orbital region of 9–11 eV exhibit similarities; however theoretical analysis shows that significant changes occur due to the involvement of the side chain in the frontier orbitals of cPP, while lesser changes are found for cLP. © *2010 American Institute of Physics*.
[doi:10.1063/1.3499740]

I. INTRODUCTION

Peptides are an important class of biomolecules with a multitude of functions in living organisms. The key chemical motif of peptides is the peptide bond, and individual peptides are distinguished by the number of amino acid residues they contain, and their respective functional side chains. A subset of this class of compounds consists of cyclic peptides, many of which are biologically active, and which are made up of a ring containing two (for a dipeptide) or more peptide bonds: the simplest is cyclo(Glycyl Glycyl) or diketopiperazine. The ring structure confers high stability[1] and indeed resistance to human digestion, so that dipeptides have been considered as scaffolds by drug designers. Many such compounds have been identified which possess biological activity including antiviral, antibiotic and anti tumor properties, and show promise as environmentally friendly marine anti fouling agents.[2,3] Due to their unique spatial conformations, their simplicity and limited conformational freedom,[4] cyclic peptides are widely used as model molecules for larger peptides in chemistry, biochemistry, pharmaceutical chemistry, and life sciences.[2] Structure dictates function, but many intrinsic properties of biomolecules are masked by their environment or by their interactions with it. However structures and properties of biomolecules in the gas phase provide us with an understanding of their intrinsic properties in the absence of the perturbations due to their interactions with their environment. The aim of this study is to determine the structure-property relationship of a number of cyclic dipeptides, using the experimental[5] and theoretical methods previously applied to modeling both valence and inner-shell electronic states of biomolecules.[6]

The compounds selected for study, Fig. 1, are the simplest cyclopeptide, cyclo(Glycyl-Glycyl), denoted diketopiperazine or cGG hereafter; cyclo(Leucyl-Prolyl), which has an alkyl group bonded at the C(9) carbon site and a fused pyrrolidine ring connected at the C(6)—N(2) site, denoted cLP; and cyclo(Phenylalanyl-Prolyl), which contains a phenyl side chain in place of the alkyl group of cLP, denoted cPP. The six member piperazine ring in cGG can be flexible as it is formed by single bonds, whereas in cLP and cPP, the C(6)—N(2) bond is shared between the hexagonal piperazine ring and the pentagonal pyrrolidine ring, making the structure more rigid.

Previous conformational studies of cyclo dipeptides in-

[a)]Present address: Aarhus University, Department of Physics and Astronomy, Ny Munkegade 120, 8000 Aarhus C, Denmark.
[b)]Electronic mail: prince@elettra.trieste.it.

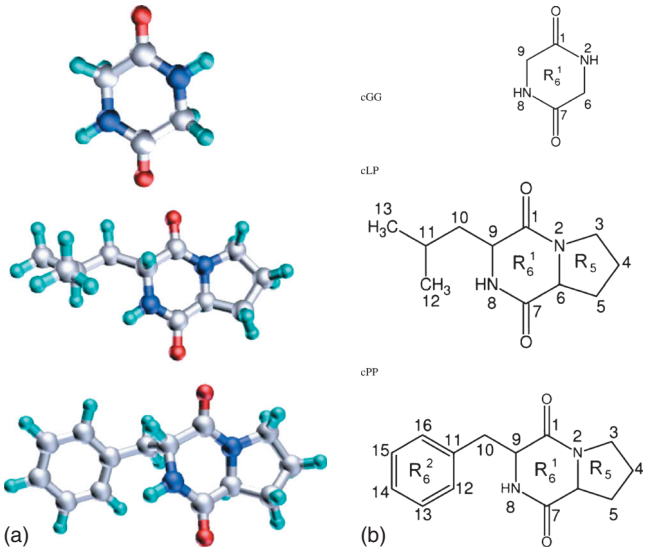

FIG. 1. Three dimensional view (a) and two dimensional (2D) view (b) of the optimized structures of cyclo(Glycyl-Glycyl), (cGG), cyclo(Leucyl-Prolyl), (cLP), and cyclo(Phenylalanyl-Prolyl), (cPP) with nomenclature. Double click the figure [part (a)] to view the interactive 3D structures.

clude theoretical studies of cyclo(His-Pro),[7] cyclo(His-His),[8] a range of dipeptides,[9] Nuclear Magnetic Resonance studies,[10] laser studies of cyclo(Phe-Ser),[11] Raman and vibrational spectroscopic studies.[12] Early studies of dipeptides[13] found that the hexagonal diketopiperazine ring formed by a pair of amino acid residues, cyclo(Thr-His), is slightly nonplanar. However, a study of the conformation of cyclo(Phe-His) in a variety of media including in vacuum, water and chloroform[14] found that the diketopiperazine ring of the preferred conformation of cyclo(Phe-His) was planar. In the solid state,[15] the diketopiperazine ring of cGG is almost planar, and the angles within the ring vary from 115° to 126°, even though the ring contains no formal double bonds. The C—C and C—N distances are shorter than the normal values for linear compounds, implying that there is significant conjugation in the ring. In a second approximation, the ring is not quite planar, but can be considered as two nearly coplanar half molecules with a slight twist. In a most recent density functional theory (DFT) study, Zhu *et al.*[9] found that the hexagonal rings of a number of dipeptides consisting of identical amino acids are preferably in boat configurations. For example, the cyclo(Gly-Gly) boat configuration is 30 kJ mol^{-1} more stable than its chair configuration. The properties and bioactivities of the cyclic dipeptides obviously depend on the electronic structures of the compounds. However, little information about the electronic structures of the dipeptides is available.

The four amino acids used to construct these dipeptides are proline (Pro), glycine (Gly), phenylalanine (Phe) and leucine (Leu), and the core level spectra and accompanying theoretical calculations of the first three of these have been reported recently,[5,16] as well as the spectra of the noncyclic peptide Glycyl-Glycine.[17] Here we investigate cyclic dimers of these amino acids, containing the diketopiperazine ring. In cPP and cLP, the pentagonal pyrrolidine ring is fused to the diketopiperazine ring to produce a rather rigid double ring structure. In proline, the four lowest energy conformers are constructed from two structural elements: the rotation of the carboxylic acid group, and the puckering of the pyrrolidine ring.[18] The carboxylic group is not present in the dipeptide, so we expect that the sole structural motif will be the puckering of the proline ring to an "up" or "down" geometry with respect to the twist in the piperazine ring, giving rise to two possible conformers. In addition to the fused piperazine and pyrrolidine ring structure, cPP contains a phenyl group, and cLP contains an alkyl (isobutyl) chain as a side group.

II. EXPERIMENTAL METHODS

The measurements were performed at the Gas Phase Photoemission beamline, Elettra Trieste, using apparatus and calibration methods described previously.[5] The cGG samples were supplied by Sigma Aldrich and the cPP and cLP from Bachem and used without further purification. They were evaporated at temperatures of 405, 385, and 365 K, respectively, and checked for signs of thermal decomposition (spectral changes as a function of time, discoloration after heating, etc.) No evidence was found for decomposition of the compounds.

III. COMPUTATIONAL METHODS

All geometry optimizations were performed using the B3LYP/cc-pVTZ model, which is incorporated in the GAMESS computational chemistry package,[19] followed by harmonic vibrational frequency calculations using the Gaussian03 computational package.[20] The optimized geometries of the species are true minimum configurations without any imaginary frequencies. Detailed instructions on generation of the interactive three dimensional (3D)-pdf structures for the molecules shown in Fig. 1 are provided by Selvam *et al.*[21] Single point calculations were then carried out based on the LB94/et-pVQZ model[22] and Statistical Averaging of (model) Orbital Potentials (SAOP)/et-pVQZ model,[23] which are incorporated in the Amsterdam Density Functional computational chemistry package[24] to produce core and valence ionization spectra, respectively. Vertical orbital ionization potentials (IPs) were produced using the "meta-Koopman's theorem"[25] without further modification and scaling. In addition, the outer valence vertical ionization energies of the dipeptides were calculated using the outer valence Green function OVGF/6-31G* model.[26] The value of the DFT approaches for calculating properties of larger, low symmetry molecules has been discussed by Stener *et al.*,[27] while the problems associated with calculating core holes with the present methods have been addressed in Refs. 28 and 29.

IV. RESULTS AND DISCUSSION

Table I compares the calculated geometric and electronic properties of the three dipeptides, using the B3LYP/cc-pVTZ model, as well as with the experimental crystal structure[30] and other theoretical results.[9] R_6^1 is the perimeter of the piperazine ring and includes the dipeptide bonds, and while R_6^2 is the perimeter of the phenyl ring.[31] The former contains single bonds only but the latter is aromatic. Isotropic geometric parameters, such as the ring perimeters,[31] R_6^1 or R_5

TABLE I. Geometric and electronic properties of the dipeptides.

		cGG			cLP		cPP	
	Molecule	cc-pVTZ[a]	6-31G[a,b]	Crystal str.[c]	cc-pVTZ[a]	Crystal str.[d,c]	cc-pVTZ[a]	Crystal str.[d,c]
Ring perimeter (Å)	R_6^1	8.66	8.68	8.59	8.68	8.65	8.68	8.60
	R_6^2						8.36	8.27
	R_5				7.54		7.54	7.50
Angles (°)	C(9)C(1)N(2)	116.10	115.62	115.1	116.28	110.9	116.01	117.2
	C(1)N(2)C(6)	126.14	125.64	126.0	125.64	113.5	125.37	126.8
	N(2)C(6)C(7)	114.42	113.96	118.9	114.00	114.0	113.92	113.8
	C(9)C(10)C(11)				115.15		113.06	113.4
	C(6)N(2)C(3)				112.15		112.25	110.8
	N(2)C(6)C(5)				102.62		102.82	102.0
Dihedral angles (°)	C(1)N(2)C(9)N(8)	−166.32	−163.02		−155.83		−154.44	
	C(7)N(8)C(6)N(2)	−166.32	−163.02		−160.96		−160.33	
	C(1)N(2)C(6)C(7)	−24.16	−28.54	−1	−23.72	−41.5	−24.96	−17.1
	C(7)N(8)C(9)C(1)	−24.16	−28.54	−1	−26.70	−41.5	−31.82	−19.0
	C(9)C(1)N(2)C(6)	8.42	9.05	1	−0.89	6.2	−0.73	0.7
	C(6)C(7)N(8)C(9)	8.42	9.05	1	5.67	6.3	6.65	2.8
Pseudorotation phase (P)(°)					18.93		1.96	
Puckering amplitude (υ_{max})(°)					38.62		62.47	
Dipole moment (Debye)		0.69	0.84		1.47		1.54	
$\langle R^2 \rangle$(a.u.)		917.18			3731.81		5371.11	

[a]The theoretical calculations have been conducted on the basis of the B3LYP model.
[b]Reference 9.
[c]Reference 30.
[d]Estimated standard deviations are in the range of 0.002–0.004 Å for bond lengths and 0.015°–0.03° for bond angles.

(Ref. 31) remain almost unchanged, regardless of the leucine or phenylalanine side groups. For example, the perimeter of the piperazine ring changes by only 0.02 Å from cGG with a single ring to cLP and cPP, whereas R_6 and R_5 of the bicyclic dipeptides cLP and cPP are equal, indicating that the dipeptide side chains do not affect the hexagonal and pentagonal rings.

Figure 1 displays the optimized structures of the three dipeptides using a recently developed interactive 3D pdf technique.[21] Double clicking the figure online or the portable document format (PDF) file on a computer will allow viewing of the embedded 3D structures in the pdf file. The interior angles describe the piperazine and pyrrolidine rings of the dipeptides and do not change significantly; there is a common C(6)—N(2) bond in cLP and cPP. Also the dihedral angles do not change significantly, except for the C(9)C(1)N(2)C(6) angle, which differentiates the point group symmetry of cGG (C_2) from cLP and cPP (C_1). It was found previously[9] that the structure of cGG is a boat configuration, which is also confirmed in the present study. We further found that the dihedral angle in cGG between the flag side and the bottom side of the boat, i.e., C(1)N(2)C(9)N(8) is −166.32° in cGG, whereas this angle in cLP and cPP is −155.83° and −155.44°, respectively. Due to the C_1 symmetry, the piperazine boat is not symmetric in cLP and cPP, so that the dihedral angle of the other side of the boat, i.e., C(7)N(8)C(6)N(2), is −160.96° and −160.33°, respectively, for cLP and cPP, indicating a larger distortion of the hexagonal ring in cLP and cPP. Nevertheless, the boat structure in cLP and cPP is very similar, regardless of the phenyl or leucine group.

The dipeptides cLP and cPP contain a five member pyrrolidine ring with a side shared with the piperazine ring. Five member rings usually undergo pseudorotation,[32] which can be defined by a pair of parameters such as the pseudorotation phase, P, and the pseudorotational amplitude, υ_m.[33] The pseudorotation of the pyrrolidine ring was examined, with N(2) being the analog of O in tetrahydrofuran in the cited work. The values of the pseudorotational phase parameter, P, of cLP and cPP, also given in Table I, are 1.96° and 18.93°, respectively, approximately the $^3_2$T (twist) and ^{3}E (envelope) configurations in the pseudorotational wheel,[34] both in characteristic *North* conformation. The pseudorotational amplitude, υ_m of cLP is nearly half of that of cPP, indicating that the phenyl ring in cPP indeed causes larger puckering of the pyrrolidine moiety in the dipeptide cPP. However, as both cLP and cPP have $\upsilon_m > 20°$,[34] namely, 38.62° and 62.47°, respectively, the pyrrolidine rings of cLP and cPP are puckered rather than flat as shown in Fig. 1. For the pyrrolidine ring of free proline, on the other hand, P=115.81°, which locates free proline between the twist $^0_1$T and envelope $_1$E, for an *East* conformation and 55.72° for its amplitude υ_m. The proline moiety of the bicyclic dipeptides changes the pseudorotational phase significantly, in order to coordinate with the hexagonal diketopiperazine ring. However, the pseudorotational amplitude, υ_m, of cLP and cPP responds to the side groups in opposite ways, indicating that the side functional groups, isobutyl or phenyl, impact differently on the proline moiety in cLP and cPP.

The potential energy has been calculated as a function of rotation of the flexible dihedral angle δ =C(11)C(10)C(9)N(8) of cLP and cPP, respectively, using the potential energy scan technique. Two minimum energy conformations are found for each of the dipeptides, that is, at

TABLE II. Vertical valence binding energies of cGG, cLP, and cPP dipeptides using outer valence Green function theory and the DFT based SAOP model. For the OVGF model, pole strengths are given in parentheses; only those with values >0.85 are shown. The orbitals calculated at higher binding energies are listed in the supporting information.

cGG			cLP			cPP		
	IP (eV)			IP (eV)			IP (eV)	
Orbital	SAOP[a]	OVGF[b]	Orbital	SAOP[a]	OVGF[b]	Orbital	SAOP[a]	OVGF[b]
15b[c]	10.86	9.34 (0.90)	57a[c]	10.46	8.53 (0.90)	65a[c]	10.42	8.60 (0.90)
15a	11.06	9.69 (0.91)	56a	10.56	8.91 (0. 90)	64a	10.48	8.89 (0.90)
14b	11.35	10.22 (0.90)	55a	11.08	9.70 (0.90)	63a	10.77	8.54 (0.90)
14a	11.51	10.26 (0.90)	54a	11.20	9.87 (0.90)	62a	10.84	9.06 (0.90)
						61a	11.11	9.73 (0.89)
						60a	11.26	9.88 (0.89)

[a]SAOP/et-pVQZ.
[b]OVGF/6-31G*.
[c]HOMO.

$\delta=61.69°$ and 185° for cLP, and for cPP $\delta=65.60°$ and 300°, separated by high energy barriers of 35.83 and 41.16 eV, respectively. The global ($\delta<100°$) and local ($\delta>100°$) minimum energy structures of cLP and cPP differ by 7.38 and 7.39 eV in energy, respectively. Thus the present study focuses on the global minimum energy configurations of cLP and cPP.

Table II gives the calculated vertical ionization potentials of the dipeptides in the outer valence space of the dipeptides in the energy region of interest here, namely, the outer 4 or 6 molecular orbitals. The full list of calculated values is listed in the supplementary material.[35] The spectroscopic pole strengths calculated using the OVGF/6-31G* model are all larger than 0.85, indicating that the single particle approximation used in the models is valid and appropriate in this study. In a previous calculation of outer valence vertical ionization potentials of benzene, for which excellent and well resolved experimental results are available,[36] the performance of the SAOP model was excellent except for the first ionization potential; this is often the case for biomolecules.[36] Contributions from satellite states are negligible in this region of interest. As seen in this table, the agreement between the SAOP and outer valence Green function theory (OVGF) models apparently improves at larger binding energy, which was also found in benzene.[36]

Figure 2 compares the measured and simulated vertical ionization spectra (both SAOP and OVGF) of the dipeptides in the outer valence region. The models used in the simulations reproduce the major peaks of the experiment in the three dipeptides. Because the compounds are related structurally, they exhibit certain similarities, in particular a common structure in the binding energy range of approximately 9–11 eV. Although the spectra are not well resolved in the experiment, the frontier occupied orbitals in the region of 9–11 eV are well separated from the other valence orbitals and are reproduced by the simulated spectra well. The OVGF model shows significantly better agreement with the experimental spectra in this energy range. The Highest Occupied Molecular Orbital (HOMO)-Lowest Unoccupied Molecular Orbital (LUMO) gap has been used as a simple indicator of kinetic stability, that is, stability with respect to the activated

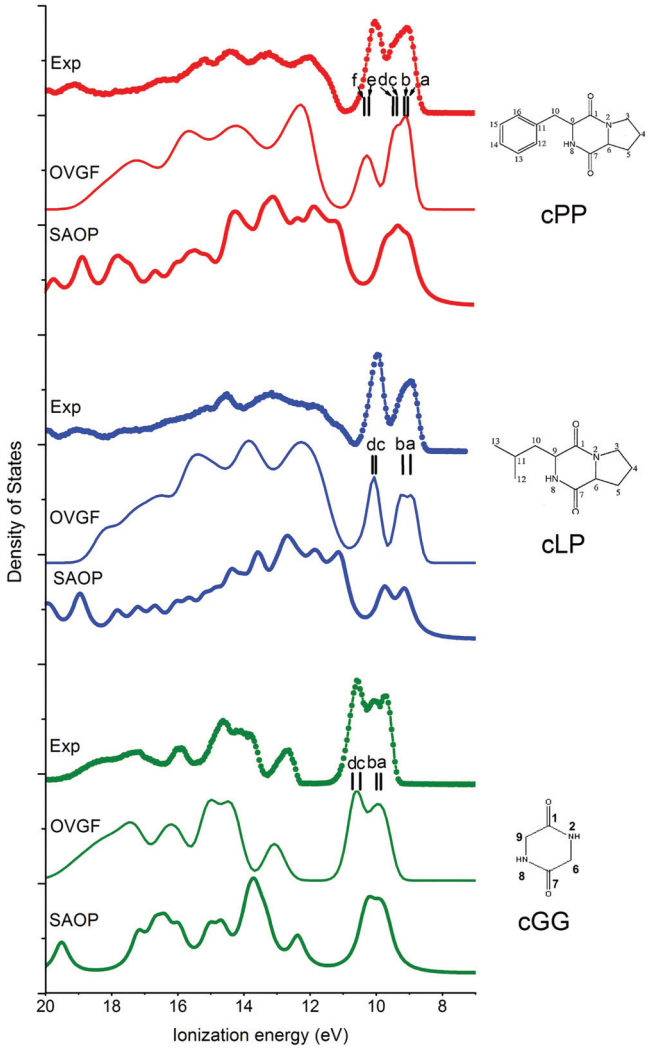

FIG. 2. Valence ionization spectra of cPP (upper), cLP (middle), and cGG (lower). Dotted lines: experiment; thin lines: theory based on the OVGF/6-31G* model (spectra are shifted by +0.46, +0.36, and +0.38 eV, respectively, for cPP, cLP, and cGG); thick lines: theory based on the SAOP/et-pVQZ model [spectra are shifted by −1.50 eV (cPP), −1.39 eV (cLP), and −1.16 eV (cGG)]. The FWHMs are adjusted in the simulation to best match the experimental spectra. The HOMO-LUMO gaps are: 5.28 (cGG), 5.13 (cLP), and 4.67 eV (cPP) based on SAOP/et-pVQZ model. The frontier valence orbital vertical ionization energies (OVGF/6-31G*) are marked a, b, c, d, e, and f using small vertical bars on the spectra.

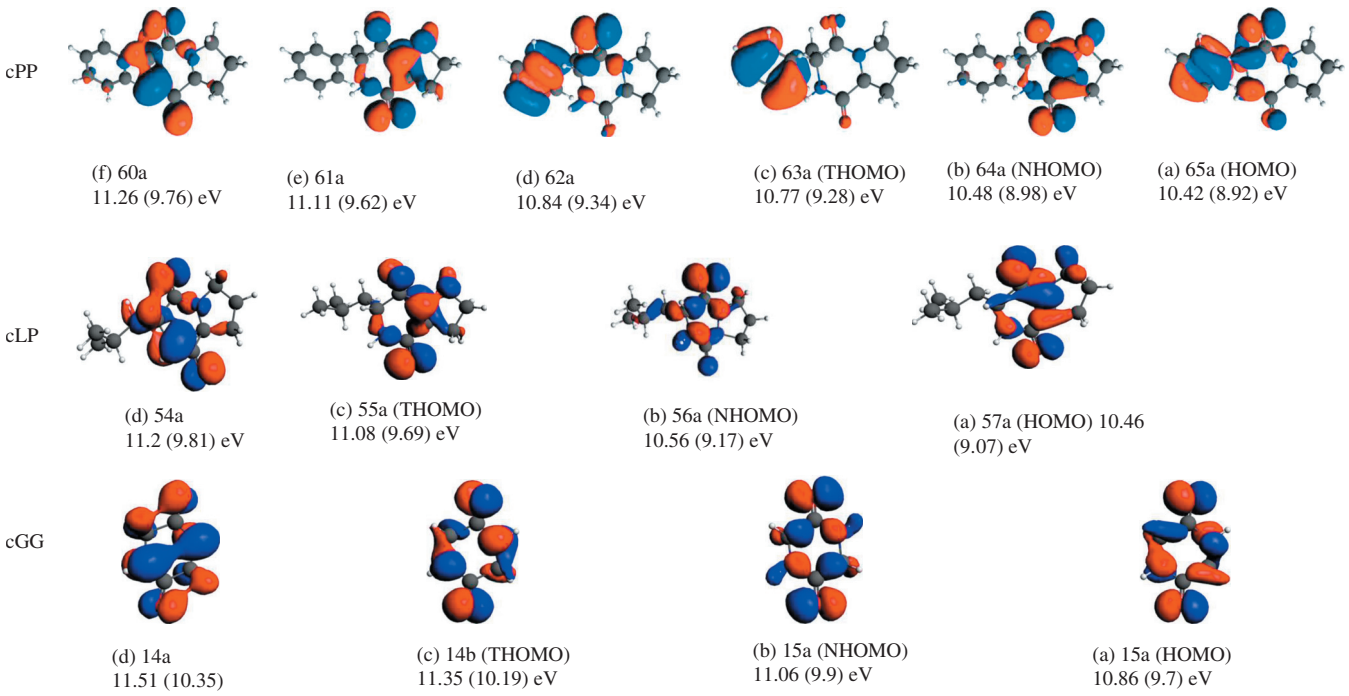

FIG. 3. The frontier valence orbitals of the dipeptides marked a–f in Fig. 2.

complex of any further chemical reactions.[37] The HOMO-LUMO gap of the dipeptides decreases as the size of the side group becomes larger. For example, the HOMO-LUMO gap is 5.28 eV in cGG, which decreases to 5.13 eV in cLP and to 4.67 eV in cPP, an energy drop of 0.15 eV from cGG to cPP but of 0.61 eV from cGG to cPP. A large HOMO-LUMO gap implies high kinetic stability and low chemical reactivity, because it is energetically unfavorable to add electrons to a high-lying LUMO to extract electrons from a low-lying HOMO, and so to form the activated complex of any potential reactions.[37]

The spectral peaks in the region of 9–11 eV shown in Fig. 2 are due to four (cGG and cLP) and six (cPP) frontier orbitals. The calculated vertical ionization energies of the frontier orbitals are marked as small vertical bars above each of the simulated spectra in Fig. 2, and the orbitals that are responsible for these spectral peaks are given in Fig. 3. While 4 frontier orbitals are responsible for the spectral peaks in the 9–11 eV region in cGG and cLP, for cPP there are 6 because two highest occupied molecular orbitals of the phenyl side chain appear in this energy range or are mixed with the diketopiperazine states. The calculated energies of the spectra are shifted to lower binding energy to match the experimental spectrum, by an average of −1.35 eV for the SAOP calculations, and by +0.40 eV for the OVGF calculations. The valence spectra show three peaks for cGG and two for cLP and cPP at lower binding energy. It is a known fact that for biomolecules,[21,36] the OVGF theory produces IPs that are closer to the measured value than the SAOP theory in the outer valence region. For example, in the energy region of 9–11 eV, the SAOP model underestimates the splitting for cGG, and predicts two partly resolved peaks, while the OVGF model gives a better simulation of spectra in this region. Similarly for cLP the SAOP model underestimates the separation of the two peaks, while the OVGF model correctly predicts three resolved peaks. For cPP, the separation of the peaks is again underestimated by the SAOP model but well reproduced by the OVGF model, as the OVGF model is based on the one-particle Green function and is applicable to the outer valence region. The OVGF model becomes unavailable when the hole state moves inwards where the SAOP model provides more accurate results and the computational costs either prevent the OVGF model from being applied to larger molecules or restrict it to small and often insufficient large basis sets. In their Table II, Ganesan *et al.*[36] compared the valence ionization energies of benzene with best available theoretical and experimental results and Fig. 4 of Selvam *et al.*[21] compared the valence ionization potentials produced using a number of models including the OVGF and SAOP models for larger molecules such as nucleosides.

In cGG, the orbitals 15b (HOMO) and 15a (NHOMO, the next HOMO) contribute to the first and second peak, and 14b (THOMO, the third HOMO) and 14a (the fourth HOMO) contribute to the third experimental peak, as indicated by the small vertical bars labeled a, b, c, and d. The orbitals are dominated by the 2p electron lone pairs on the oxygen atoms of the two keto C═O moieties. From a local symmetry point of view, the HOMO (15b) and NHOMO (15a) of cGG, are dominated by in-plane p electrons of O, N, and C, implying σ and σ^* bonding in cGG, whereas the THOMO (15a) and fourth HOMO 14a are dominated by the out-of-plane p electrons that are potentially responsible for π^* and π bonding. In addition, the HOMO (15b) and LUMO (16a) of cGG exhibit different orbital symmetry and the lowest lying virtual orbital with the same symmetry as the HOMO (15b) is the second LUMO (16b) leading to a higher energy gap of 5.28 eV for cGG.

In cLP, the contributions to frontier orbitals are again

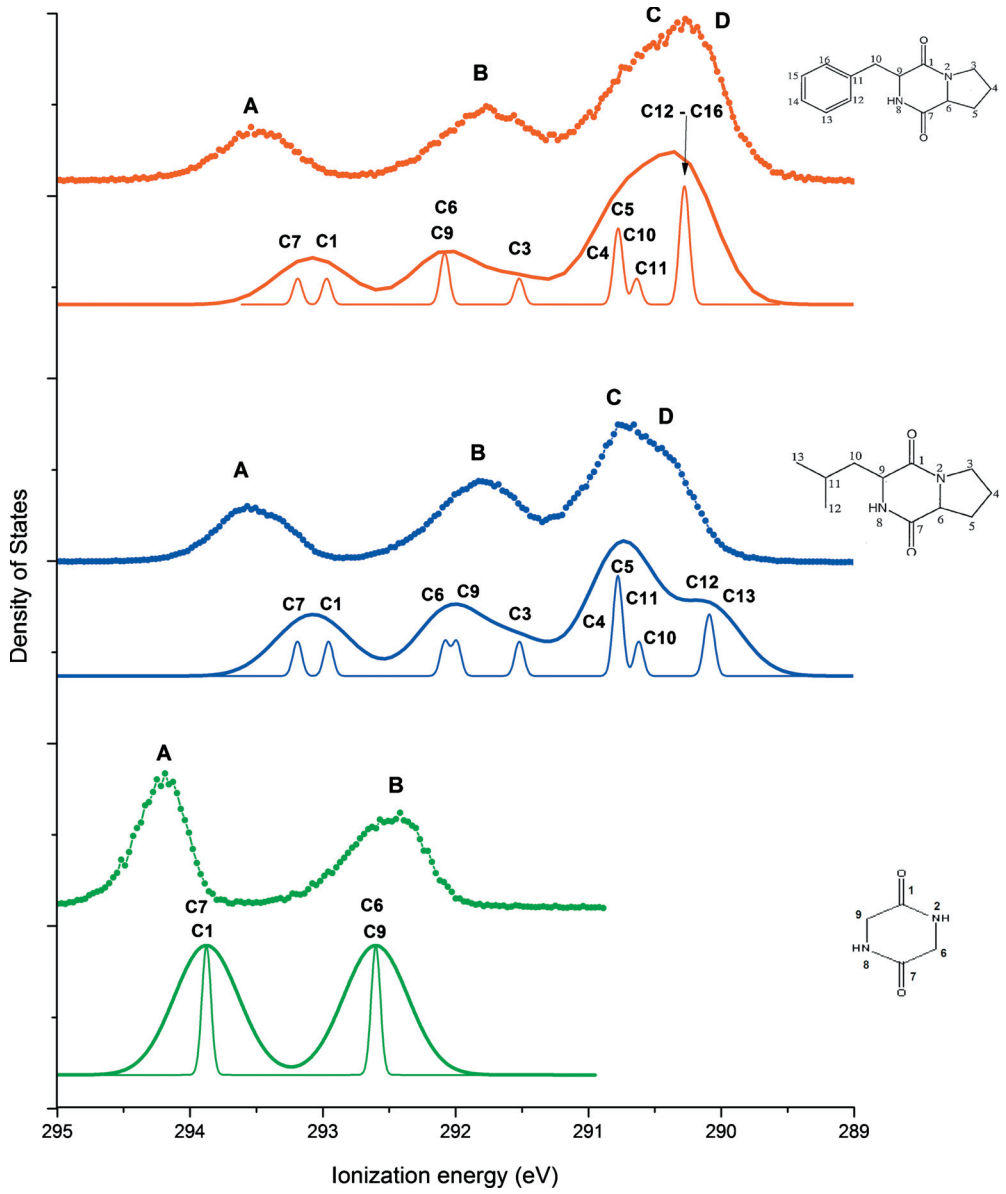

FIG. 4. C 1s spectra of cGG (lower), cLP (middle) and cPP (upper curve). Dots and lines: experiment; lines: theory at the LB94/et-pVQZ level. Thick lines: Gaussian broadening of FWHM=0.57 eV; thin lines: broadening of FWHM=0.08 eV. The theoretical spectra have been shifted in energy by +0.98 eV for cGG, and +0.62 eV for cLP and +0.58 eV for cPP.

dominated by the diketopiperazine ring with a small contribution from the leucine side chain (56a) and stronger contributions from the proline ring (57a, 55a, and 54a). For example, the HOMO (57a) and THOMO (55a) have substantial orbital character from the proline ring whereas the NHOMO and 4th HOMO receive small contributions from the isobutyl side chain of cLP. All these frontier orbitals in cLP show similarities to their respective orbitals in cGG, although the orbitals in cLP exhibit a certain distortion due to the reduction of its point group symmetry. The top keto C(1)=O in the frontier orbitals of cLP exhibits π-like bonding character whereas the bottom keto C(7)=O exhibits in-plane σ-like symmetric bonding character, except for the NHOMO (56a) that is dominated by σ^* bonding character.

The frontier orbitals of cPP are, however, very different, although the orbitals are dominated by the diketopiperazine ring and retain certain similarities to those of cLP. The most significant differences are (a) six frontier orbitals contribute to the spectral peaks at 9–11eV, rather than four orbitals in cLP and cGG, and (b) three phenyl dominant frontier orbitals in cPP, including the HOMO (65a), THOMO (63a) and fourth HOMO (62a) as shown in Fig. 3. The latter makes cPP significantly different from cGG and cLP as the frontier orbitals in cGG and cLP are all dominated by the diketopiperazine ring. The three phenyl dominant orbitals in cPP stem from the doubly degenerate HOMO ($1e_{1g}$) of benzene interacting with the diketopiperazine moiety. The phenyl dominant HOMO (and LUMO) in cPP explains the significantly larger decrease of the HOMO-LUMO gap in cPP among the dipeptides. The THOMO (63a) of cPP is almost entirely dominated by the phenyl π-orbital, whereas the HOMO (65a) and the fourth HOMO (62a) show interaction with the top keto C(1)=O(1), which is located closer to the phenyl moiety and therefore enhances the intramolecular interaction,

TABLE III. Experimental and theoretical (LB94/et-pVQZ) C 1s core level binding energies and intensities.

Sample	Expt., (eV, ±0.05 eV)	Intensity (stoichiometric ratio)	Theor. energies (eV)	Hirshfeld charges (a.u.)
cGG	A 294.20	2 (2)	C(7) 292.90 (B), C(1) 292.90 (A)	C(7), C(1) 0.169
	B 292.50	1.93 (2)	C(6) 291.62 (B), C(9) 291.62 (A)	C(6), C(9) −0.006
cLP	A 293.52	1.84 (2)	C(7) 292.60, C(1) 292.33	C(7) 0.164, C(1) 0.157
	B 291.79	3.4 (3)	C(6) 290.16, C(9) 291.37, C(3) 291.46	C(6) 0.024, C(9) 0.027, C(3) −0.01
	C 290.63	5.0 (4)	C(4) 290.89, C(5) 290.14, C(11) 290.16, C(10) 290.00	C(4) −0.061, C(5) −0.059, C(11) −0.016, C(10) −0.065
	D 290.27	0.74 (2)	C(12) 289.45, C(13) 289.48	C(12) −0.113, C(13) −0.115
cPP	A 293.40	1.9 (2)	C(7) 292.60, C(1) 292.39	C(7) 0.164, C(1) 0.157
	B 291.74	3.4 (3)	C(6) 291.50, C(9) 291.50, C(3) 290.94	C(6) 0.024, C(9) 0.03, C(3) −0.009
	C 290.60	7.0 (7)	C(4) 290.20, C(5) 290.19, C(10) 290.18, C(11) 290.05, C(13) 289.71, C(15) 289.70, C(12) 289.70	C(4) −0.061, C(5) −0.059, C(10) −0.064, C(11) 0.008, C(13) −0.043, C(15) −0.043, C(12) −0.047
	D 290.20	1.7 (2)	C(16) 289.69, C(14) 289.67	C(16) −0.045, C(14) −0.045

through the C(10)—C(9) bridge. The bottom C(7)=O(7) keto in cPP plays a minor role in these phenyl dominated orbitals. The other frontier orbitals are composed of contributions from the diketopiperazine and proline rings, almost in one-to-one correspondence with cLP. For example, apart from the leucine influenced NHOMO (56a) in cLP, the HOMO (57a), THOMO (55a) and fourth HOMO (54a) of cLP are similar to the NHOMO (64a), fifth HOMO (61a), and sixth HOMO (60a) in cPP.

At higher binding energy than about 11 eV, the OVGF calculations match the overall shape of the cGG spectrum better than SAOP, while the latter method seems to be a little better for cLP and cPP. However given the complexity and high density of states in the spectra we do not enter into a detailed discussion here.

In Fig. 4, we show the experimental and theoretical C 1s photoemission spectra of the three compounds, and the data is summarized in Table III, together with their Hirshfeld charges calculated using the same model. The spectrum of cGG is relatively simple and consists of two peaks, due to core emission from carbon atoms in the C=O and peptide chemical environments, as assigned by our simulated spectrum. A global energy shift of +0.98 eV for cGG, +0.58 eV for cLP and +0.62 eV for cPP has been applied to the simulated spectra in order to correct systematic errors.[17,38] The agreement with theory is satisfactory. The DFT LB94/et-pVQZ model is known to underestimate the core ionization potential of carbon[39] when doubly bonded with O as in the keto group (Table IV). Nevertheless, the spectral peak A at 294.20 eV of cGG is clearly assigned to the carbon atoms doubly bonded to oxygen, C(1)=O, C(7)=O, whereas peak B at 292.50 eV is assigned to the carbon atoms singly bonded to nitrogen and carbon that is, C(6)—N and C(9)—N. The Hirshfeld charges given in the same table support such an assignment. In addition, the ratio of intensity is 2:1.93, very near to the stoichiometric ratio of 2:2, and the small discrepancy is due to the limitations of defining the peak tails, and probably also to different cross-sections for excitation of satellites.

Cyclic cGG is quite different from the linear dipeptide Gly-Gly.[17] All the carbon atoms in linear Gly-Gly (Ref. 17) are in different chemical environments due to its C_1 point group symmetry, although the amino carbon atoms, C(6)—N and C(9)—N, have only small chemical shifts that cannot be resolved experimentally.[17] Their energy, 292.50 eV, is very similar to that of the corresponding carbon atoms in cGG, 292.32 eV. The carbon atom bonded to O in the linear peptide has a binding energy of 293.85 compared with 294.20 eV in the cyclic compound, which is also a small difference.

For cLP and cPP, the agreement with theory is quite satisfactory and the overall shapes of the spectra are well reproduced, although the energy of peak A from C(1) and C(7) is still underestimated. In cLP, the separation of the main peak C from its shoulder D is slightly overestimated, and the positions of peaks B and C are shifted from the experimental positions. A similar effect is seen for these features in the spectrum of cPP. The three C 1s spectra exhibit certain similarities. Although they are energy shifted, peaks

TABLE IV. Experimental and theoretical (LB94/et-pVQZ) N and O 1s core level binding energies.

Sample	Expt. N 1s (eV, ±0.05 eV)	Theor. N 1s (eV)	Hirshfeld charges (a.u.)	Expt. O 1s (eV, ±0.05 eV)	Theor. O 1s (eV)	Hirshfeld charges (a.u.)
cGG	406.29	404.644 (B) N(2)	−0.108	537.45	534.289 (B) O(1)	−0.331
		404.644 (A) N(8)	−0.108		534.289 (A) O(7)	−0.335
cLP	405.53	404.343 N(2)	−0.062	536.95	533.867 O(1)	−0.334
		404.276 N(8)	−0.110		534.046 O(7)	−0.331
cPP	405.53	404.40 N(2)	−0.062	536.95	533.93 O(1)	−0.333
		404.30 N(8)	−0.111		534.09 O(7)	−0.331

A and B in cLP and cPP are dominated by the diketopiperazine moiety, with peaks A due to the doubly bonded carbons, C(1) and C(7), and peaks B mostly due to the singly bonded carbon atoms of the diketopiperazine moiety, C(6) and C(9). Peak B in cLP and cPP has an asymmetric shape, due to a contribution from C(3) in the pyrrolidine moiety, approximately 0.5 eV away from the C(6) and C(9) peaks. This C(3) site has a chemical environment that is similar to the other single C—N bonded sites, C(6) and C(9). Therefore, peaks B in the cyclic dipeptides are composed of contributions from the singly bonded carbon atoms bonded to nitrogen atoms and the asymmetric shape of B indicates the difference in the chemical environments of carbon atoms in the diketopiperazine and pyrrolidine moieties.

The peaks C and D of the bicyclic compounds are due to contributions of the side chain functional groups. The positions of the peaks in cLP and cPP are not the same, and the asymmetry of peaks C and D is opposite. For example, in cLP, C is the main peak and D represents the shoulder on the lower energy side. The shoulder D is associated with the isobutyl side chain C(12) and C(13), i.e., the two methyl groups at the end of the leucine functional group. However, in cPP, the peak D, which is located at the higher energy side of the spectrum, becomes the main peak. A more detailed inspection of the simulated spectra reveals that in fact, the main peak C in cLP becomes the shoulder in cPP as five of the phenyl carbon atoms contribute to the larger peak D at the low energy peak. In the cLP spectrum, the main peak C is due to the bridge carbon atoms: C(10) and C(11) of the leucine group, and C(4) and C(5) of the proline group. In the case of cPP, the shoulder C has a similar origin to peak C of cLP.

The experimental and simulated N 1s core level spectra are shown in Fig. 5. Diketopiperazine possesses a center of symmetry[15] and thus the nitrogen atoms are symmetrically equivalent, and the core levels have the same binding energy. The asymmetry is therefore not due to the presence of different components, but is determined by the Franck–Condon envelope. The experimental N 1s energies of cLP and cPP are equal, and both are 0.8 eV lower than that of diketopiperazine, similar to the shift of the outer valence orbitals. In fact, the differences between the single cyclic cGG and bicyclic cLP and cPP are not just a simple global energy shift, but also splitting of the two nonequivalent nitrogen atoms in cLP and cPP. We attribute the shift to the effect of different chemical environment caused by the side functional groups. The theoretical N 1s energies of cLP and cPP are 1.20 eV lower than the experimental values, and theory predicts that the two nitrogen core levels are very close together, separated by 70 meV. In cPP, the binding energy splitting is slightly larger, 100 meV. Thus, the asymmetric line shape is almost entirely due to the Franck–Condon envelope.

The oxygen core level spectra are shown in Fig. 6 and their peak shapes are similar to but more nearly symmetric than the N 1s peaks. The predicted splittings for cLP and cPP are again small, 180 meV for cLP and 160 meV for cPP, and the theoretical energies are about 2.9 eV lower than the experimental values. As in the case of the nitrogen core levels, the O 1s binding energies of cLP and cPP are equal, and

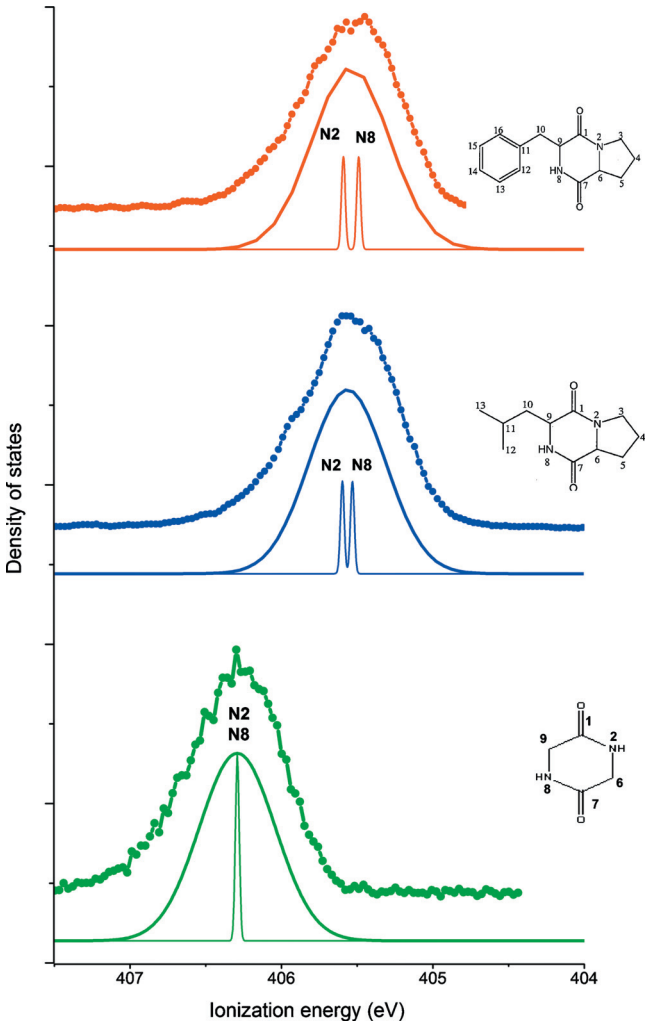

FIG. 5. N 1s spectra of cGG (lower), cLP (middle), and cPP (upper curve). Dots and lines: experiment; lines: theory at the LB94/et-pVQZ level. Thick lines: Gaussian broadening of FWHM=0.59 eV; thin lines: broadening of FWHM=0.03 eV. The theoretical spectra have been shifted in energy by +1.65 for cGG, +1.25 eV for cLP and +1.19 eV for cPP.

shifted from the value for diketopiperazine. The shift is smaller, 0.5 eV, and is again attributed to the effect of different chemical environments caused by the side functional groups. From the theoretical results, it appears that the functional groups of the side chains of the amino groups do not affect the piperazine ring strongly, as only small splittings are calculated.

V. CONCLUSIONS

Outer valence and core photoemission spectra of the three cyclic dipeptides cyclo(Glycyl-Glycyl), cyclo(Leucyl-Prolyl), and cyclo(Phenylalanyl-Prolyl) have been measured using synchrotron radiation soft x-ray spectroscopy in the gas phase. The spectra have also been simulated quantum mechanically using density functional theory, and agree reasonably well with the measurements. The present study correlates the observed spectra with the electronic structures of the cyclic dipeptides through the simulated spectra. The isotropic geometric properties of the dipeptides such as the perimeters of the diketopiperazine ring and pyrrolidine ring of proline remain almost unchanged. All the hexagonal dike-

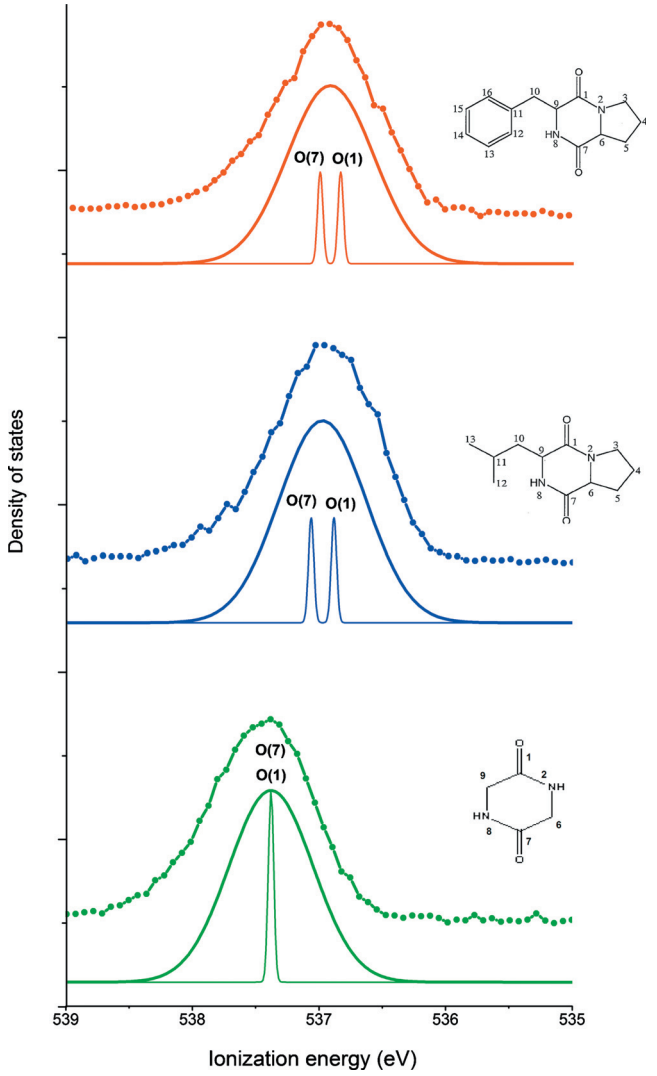

FIG. 6. O 1s spectra of cGG (lower), cLP (middle) and cPP (upper curve). Dots and lines: experiment; lines: theory at the LB94/et-pVQZ level. Thick lines: Gaussian broadening of FWHM=0.78 eV; thin lines: broadening of FWHM=0.05 eV. The theoretical spectra have been shifted in energy by +3.09 eV for cGG, +3.01 eV for cLP and +2.91 for cPP.

topiperazine rings are present in the boat conformation and the pyrrolidine ring has various degrees of puckering. The bicyclic dipeptides, cLP and cPP in fact enhance the boat configuration, compared to cGG. However, the pyrrolidine pseudorotational amplitudes split into opposite directions when compared to free proline.

Similarities in the outer valence and core photoemission spectra of the cyclic dipeptides are observed, indicating that the diketopiperazine structure dominates their electronic structures. In the outer valence space, a group of frontier orbitals in the energy region of 9–11 eV is well separated from other valence orbitals. The HOMO-LUMO gap decrease reflects changes in their chemical bonding characters. More importantly, although similar in their IPs and spectra, the energy region of 9–11 eV of cLP and cPP reveals significant bonding character changes from minor contributions of the alkyl side chain in cLP (orbital 56a) to dominant aromatic phenyl side chain contributions in half of the six frontier orbitals in cPP including the HOMO (orbitals 65a) and orbitals 63a and 62a. For the C 1s spectra, the major spectral peaks at higher binding energy are dominated by the diketopiperazine ring, whereas the asymmetry and the large peaks at lower energy are contributions of side isobutyl, phenyl and pyrrolidine groups. For the N 1s and O 1s spectra, the measured spectra present one peak, while theory indicates that the peaks have weak splittings not resolved experimentally. The effects of the side isobutyl and phenyl functional groups on the N 1s and O 1s spectra are not the same. The splitting in the N 1s spectra of cLP is smaller than that of cPP, but is opposite for the O 1s spectra. For diketopiperazine, due to the symmetry, the two nitrogen and two oxygen atoms are symmetrically equivalent, so only a single peak is expected and found experimentally.

ACKNOWLEDGMENTS

We acknowledge the National Computational Infrastructure (NCI) at the Australian National University for an award under the Merit Allocation Scheme and Swinburne University Supercomputing Facilities, and Victorian Partnership for Advanced Computing (VPAC) for supercomputing facilities. One of the authors (A.P.W.A.) acknowledges a Swinburne University Centenary Postgraduate Research Award.

[1] P. Wipf, Chem. Rev. (Washington, D.C.) **95**, 2115 (1995).
[2] C. Prasad, Peptides **16**, 151 (1995).
[3] M. B. Martins and I. Carvalho, Tetrahedron **63**, 9923 (2007); K. H. Rhee, J. Gen. Appl. Microbiol. **48**, 321 (2002); M. Li, Ph.D. thesis, Hong Kong University of Science and Technology, 2006; http://repository.ust.hk/dspace/handle/1783.1/2759; T. F. N. Kwong, L. Miao, X. Li, and P. Y. Qian, Marine Biotechnology **8**, 634 (2006).
[4] P. J. Milne, A. L. Hunt, K. Rostoll, J. J. van der Walt, and C. J. Graz, J. Pharm. Pharmacol. **50**, 1331 (1998).
[5] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, and V. Carravetta, Chem. Phys. Lett. **442**, 429 (2007); J. Electron Spectrosc. Relat. Phenom. **155**, 47 (2007); J. Phys. Chem. A **111**, 10998 (2007).
[6] A. Thompson, S. Saha, F. Wang, T. Tsuchimochi, A. Nakata, Y. Imamura, and H. Nakai, Bull. Chem. Soc. Jpn. **82**, 187 (2009); S. Saha, F. Wang, J. B. MacNaughton, A. Moewes, and D. P. Chong, J. Synchrotron Radiat. **15**, 151 (2008); F. Wang, Q. Zhu, and E. Ivanova, *ibid.* **15**, 624 (2008).
[7] A. Abiram and P. Kolandaivel, J. Mol. Model. **16**, 193 (2010).
[8] A. Abiram and P. Kolandaivel, Mol. Simul. **35**, 409 (2009).
[9] Y. Zhu, M. Tang, X. Shi, and Y. Zhao, Int. J. Quantum Chem. **107**, 745 (2007).
[10] F. R. Lucietto, P. J. Milne, G. Kilian, C. L. Frost, and M. van de Venter, Peptides **27**, 2706 (2006).
[11] A. G. Abo-Riziq, B. Crews, J. E. Bushnell, M. P. Callahan, and M. S. De Vries, Mol. Phys. **103**, 1491 (2005).
[12] A. P. Mendham, T. J. Dines, M. J. Snowden, B. Z. Chowdhry, and R. Withnall, J. Raman Spectrosc. **40**, 1478 (2009); A. P. Mendham, T. J. Dines, M. J. Snowden, R. Withnall, and B. Z. Chowdhry, *ibid.* **40**, 1508 (2009); A. P. Mendham, T. J. Dines, R. Withnall, J. C. Mitchell, and B. Z. Chowdhry, *ibid.* **40**, 1498 (2009); A. P. Mendham, B. S. Potter, R. A. Palmer, T. J. Dines, J. C. Mitchell, R. Withnall, and B. Z. Chowdhry, *ibid.* **41**, 148 (2010).
[13] M. Cotrait, M. Ptak, B. Busetta, and A. Heitz, J. Am. Chem. Soc. **98**, 1073 (1976).
[14] M. North, Tetrahedron **48**, 5509 (1992).
[15] R. Degeilh and R. E. Marsh, Acta Crystallogr. **12**, 1007 (1959).
[16] W. H. Zhang, V. Carravetta, O. Plekan, V. Feyer, R. Richter, M. Coreno, and K. C. Prince, J. Chem. Phys. **131**, 035103 (2009).
[17] V. Feyer, O. Plekan, R. Richter, M. Coreno, K. C. Prince, and V. Carravetta, J. Phys. Chem. A **113**, 10726 (2009).
[18] W. D. Allen, E. Czinki, and A. G. Csaszar, Chem.-Eur. J. **10**, 4512 (2004).
[19] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L.

Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).
[20] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, Gaussian, Inc., Wallingford, CT, 2004.
[21] L. Selvam, V. Vasilyev, and F. Wang, J. Phys. Chem. B **113**, 11496 (2009).
[22] D. P. Chong, E. van Lenthe, S. van Gisbergen, and E. J. Baerends, J. Comput. Chem. **25**, 1030 (2004); R. van Leeuwen and E. J. Baerends, Phys. Rev. A **49**, 2421 (1994).
[23] P. R. T. Schipper, O. V. Gritsenko, S. J. A. van Gisbergen, and E. J. Baerends, J. Chem. Phys. **112**, 1344 (2000).
[24] E. J. Baerends, J. Autschbach, A. Bérces, C. Bo, P. M. Boerrigter, L. Cavallo, D. P. Chong, L. Deng, R. M. Dickson, D. E. Ellis, M. van Faassen, L. Fan, T. H. Fischer, C. F. Guerra, S. J. A. van Gisbergen, J. A. Groeneveld, O. V. Gritsenko, M. Grüning, F. E. Harris, P. van den Hoek, H. Jacobsen, L. Jensen, G. van Kessel, F. Kootstra, E. van Lenthe, D. McCormack, A. Michalak, V. P. Osinga, S. Patchkovskii, P. H. T. Philipsen, D. Post, C. C. Pye, W. Ravenek, P. Ros, P. R. T. Schipper, G. Schreckenbach, J. G. Snijders, M. Solà, M. Swart, D. Swerhone, G. te Velde, P. Vernooijs, L. Versluis, O. Visser, F. Wang, E. van Wezenbeek, G. Wiesenekker, S. K. Wolff, T. K. Woo, A. Yakovlev, and T. Ziegler, Theoretical Chemistry, ADF 2006.01, SCM; Vrije Universiteit: Amsterdam, The Netherlands (2006).
[25] O. V. Gritsenko, R. van Leeuwen, and E. J. Baerends, Phys. Rev. A **52**, 1870 (1995).
[26] L. S. Cederbaum J. Phys. B **8**, 290 (1975); L. S. Cederbaum and W. Domcke, Adv. Chem. Phys. **36**, 205 (1977); W. von Niessen, J. Schirmer, and L. S. Cederbaum, Comput. Phys. Rep. **1**, 57 (1984); V. G. Zakrzewski and W. von Niessen, J. Comput. Chem. **14**, 13 (1993); V. G. Zakrzewski and J. V. Ortiz, Int. J. Quantum Chem. **53**, 583 (1995).
[27] M. Stener, G. Fronzoni, and M. de Simone, Chem. Phys. Lett. **373**, 115 (2003).
[28] D. P. Chong, J. Chem. Phys. **128**, 084112 (2008).
[29] M. Segala and D. P. Chong, J. Electron Spectrosc. Relat. Phenom. **171**, 18 (2009).
[30] R. Ramani, V. Sasisekharan, and K. Venkatesan, Int. J. Pept. Protein Res. **9**, 277 (1977); R. Ramani, K. Venkatesan, R. E. Marsh, and W.-J. H. Kung, Acta Crystallogr. **32**, 1051 (1976); K. Suguna, S. Ramakumar, N. Shamala, B. V. V. Prasad, and P. Balaram, Biopolymers **21**, 1847 (1982).
[31] F. Wang, M. T. Downton, and N. Kidwani, J. Theor. Comput. Chem. **4**, 247 (2005).
[32] T. Ikeda, R. C. Lord, T. B. Malloy, and T. Ueda, J. Chem. Phys. **56**, 1434 (1972); R. Davidson and P. A. Warsop, J. Chem. Soc., Faraday Trans. 2 **68**, 1875 (1972).
[33] M. Tomimoto and N. Go, J. Phys. Chem. **99**, 563 (1995).
[34] G. Sun, J. H. Voigt, I. V. Filippov, V. E. Marquez, and M. C. Nicklaus, J. Chem. Inf. Comput. Sci. **44**, 1752 (2004).
[35] See supplementary material at http://dx.doi.org/10.1063/1.3499740 for a full list of the calculated orbital energies.
[36] A. Ganesan, F. Wang, and C. Falzon, "Intermolecular Interactions of L-Phenylalanine: Valence Ionization Spectra and Orbital Momentum Distributions of its Fragment Molecules," J. Comput. Chem. (to be published).
[37] D. E. Manolopoulos, J. C. May, and S. E. Down, Chem. Phys. Lett. **181**, 105 (1991).
[38] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, A. B. Trofimov, E. V. Gromov, I. L. Zaytseva, and J. Schirmer, Chem. Phys. **347**, 360 (2008).
[39] A. Ganesan and F. Wang, J. Chem. Phys. **131**, 044321 (2009).