Journal of
Synchrotron Radiation

ISSN 0909-0495

Received 21 June 2011
Accepted 21 July 2011

Effects of alkyl side chains on properties of aliphatic amino acids probed using quantum chemical calculations

Aravindhan Ganesan,[a] Feng Wang,[a]* Michael Brunger[b] and Kevin Prince[c]

[a]Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Victoria 3122, Australia, [b]School of Chemical and Physical Sciences, Flinders University, Adelaide, South Australia 5001, Australia, and [c]Sincrotrone Trieste, Area Science Park, I-34012 Basovizza, Trieste, Italy. E-mail: fwang@swin.edu.au

Effects of alkyl side chains (*R*-) on the electronic structural properties of aliphatic amino acids are investigated using quantum mechanical approaches. The carbon (C 1*s*) binding energy spectra of the aliphatic amino acids are derived from the C 1*s* spectrum of glycine (the parent spectrum) by the addition of spectral peaks, depending on the alkyl side chains, appearing in the lower energy region IP < 290 eV (where IP is the ionization potential). The two glycyl parent spectral peaks of the amide 291.0 eV [$C_{(2)}$] and carboxylic 293.5 eV [$C_{(1)}$] C atoms are shifted in the aliphatic amino acids owing to perturbations depending on the size and structure of the alkyl chains. The pattern of the N 1*s* and O 1*s* spectra in glycine is retained in the spectra of the other amino acids with small shifts to lower energy, again depending on the alkyl side chain. The Hirshfeld charge analyses confirm the observations. The alkyl effects on the valence binding energy spectra of the amino acids are concentrated in the middle valence energy region of 12–16 eV, and hence this energy region of 12–16 eV is considered as the 'fingerprint' of the alkyl side chains. Selected valence orbitals, either inside or outside of the alkyl fingerprint region, are presented using both density distributions and orbital momentum distributions, in order to understand the chemical bonding of the amino acids. It is also observed that the HOMO–LUMO energy gaps of the aliphatic amino acids are reduced with the growth of the alkyl side chain.

Keywords: aliphatic amino acids; alkyl side chain effects; ionization energy spectra; dual space analysis; orbital momentum distributions.

1. Introduction

The attraction of gas phase chemistry has grown over the last decade, since important intramolecular features specific to individual molecules are best observed in the gas phase rather than in the liquid or solid environment (Lesarri *et al.*, 2004). Further, such gas phase studies also help in the recognition of the intrinsic geometries and electronic structures of biomolecules (Feyer *et al.*, 2008). Amino acids are biomolecules that are of much interest in various fields including chemistry, protein biochemistry, medicinal chemistry and drug research, as they remain the fundamental units in describing protein structures and functions. These amino acids may generally exist as ionized, charged zwitterionic [NH_3^+–CH(*R*)–COO^-] or neutral species in crystals or solution, depending on the pH of the environment. Nonetheless, studying amino acids in the gas phase with their neutral structures, NH_2–CH(*R*)–COOH, can provide better insights into their intrinsic physical and chemical properties, thus unfolding some of the chemistry behind protein dynamics.

Aliphatic α amino acids are the class of amino acids whose members include glycine (*R* = H), alanine (*R* = CH_3), valine [Val, *R* = $CH(CH_3)_2$], leucine [Leu, *R* = $CH_2CH(CH_3)_2$] and isoleucine [Isoleu, *R* = $CH(CH_3)(CH_2CH_3)$], of which glycine and alanine are naturally produced inside the human body and hence are 'non-essential' amino acids, whereas the other three branched-chain amino acids, *i.e.* Val, Leu and Isoleu, are all considered to be essential. In the gas phase the amino acids exist in a number of low-energy configurations or conformations that are mainly dependent on the type of intramolecular hydrogen-bonding interactions. Numerous experimental and theoretical studies and conformational analyses of these aliphatic amino acids have been reported (Lesarri *et al.*, 2004; Falzon & Wang, 2005; Jensen & Gordon, 1991; Hu *et al.*, 1993;

Herrera *et al.*, 2004; Császár, 1995; Shirazian & Gronert, 1997; Cocinero *et al.*, 2007; Lesarri *et al.*, 2005; Stepanian *et al.*, 1998; Stepanian, Reva, Radchenko, Rosado *et al.*, 1998; Stepanian *et al.*, 1999). The lowest-energy conformations and ground-state electronic structures are used in the present work. The lowest-energy conformations are generally stabilized by bifurcated hydrogen bonds from the amide to carbonyl groups (N—H···O═C), and *cis*-carboxylic functional group interactions (Lesarri *et al.*, 2004, 2005; Cocinero *et al.*, 2007; Stepanian *et al.*, 1998; Stepanian, Reva, Radchenko, Rosado *et al.*, 1998; Stepanian *et al.*, 1999; Falzon *et al.*, 2006).

The non-ionized aliphatic amino acids have been the subject of a variety of studies (Ji *et al.*, 2010; Danecek *et al.*, 2007; Cocinero *et al.*, 2007; Császár, 1995; Falzon & Wang, 2005; Falzon *et al.*, 2006; Feyer *et al.*, 2008; Herrera *et al.*, 2004; Hu *et al.*, 1993; Jensen & Gordon, 1991; Lesarri *et al.*, 2004, 2005; Shirazian & Gronert, 1997; Stepanian *et al.*, 1998; Stepanian *et al.*, 1999; Stepanian, Reva, Radchenko, Rosado *et al.*, 1998; Derbel *et al.*, 2007; Hu & Bernstein, 2008; Gontrani *et al.*, 2000; Tulip & Clark, 2004; Linder *et al.*, 2005, 2008; Kumar *et al.*, 2005, 2006; Sabino *et al.*, 2009), reporting theoretical computations and experimental measurements that include, but are not limited to, rotational spectroscopy, vibrational spectroscopy, optical activity and photoemission spectroscopy. Comparative studies of related molecules have improved the understanding of the structures of a number of biomolecules, although some results remain contradictory (Falzon *et al.*, 2006; Ganesan & Wang, 2009; Ganesan *et al.*, 2010, 2011; Chen & Wang, 2009; Selvam *et al.*, 2009). For example, Klasinc (1976) measured and assigned the valence spectra of some amino acids and suggested that the four outermost valence orbitals of an amino acid have dominant molecular orbital character of nitrogen lone pair, oxygen lone pair (*s*-like), oxygen lone pair (*p*-like) and π_{CO}. In contrast to the results of Klasinc, in a more recent study of the valence ionization spectrum of L-alanine, Powis *et al.* (2003) suggested that the third highest molecular orbital of Ala was assigned to π_{CO}, which was in agreement with an earlier measurement of Cannington & Ham (1983). Orbital 19a at about 14 eV was singled out as having particularly clear methyl character. Dehareng & Dive (2004) calculated the ionization energies and orbital character of the outer orbitals of the compounds, and found that their orbital characters were dominated by contributions from the amino-carboxylic acid moiety, in agreement with Powis *et al.* (2003). Some orbitals reveal certain contributions from the alkyl atoms, but not the dominant contributions.

In our previous work we have studied the effects of substitution of hydrogen in glycine by the methyl group in L-alanine (Falzon *et al.*, 2006) and by the 'fragments-in-molecules' model for L-phenylalanine (Ganesan *et al.*, 2011). In the present work, comparative quantum mechanical analyses of the aliphatic amino acids are carried out in order to observe the effects of the alkyl side chains on the amino acids and their trends of ionization energies in the gas phase. Further, information from the position and momentum spaces is combined, using dual space analysis (Wang, 2003), to explore the influence of the different species of the alkyl side chains of the aliphatic amino acids on their ionization and chemical bonding characteristics.

2. Methods and computational details

Geometry optimization of the most stable conformers of the aliphatic amino acids was performed using the density functional theory (DFT) based B3LYP/TZVP model, where TZVP stands for the triple zeta valence polarized basis set of Godbout *et al.* (1992). This B3LYP/TZVP model has been recognized to provide highly accurate geometries that are close to corresponding values from experiments (Falzon & Wang, 2005; Falzon *et al.*, 2006; Ganesan & Wang, 2009).

Molecular wavefunctions obtained from position or coordinate space (*r*-space), using the B3LYP/TZVP model, were directly transformed into momentum space (*k*-space) as orbital momentum distributions. Several approximations such as the plane-wave impulse approximation (McCarthy & Weigold, 1991), Born–Oppenheimer approximation and independent particle approximation were applied when the position space information was Fourier-transformed into *k*-space as momentum distributions (σ),

$$\sigma \propto \int \mathrm{d}\Omega \left|\psi_j(p)\right|^2. \quad (1)$$

Here, $\psi_j(p)$ represents the transformed Kohn–Sham orbitals in momentum space and p is the momentum of the target electron at the instant of ionization. The theoretical momentum distributions are generated using the *NEMS* code (Ning *et al.*, 2008). Note that theoretical momentum distributions for benzene (Ganesan *et al.*, 2011) and glycine (Falzon & Wang, 2005) in previous studies indicated that the B3LYP/TZVP model is an appropriate model for molecules of this size and reliable for calculating momentum distributions of biomolecules.

Other quantum mechanical models such as LB94/et-pVQZ (van Leeuwen & Baerends, 1994) for the core-shell calculations, OVGF/TZVP, *i.e.* outer valence Green's function (Cederbaum & Domcke, 2007) coupled with the TZVP basis set, and SAOP/et-pVQZ (Schipper *et al.*, 2000) for valence calculations were employed. Here, et-pVQZ is an even-tempered polarized valence quadruple-zeta Slater-type basis set (Chong *et al.*, 2004). Calculations based on the B3LYP/TZVP and OVGF/TZVP models were carried out using the *Gaussian03* computational chemistry program (Frisch *et al.*, 2004), whereas the LB94/et-pVQZ and SAOP/et-pVQZ models are from the Amsterdam density functional package (Velde *et al.*, 2001). *Molden* (Schaftenaar & Noordik, 2000) and *Jamberoo* (http://sf.anu.edu.au/~vvv900/cct/appl/jmol editor/) were used for molecular visualization.

Except for glycine, the aliphatic amino acids are chiral, and the naturally occurring L-enantiomers of alanine, valine, leucine and isoleucine were calculated. This does not influence the energies of the valence orbitals and core levels, as only the intensity is affected in chiral experiments (such as photo-

Table 1
Selected geometric parameters of the aliphatic amino acids with relevant experimental data and results from other studies.

	Glycine			Alanine			Valine			Leucine		Isoleucine	
	This work†	Exp‡	Other work§	This work†	Exp¶	Other work§	This work†	Exp††	Other work‡‡	This work†	Exp††	This work†	Exp††
O—H (Å)	0.968			0.970			0.970			0.970		0.970	
$O_{(1)}—C_{(1)}$ (Å)	1.357	1.354	1.356	1.357	1.347	1.356	1.357		1.357	1.358		1.358	
$O_{(2)}—C_{(1)}$ (Å)	1.209	1.204	1.209	1.211	1.192	1.211	1.212		1.214	1.212		1.212	
$C_{(1)}—C_{(2)}$ (Å)	1.520	1.529	1.519	1.523	1.507	1.521	1.522	1.524	1.531	1.521	1.522	1.523	1.525
$C_{(2)}—N$ (Å)	1.449	1.466	1.447	1.454	1.471	1.452	1.456	1.469	1.457	1.454	1.480	1.456	1.484
$C_{(2)}—C_{(3)}$ (Å)	1.095§§			1.532	1.544	1.530	1.547	1.537	1.561	1.538	1.526	1.547	1.547
$O_{(1)}—C_{(1)}—O_{(2)}$ (°)	123.4			123.1		122.5¶¶	123.0			123.1		122.9	
H—N—H (°)	106.1			106.0		109.1¶¶	105.7			105.7		105.6	
$C_{(1)}—C_{(2)}—N$ (°)	115.6	113.0	115.6	113.7	110.1	113.7	113.0	108.4	112.8	113.6	110.1	113.1	108.3
$O_{(1)}—C_{(1)}—C_{(2)}$ (°)	110.9	111.5	110.9	111.4	110.3	111.4	111.6		112.3	111.5		111.5	
$O_{(2)}—C_{(1)}—C_{(2)}$ (°)	125.7	125.0	125.7	125.4	125.6	125.4	125.4		125.5	125.3		125.5	
$N—C_{(2)}—C_{(3)}$ (°)	109.9§§			109.6		110.7¶¶	111.1	111.3		110.2	109.5	111.2	111.9
$C_{(1)}—C_{(2)}—C_{(3)}$ (°)	107.5§§			108.4			109.2			107.6		109.6	
$O_{(1)}—C_{(1)}—O_{(2)}—C_{(2)}$ (°)	180.0			178.1			178.2			178.1		178.7	
$O_{(1)}—C_{(1)}—C_{(2)}—N$ (°)	180.0	180.0	180.0	161.6		161.3	165.1			161.3		168.0	
$O_{(2)}—C_{(1)}—C_{(2)}—N$ (°)	0.0	0.0	0.0	−20.0			−16.5			−20.4		−13.2	
$O_{(1)}—C_{(1)}—C_{(2)}—C_{(3)}$ (°)	−56.9§§			−76.2			−70.7			−76.4		−67.3	
$O_{(2)}—C_{(1)}—C_{(2)}—C_{(3)}$ (°)	123.1§§			102.1			107.7			101.8		111.5	
$H—O_{(1)}—C_{(1)}—C_{(2)}$ (°)	180.0			176.9			177.8			177.0		178.6	
$H—O_{(1)}—C_{(1)}—O_{(2)}$ (°)	0.0			−1.4			−0.7			−1.3		−0.2	

† B3LYP/TZVP. ‡ Iijima *et al.* (1991). § MP2 calculations (Császár, 1995). ¶ Iijima & Beagley (1991). †† Gould *et al.* (1985). ‡‡ B3LYP/6-31++G** (Stepanian *et al.*, 1999). §§ $C_{(3)}$ is replaced by H in glycine. ¶¶ B3LYP/6-31++G** (Kumar *et al.*, 2006).

ionization with circularly polarized light), which is out of scope of the present study.

3. Results and discussion

3.1. Optimized geometries

The lowest energy structures of the aliphatic amino acids are those which have been found to be the most stable conformers in previous experiments and theoretical calculations (Cocinero *et al.*, 2007; Császár, 1995; Falzon & Wang, 2005; Herrera *et al.*, 2004; Hu *et al.*, 1993; Jensen & Gordon, 1991; Lesarri *et al.*, 2004, 2005; Shirazian & Gronert, 1997; Stepanian, Reva, Radchenko, Rosado & Adamowicz, 1998, 1999; Stepanian, Reva, Radchenko, Rosado *et al.*, 1998). The conformer of glycine in this work has a C_s symmetry. However, when a H atom on $C_{(2)}$ in glycine is substituted by the alkyl groups to form the other amino acids, the point group symmetry is reduced to the lower C_1 point group. Table 1 compares geometrical parameters of the aliphatic amino acids with available experiment and other calculations.

As can be observed from Table 1, the calculated geometries agree well with those of the available experiments (Iijima *et al.*, 1991; Iijima & Beagley, 1991; Gould *et al.*, 1985) and other theoretical calculations (Császár, 1995; Kumar *et al.*, 2006; Stepanian *et al.*, 1999). In particular, all the bond lengths are in excellent agreement, with the maximum discrepancy being approximately 0.03 Å, whereas the bond angles display slightly larger differences compared with the measured angles, with the largest deviations being observed for the bond angle of $C_{(1)}—C_{(2)}—N$ in all amino acids with a maximum discrepancy of 3.7°. Nevertheless, bond angles are known to be less sensitive to energy variations and hence are much more difficult to predict in quantum mechanics (Falzon *et al.*, 2006).

On comparing the geometric parameters of the respective amino acids, the bond lengths are found to be hardly affected by the changes in the alkyl side chains. However, the bond angles and torsional angles display changes in response to the alkyl chain modifications; in particular, the $C_{(1)}—C_{(2)}—N$ bond angle and the $O—C_{(1)}—C_{(2)}—N$ dihedral angle are significantly affected. Although these latter angles show differences in all the species, the largest disparities are observed between glycine and alanine, where the $C_{(1)}—C_{(2)}—N$ bond angle difference is almost 2° and the $O—C_{(1)}—C_{(2)}—N$ dihedral angle difference is as large as 20°. Similar geometric trends were reported in our previous work (Falzon *et al.*, 2006) and by Csaszar (1995). Such large angular differences originate from the breaking of the symmetry plane of the most stable conformer of glycine, as one of the H atoms on the $C_{(2)}$ site is substituted by a methyl group in alanine.

Another interesting observation regarding geometry concerns the changes in the $C_{(2)}—C_{(3)}$ bond length and the $N—C_{(2)}—C_{(3)}$ bond angle, in the amino acids other than glycine (in glycine these bonds do not exist). Both these parameters in leucine (1.538 Å, 110.2°) are smaller than those of valine (1.547 Å, 111.1°) and isoleucine (1.547 Å, 111.2°). These geometric trends are due to the effect of intramolecular steric repulsion from the methyl group attached to $C_{(3)}$ of valine and isoleucine (Gould *et al.*, 1985), whereas in leucine this methyl group is attached to $C_{(4)}$ [refer to Fig. 1 for the structures and Fig. 2(*a*) for their numbering]. Thus, overall the modification of the alkyl side chains in the aliphatic amino acids exhibit very small effects on the geometries of these species.

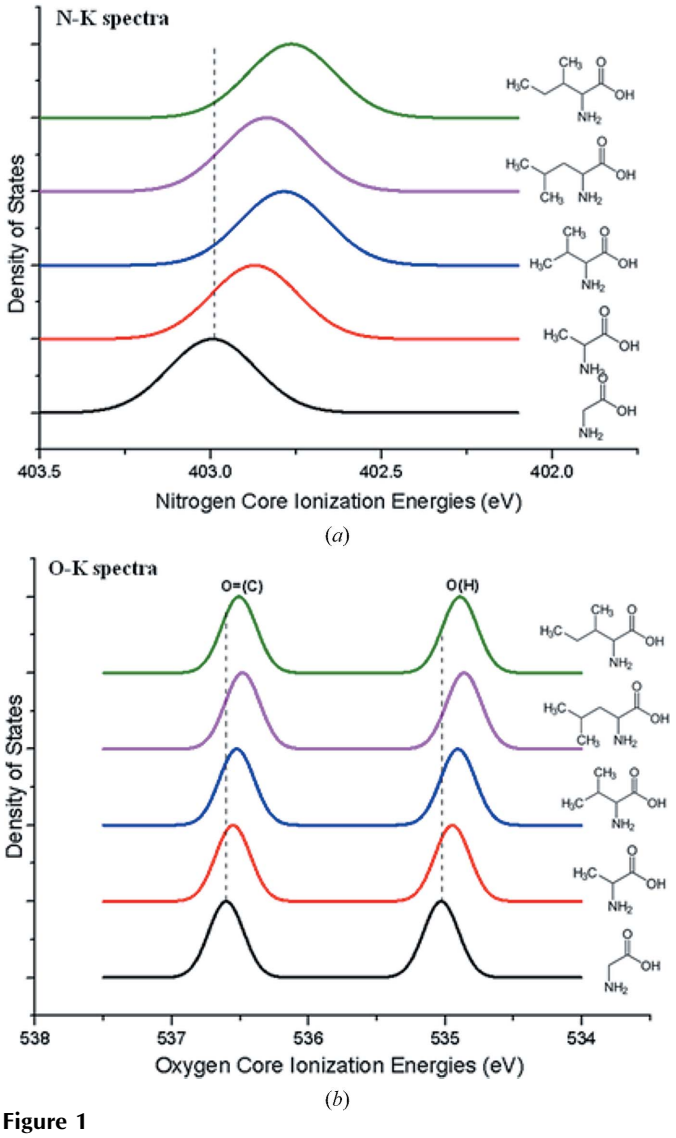

Figure 1
Comparison of (*a*) the N 1*s* spectra and (*b*) the O 1*s* spectra of the aliphatic amino acids simulated using the LB94/et-pVQZ model with a FWHM of 0.57 eV.

3.2. Hirshfeld charge distributions

Hirshfeld charges (Hirshfeld, 1991) are important anisotropic properties, useful for understanding atomic behaviour within molecules. Hirshfeld charges (Q^{H}) of the aliphatic amino acids, calculated based on the LB94/et-pVQZ wavefunction, are presented in Table 2. From the table it is noted that all the nitrogen and oxygen sites in the amino acids possess negative charges, whereas the carbon sites in the aliphatic amino acids display charge re-distribution related to their respective side chains. The small negative charge on $C_{(2)}$, −0.014 a.u., becomes positive in the other amino acids, while all the other carbon atoms within the alkyl side chains are negative.

Although all the carbon sites in the alkyl side chains of the amino acids possess negative charges, they still show some subtle differences depending upon their local interactions. For example, despite the $C_{(3)}$ sites being negative in all the molecules, the charges in valine and isoleucine are very small, with values of −0.014 and −0.016 atomic units, respectively. This is because the weakly electron-donating methyl groups attached at this carbon site [$C_{(3)}$] in valine and isoleucine lead to a charge redistribution, thereby making $C_{(3)}$ almost neutral. This observation is very similar to that of 1-(β-D-ribofuranosyl)-5-methyl-2-pyrimidinone (d5) which showed a reduced negative charge at the carbon site attached to a methyl group (Selvam *et al.*, 2009).

Table 2
Hirshfeld charges of the C, N and O sites of the aliphatic amino acids (in a.u.).

Values in bold face indicate negative charge accumulation.

	Glycine	Alanine	Valine	Leucine	Isoleucine
$C_{(1)}$	0.216	0.215	0.215	0.214	0.215
$C_{(2)}$	**−0.014**	0.025	0.023	0.021	0.023
$C_{(3)}$		**−0.106**	**−0.014**	**−0.061**	**−0.016**
$C_{(4)}$			**−0.116**	**−0.117**	**−0.066**
$C_{(5)}$			**−0.116**	**−0.115**	**−0.113**
$C_{(6)}$				**−0.115**	**−0.116**
$O_{(H)}$	**−0.201**	**−0.198**	**−0.196**	**−0.198**	**−0.196**
O(=C)	**−0.295**	**−0.296**	**−0.292**	**−0.299**	**−0.294**
N	**−0.241**	**−0.24**	**−0.231**	**−0.23**	**−0.231**

3.3. Vertical core ionization energies and spectra

The present core-shell calculations were carried out using the LB94/et-pVQZ model and based on the 'meta-Koopman' theorem (Gritsenko *et al.*, 2003). Fig. 1 presents the simulated N 1*s* (*a*) and O 1*s* (*b*) ionization binding energy spectra of the amino acids. The O sites and the N sites are not affected much by the alkyl side chain modification, and hence their ionization potentials (IPs) exhibit small shifts to lower energy in both N-*K* and O-*K* spectra. Fig. 2(*a*) compares the C 1*s* spectra of the aliphatic amino acids simulated using a full width at half-maximum (FWHM) of 0.57 eV, reproducing the experimental conditions of the measurement of glycine (Plekan *et al.*, 2007). The C 1*s* spectra are also marked using individual ionization energies of the carbon sites to indicate the spectral line positions. As seen in this figure, the energies of the peaks fall into three classes: $C_{(1)}$═O, $C_{(2)}$−N and the alkyl C atoms. Basically, the $C_{(1)}$ 1*s* and $C_{(2)}$ 1*s* spectra of all the amino acids are similar to the spectra of glycine, with small energy shifts to lower energy, except for $C_{(2)}$ of alanine, which exhibits a small energy shift to higher energy. The alkyl C atoms form the structure-dependent C 1*s* spectra in the lower energy side of C 1*s* with IP < 290 eV. The core binding energies of the C, N and O 1*s* sites of the amino acids together with the available experimental ionization energies for glycine (Plekan *et al.*, 2007) and alanine (Feyer *et al.*, 2008) are given in the supplementary Table S1.[1]

Fig. 2(*b*) gives the carbon core energy diagram of the aliphatic amino acids. The carboxyl carbon [$C_{(1)}$] and the α carbon [$C_{(2)}$] of the molecules remain in a similar chemical

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: VV5018). Services for accessing these data are described at the back of the journal.

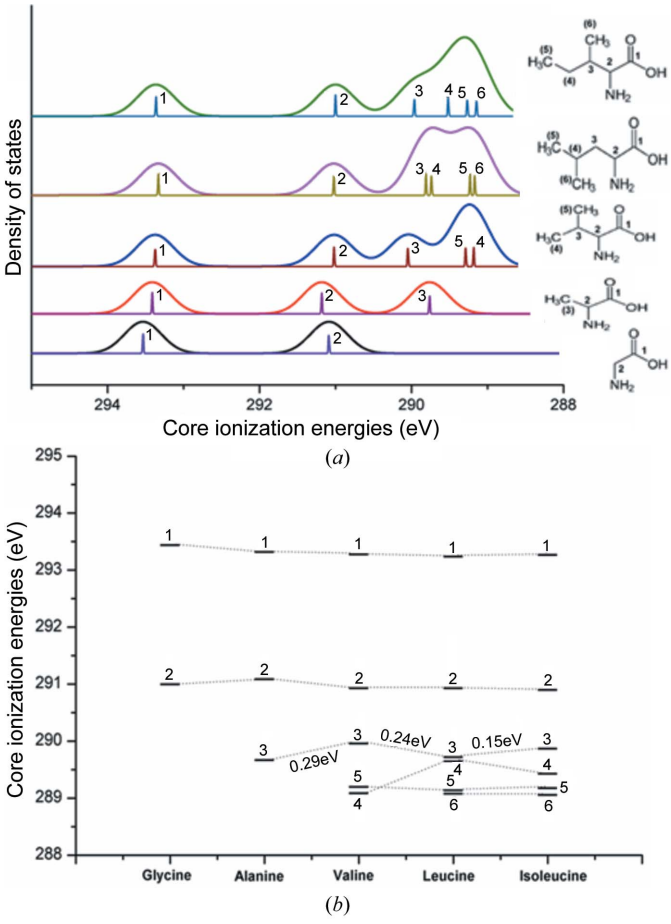

Figure 2
(*a*) Comparison of the C 1*s* spectra of the aliphatic amino acids simulated using the LB94/et-pVQZ model with a FWHM of 0.57 eV. (*b*) C 1*s* level energy diagram of the aliphatic amino acids.

environment and the alkyl side chains have minimal impact on their ionization potentials. For the alkyl side chains in valine, leucine and isoleucine, the chemical environment of the terminal methyl groups are very similar, and therefore their related ionization potentials remain very similar too. The $C_{(3)}$ atoms in valine and isoleucine are both associated with a 1-methylpropyl moiety which is chemically different from the butyl moiety in leucine. As a result, the IPs of $C_{(3)}$ in valine and isoleucine are close in value, but quite different from the IP of $C_{(3)}$ in leucine, as shown in Fig. 2(*b*). The IP change, ΔIP, of the $C_{(4)}$ atom between valine and leucine is almost double (+0.56 eV) the ΔIP between leucine and isoleucine (−0.22 eV). The $C_{(4)}$ site in leucine is bonded to three carbon atoms, *i.e.* $C_{(3)}$, $C_{(5)}$ and $C_{(6)}$, but $C_{(4)}$ in isoleucine is bonded to only two carbon atoms, *i.e.* $C_{(3)}$ and $C_{(5)}$. The IP changes in the opposite directions on the $C_{(4)}$ site with respect to the $C_{(3)}$ site in this figure.

3.4. Valence ionization and spectral properties

Valence shell ionization potentials directly link the valence electronic structural information with molecular orbital theory, and provide details of intramolecular interactions within the molecules. Table 3 compares vertical valence IPs of the aliphatic amino acids, obtained using the DFT-based SAOP/et-pVQZ and OVGF/TZVP models, along with the available experimental data of glycine (Plekan *et al.*, 2007) and results from other theoretical calculations based on the ROVGF/cc-pVDZ (for alanine) (Powis *et al.*, 2003) and OVGF (for valine, leucine and isoleucine) (Dehareng & Dive, 2004) models. A number of previous studies have recognized the SAOP/et-pVQZ model as an accurate model for valence space calculations (Ganesan *et al.*, 2011; Schipper *et al.*, 2000; Slavícek *et al.*, 2009; Wang, 2007; Wang & Pang, 2007), and the present IPs for the amino acids are also in good agreement with available results. However, those earlier studies also contended that the SAOP model is not more accurate than the OVGF model for frontier orbitals including the highest occupied molecular orbital (HOMO) (Falzon & Wang, 2005; Ganesan *et al.*, 2011; Saha *et al.*, 2005; Wang, 2007). This is similarly true in the case of aliphatic amino acids.

As seen in Table 3, the SAOP/et-pVQZ model slightly overestimates the ionization energies of the HOMOs. For example, the HOMO in glycine (16a′) has a calculated energy of 10.53 eV and 9.89 eV using the SAOP and OVGF models, respectively, compared with the measured value of 10.0 eV (Plekan *et al.*, 2007). However, the OVGF model only applies to the outer valence space of a molecule, but is unable to predict the valence IPs at energies larger than 20 eV where the SAOP model becomes attractive as it is able to calculate the IPs in the complete valence space of a molecule and it agrees more accurately with experiment. Thus the combination of the two models, SOAP/et-pVQZ (for inner valence regions) and OVGF (for outer valence regions), is able to provide a more extensive valence space prediction.

Fig. 3 compares the simulated valence ionization spectra of glycine, calculated using both the SAOP and OVGF models, with a recent high-resolution gas-phase synchrotron-radiation-based photoelectron spectroscopy (PES) measurement of Plekan *et al.* (2007). The simulated valence binding energy spectra of glycine have been globally shifted in order to align

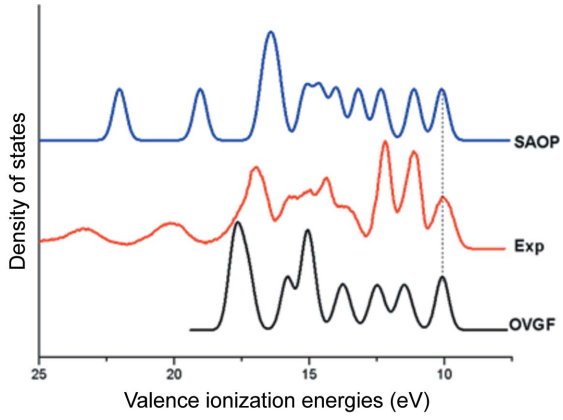

Figure 3
Comparison of the experimental photoelectron spectra of glycine in the outer valence space with the simulated spectra using the OVGF/TZVP model (bottom, $\Delta\varepsilon = 0.20$ eV) and the SAOP/et-pVQZ model (top, $\Delta\varepsilon = -0.42$ eV). The simulated spectra are convoluted with a FWHM of 0.70 eV.

Table 3
Comparison of the valence orbital ionization energies (eV) of the aliphatic amino acids with other theoretical and experimental energies.

Glycine			Alanine			Valine			Leucine			Isoleucine	
Orbital	SAOP (OVGF)†	Exp‡	Orbital	SAOP (OVGF)†	ROVGF§	Orbital	SAOP (OVGF)†	OVGF¶	Orbital	SAOP/ OVGF†	OVGF¶	SAOP (OVGF)†	OVGF¶
HOMO			HOMO			HOMO			HOMO				
(16a′)	10.53 (9.89)	10.0	(24)	10.49 (9.75)	9.58	(32)	10.43 (9.59)	9.50	(36)	10.45 (9.61)	9.51	10.41 (9.56)	9.45
15a′	11.56 (11.29)	11.2	23	11.43 (11.08)	10.77	31	11.37 (10.93)	10.80	35	11.36 (10.92)	10.77	11.34 (10.96)	10.79
4a″	12.78 (12.29)	12.2	22	12.71 (12.14)	11.88	30	12.23 (11.63)	11.39	34	12.08 (11.40)	11.27	11.99 (11.20)	11.10
3a″	13.61 (13.57)	13.7	21	13.15 (12.81)	12.58	29	12.36 (11.80)	11.67	33	12.11 (11.61)	11.44	12.14 (11.60)	11.39
14a′	14.42 (14.73)	14.4	20	13.60 (13.43)	13.22	28	12.63 (12.18)	12.03	32	12.27 (11.63)	11.58	12.26 (11.76)	11.63
13a′	15.04 (14.95)	15.0	19	13.74 (13.74)	13.59	27	12.79 (12.23)	12.01	31	12.68 (12.13)	11.90	12.74 (12.16)	11.90
2a″	15.57 (15.61)	15.8	18	14.68 (15.09)	14.82	26	13.43 (13.47)		30	13.00 (12.74)		13.01 (12.74)	
12a′	16.60 (16.61)	16.6	17	15.09 (15.00)	14.84	25	13.55 (13.50)		29	13.38 (13.12)		13.39 (13.41)	
11a′	16.85 (17.05)	16.9	16	15.66 (15.66)	15.37	24	14.35 (14.34)		28	13.58 (13.59)		13.52 (13.23)	
1a″	17.11 (17.42)	17.6	15	16.49 (16.84)	16.60	23	14.54 (14.61)		27	13.89 (13.79)		13.92 (13.90)	
10a′	19.46	20.2	14	16.66 (17.17)	17.00	22	14.78 (14.86)		26	14.43 (14.60)		14.36 (14.31)	
9a′	22.45	23.3	13	17.20 (17.57)	17.30	21	15.16 (15.07)		25	14.85 (14.93)		14.68 (14.97)	
8a′	26.80	28.3	12	18.97	19.64	20	15.70 (15.66)		24	15.02 (14.93)		15.03 (14.96)	
7a′	30.33	32.3	11	21.03	22.03	19	16.54 (16.89)		23	15.51 (15.64)		15.30 (15.24)	
6a′	32.57	34.3	10	23.22	24.56	18	16.72 (17.15)		22	15.58 (15.48)		15.83 (15.86)	
			9	26.88		17	17.16 (17.46)		21	16.52 (16.84)		16.53 (16.82)	
			8	30.28		16	18.18 (18.92)		20	16.61 (17.01)		16.68 (17.13)	
			7	32.53		15	19.54		19	17.15 (17.48)		17.05 (17.30)	
						14	21.24		18	18.09 (18.83)		18.04 (18.75)	
						13	22.07		17	19.03		19.29	
						12	24.14		16	20.85		20.13	
						11	26.86		15	21.18		21.81	
						10	30.26		14	22.77		22.60	
						9	32.52		13	24.432		24.34	
									12	26.88		26.84	
									11	30.22		30.24	
									10	32.47		32.50	

† SAOP/et-pVQZ model. Energies based on the OVGF/TZVP model are given in parentheses. ‡ Plekan *et al.* (2007). § ROVGF/cc-pVDZ (Powis *et al.*, 2003). ¶ OVGF model (Dehareng & Dive, 2004).

the first experimental ionization energy at 10.0 eV (Plekan *et al.*, 2007). The PES spectrum simulated using the OVGF model has been blue shifted by ΔIP = 0.20 eV, whereas the energy shift is ΔIP = −0.42 eV for the SAOP model. After taking account of these energy shifts, the OVGF and SAOP models agree well with the measurement in the valence space. Indeed the OVGF model reproduces the measured data in the outer valence space very well, whereas the SAOP model is able to continue providing a good estimation in the region beyond 20 eV. This is clearly seen in this figure. A combination of these two models thus provides a reliable tool for valence ionization energies of molecules.

Fig. 4 compares the outer valence vertical ionization spectra of the aliphatic amino acids, simulated now using a FWHM of 0.70 eV and the OVGF model. The valence spectra are more complicated than the core spectra, as there are many more valence electrons than core electrons, and, moreover, valence electrons exhibit a more delocalized nature than the core electrons. The alkyl side chain effects on the inner shell are different from the valence shell: the spectra in the core shell indeed exhibit a localized feature, whereas the spectra in the valence shell are largely delocalized. For example, all the alkyl side chain dominant spectral peaks in the C 1*s* spectra appear below 290.50 eV as shown in Fig. 2(*a*). However, alkyl effects

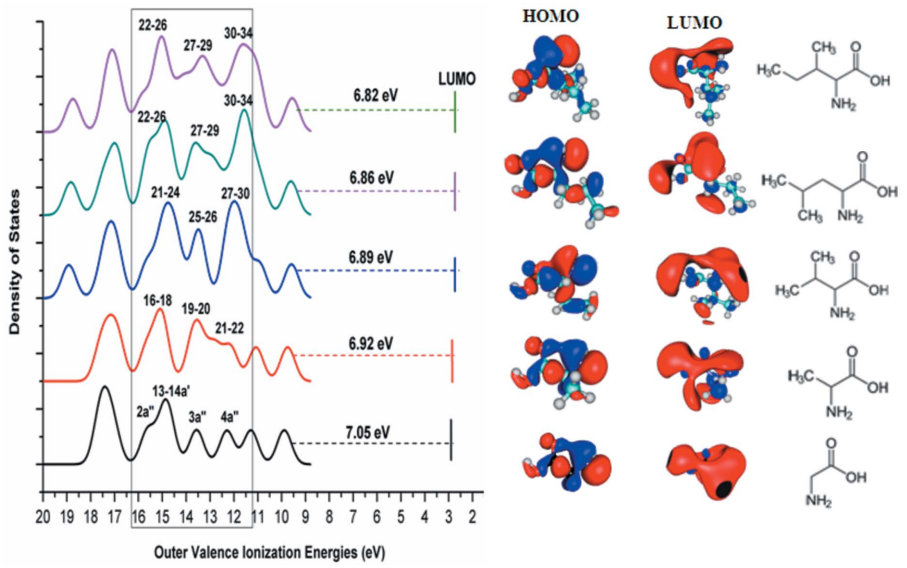

Figure 4
Comparison of the simulated outer valence vertical ionization spectra of the aliphatic amino acids simulated using the OVGF/TZVP model, convoluted with a FWHM of 0.70 eV.

on the valence spectra mix with the main groups in the energy region 12 eV < IP < 16 eV, as shown in Fig. 4. The inner valence spectral region IP > 16 eV is dominant by the main groups again. As a result the alkyl signatures of the aliphatic amino acids concentrate in the energy region 12 eV < IP < 16 eV, as well as the decrease of the HOMO–LUMO energy gaps, as shown in Fig. 4. Indeed, the spectra of the aliphatic amino acids seem similar in the frontier orbital region IP < 12 eV and in the higher-energy region IP > 16 eV.

The HOMO and the next HOMO (NHOMO), appearing within the valence energy range <2 eV, display similarities in their ionization energies, indicating that the spectral changes owing to the side chains of these amino acids are not in the frontier orbital region, in agreement with previous results (Falzon & Wang, 2005; Ganesan *et al.*, 2010). This is due to the fact that the HOMOs and NHOMOs [see Figs. 5(*a*) and 5(*b*)] of the amino acids are not localized on the alkyl moieties. In the higher-energy region IP > 16 eV, similarities appear again, but the small spectral peak at ∼19 eV, which is missing in glycine and alanine, is associated with the growth and branch of the aliphatic chain in valine, leucine and isoleucine. The energy gaps between the HOMOs and the lowest unoccupied molecular orbitals (LUMOs), or HOMO-LUMO gaps, become smaller with increasing size of the alkyl side chain of the molecules, that is: $\Delta\varepsilon_{\text{HOMO-LUMO}}$ (glycine) > $\Delta\varepsilon_{\text{HOMO-LUMO}}$ (alanine) > $\Delta\varepsilon_{\text{HOMO-LUMO}}$ (valine) > $\Delta\varepsilon_{\text{HOMO-LUMO}}$ (leucine) > $\Delta\varepsilon_{\text{HOMO-LUMO}}$ (isoleucine), as also shown in Fig. 4. Thus, the side chains contribute to the energy reduction of the HOMO–LUMO gaps, which is also found in 1-(β-D-ribofuranosyl)-5-methyl-2-pyrimidinone (d5) when compared with its parent molecule, 1-(β-D-ribofuranosyl)-2-pyrimidinone (zebularine) (Selvam *et al.*, 2009). In summary, the side chain effects on the aliphatic amino acids are largely revealed in the energy region 12 eV < IP < 16 eV of the binding energy spectra, indicating strong chemical bonds involving alkyl contributions in this region, whereas the alkyl does not contribute significantly in the outer and inner valence regions of the amino acids.

3.5. Orbitals and chemical bonding characteristics

Position space analyses provide orbital information in a qualitative manner, whereas momentum space studies are able to reveal subtle information about orbitals and, therefore, chemical bonding. The complementary nature of information in position and momentum space is revealed by dual space analysis (Wang, 2003). Figs. 5(*a*)–5(*d*) present selected valence orbitals including frontier orbitals with similar chemical bonding characters and their momentum distributions. The orbital electron density of the HOMOs (see Fig. 5*a*) is concentrated on the amine group (NH_2) in the HO—C(═O)C—N moiety, with only minor density appearing in the aliphatic extension of the amino acids, which is consistent with the fact that the alkyl chains contribute little to those orbitals. In their orbital momentum distributions it is clearly demonstrated that the HOMOs are dominated by *p*-like electrons from the –NH_2 moiety. In fact, the two hydrogen 1*s* electrons bonding with the nitrogen 2*p* electrons in –NH_2 enhances the *p*-like distribution. The residual density in the HO—C(═O)C—N moiety as well as the aliphatic chain contribute to the small increase in the lower momentum region of the *sp*-like orbital momentum distribution of the HOMOs. Although the NHOMOs in Fig. 5(*b*) also exhibit similar trends as the HOMOs among the amino acids, their orbital momentum distributions are not the same as the HOMOs. The orbital density distributions in this figure reveal that the HO—C(═O)C—N bond network dominates, but in the NHOMOs the electron lone pair on the O atom of the keto C═O moiety as well as the N atom of the amine NH_2 both also contribute to this orbital. As a result, the orbital momentum distribution of the NHOMOs exhibit strong *p*-like character with two *p*-like peaks.

The other important orbitals in the valence space, selected by their similarities, are the innermost and the third innermost valence orbitals of the amino acids, as shown in Figs. 5(*c*) and 5(*d*), respectively. The innermost valence orbitals are dominated by the 2*s* electrons from the two O atoms in the carboxylic acid group, HO—C(═O). The side chains do not affect this orbital significantly, so that the angularly averaged orbital momentum distributions of these innermost valence orbitals are almost identical (Fig. 5*c*). The orbital momentum distributions of the third innermost orbitals, assigned to N 2*s*, are again similar to each other. However, although dominated by the NH_2 moiety in this orbital, the alkyl side chains show some effect of mixing. Therefore, the orbital momentum

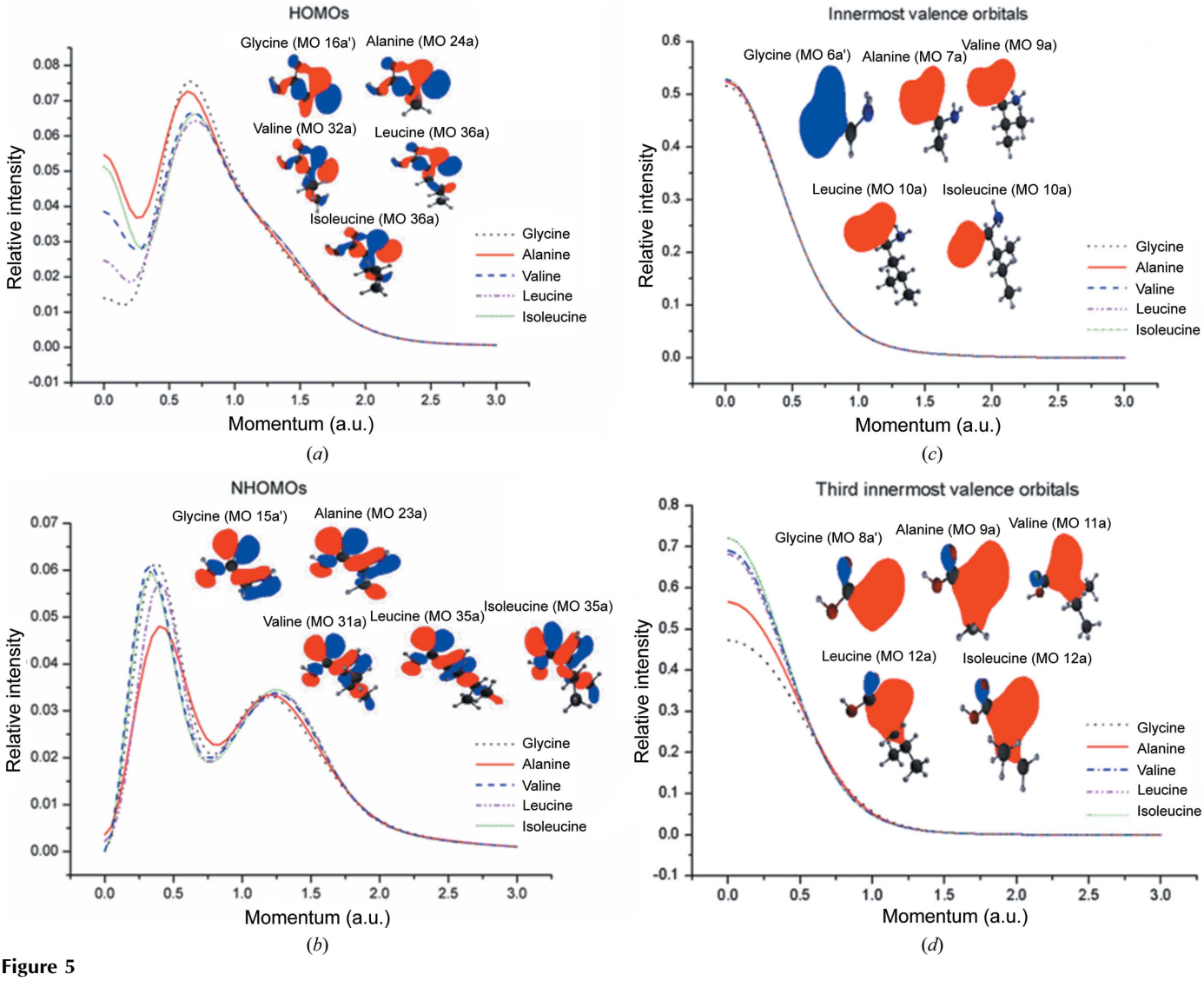

Figure 5
Selected valence orbital electron density and momentum distributions of the aliphatic amino acids: (*a*) the HOMOs, (*b*) the NHOMOs, (*c*) the innermost valence orbitals and (*d*) the third innermost valence orbitals.

distributions are *s*-like ('half-bell shaped') but differ in the low momentum region owing to the side chains as shown in Fig. 5(*d*).

Apart from the similarities of the few selected valence orbitals which are concentrated on the common HO—C(═O)C—N moiety, and the inner valence orbitals, the majority of the valence orbitals of the molecules have contributions from their aliphatic side chains and therefore their orbitals are very different, depending on their structures. Figs. 6(*a*)–6(*d*) display the orbital electron density distributions and momentum distributions of the molecules as the aliphatic side chain grows. It is obvious that these energy levels of the amino acids are correlated. Figs. 6(*a*) and 6(*b*) report the orbitals where there is a correlation in the glycine–alanine–valine series, whereas Figs. 6(*c*) and 6(*d*) display the valine–leucine–isoleucine series. The related orbital momentum distributions also show a certain association among the three orbitals, which exhibit similar orbital distributions in the larger momentum region but change significantly in the low momentum region.

4. Conclusions

The effects of the alkyl side chain of the aliphatic amino acids were reported using quantum mechanically simulated ionization spectra in both valence and core shells. The effects on the N 1*s* and O 1*s* core ionization spectra exhibit small perturbations of the spectra of glycine, whereas in the C 1*s* spectra there are additional peaks at lower energy, IP < ∼290 eV, owing to the additional C atoms in the alkyl chains. In the valence ionization region, spectral changes are significant in the middle energy region 12 eV < IP < 16 eV, while the outermost and innermost valence peaks are less affected with respect to the alkyl chains. This is because the electron density distributions of these orbitals are concentrated on the common HO—C(═O)C—N moiety of the amino acids and

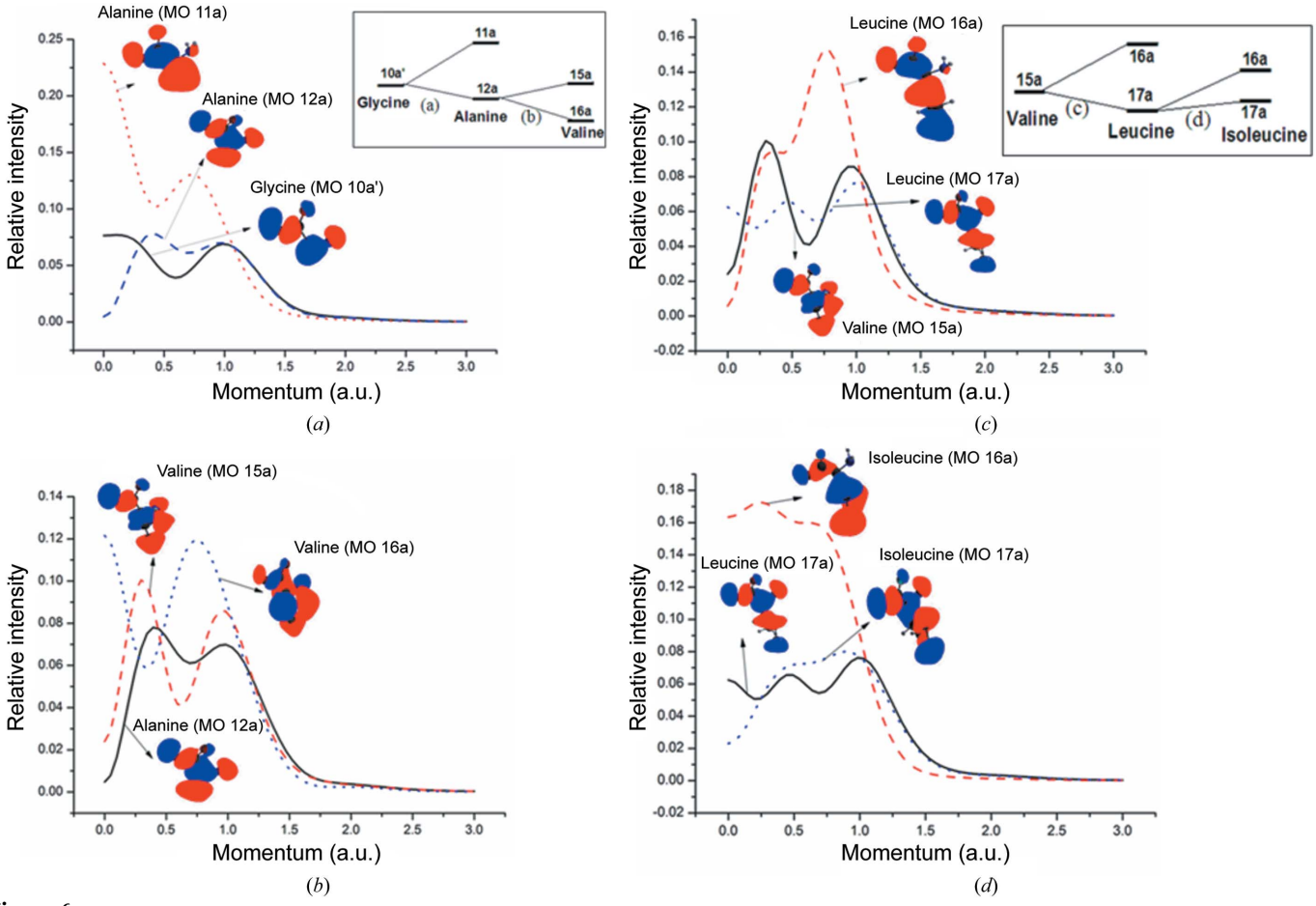

Figure 6
Orbital momentum densities and charge densities of (*a*) glycine (10a′ MO), alanine (11a and 12a MOs); (*b*) alanine 12a, valine (15a and 16a MOs), and (*c*) and (*d*) valine (15a MO), leucine (16a and 17a MOs) and isoleucine (16a and 17a MOs).

are largely 2*s* in character. As a result, the growth of the alkyl side chain does not significantly affect the frontier orbitals of the aliphatic amino acids. However, other properties, such as the HOMO–LUMO energy gap, displayed changes in relation to the side chains. The HOMO–LUMO energy decreases with increasing side chain length, and we found that the energy order was: glycine > alanine > valine > leucine > isoleucine. The valence energy region of approximately 12–16 eV remains the signature region for the aliphatic amino acids, as the orbitals of the side chains appeared in this region.

FW acknowledges her Vice-Chancellor's Research Award at Swinburne University of Technology, making the PhD scholarship of AG possible. AG thanks the ARC Centre of Excellence for Antimatter–Matter studies for additional financial support. The authors would like to thank Dr N. G. Ning of Tsinghua University for providing the *NEMS* code. Supercomputing facilities at NCI, VPAC, VLSCI and Swinburne University's Green Machine are also acknowledged.

References

Cannington, P. H. & Ham, N. S. (1983). *J. Electron Spectrosc. Relat. Phenom.* **32**, 139–151.
Cederbaum, L. S. & Domcke, W. (2007). *Adv. Chem. Phys.* **36**, 205–344.
Chen, F.-F. & Wang, F. (2009). *Molecules*, **14**, 2656–2668.
Chong, D. P., van Lenthe, E., Van Gisbergen, S. & Baerends, E. J. (2004). *J. Comput. Chem.* **25**, 1030–1036.
Cocinero, E. J., Lesarri, A., Grabow, J. U., López, J. C. & Alonso, J. L. (2007). *ChemPhysChem*, **8**, 599–604.
Császár, A. G. (1995). *J. Mol. Struct.* **346**, 141–152.
Danecek, P., Kapitán, J., Baumruk, V., Bednárová, L., Kopecký, V. & Bour, P. (2007). *J. Chem. Phys.* **126**, 224513.
Dehareng, D. & Dive, G. (2004). *Int. J. Mol. Sci.* **5**, 301–332.
Derbel, N., Hernández, B., Pflüger, F., Liquier, J., Geinguenaud, F., Jaïdane, N., Lakhdar, Z. B. & Ghomi, M. (2007). *J. Phys. Chem. B*, **111**, 1470–1477.
Falzon, C. T. & Wang, F. (2005). *J. Chem. Phys.* **123**, 214307.
Falzon, C. T., Wang, F. & Pang, W. (2006). *J. Phys. Chem. B*, **110**, 9713–9719.
Feyer, V., Plekan, O., Richter, R., Coreno, M., Prince, K. C. & Carravetta, V. (2008). *J. Phys. Chem. A*, **112**, 7806–7815.
Frisch, M. J. *et al.* (2004). *Gaussian03*, revision D.01. Gaussian Inc., Wallingford, CT, USA.
Ganesan, A., Brunger, M. J. & Wang, F. (2010). *Nucl. Instrum. Methods Phys. Res. A*, **619**, 143–146.
Ganesan, A. & Wang, F. (2009). *J. Chem. Phys.* **131**, 044321.
Ganesan, A., Wang, F. & Falzon, C. (2011). *J. Comput. Chem.* **32**, 525–535.
Godbout, N., Salahub, D. R., Andzelm, J. & Wimmer, E. (1992). *E. Can. J. Chem.*, **70**, 560–571.

Gontrani, L., Mennucci, B. & Tomasi, J. (2000). *J. Mol. Struct. (Theochem)*, **500**, 113–127.
Gould, R. O., Gray, A. M., Taylor, P. & Walkinshaw, M. D. (1985). *J. Am. Chem. Soc.* **107**, 5921–5927.
Gritsenko, O. V., Braida, B. & Baerends, E. J. (2003). *J. Chem. Phys.* **119**, 1937–1950.
Herrera, B., Dolgounitcheva, O., Zakrzewski, V. G., Toro-Labbe, A. & Ortiz, J. V. (2004). *J. Phys. Chem. A*, **108**, 11703–11708.
Hirshfeld, F. L. (1991). *Crystallogr. Rev.* **2**, 169–200.
Hu, C. H., Shen, M. & Schaefer, H. F. (1993). *J. Am. Chem. Soc.* **115**, 2923–2929.
Hu, Y. & Bernstein, E. R. (2008). *J. Chem. Phys.* **128**, 164311.
Iijima, K. & Beagley, B. (1991). *J. Mol. Struct.* **248**, 133–142.
Iijima, K., Tanaka, K. & Onuma, S. (1991). *J. Mol. Struct.* **246**, 257–266.
Jensen, J. H. & Gordon, M. S. (1991). *J. Am. Chem. Soc.* **113**, 7917–7924.
Ji, Z., Santamaria, R. & Garzón, I. L. (2010). *J. Phys. Chem. A*, **114**, 3591–3601.
Klasinc, L. (1976). *J. Electron Spectrosc. Relat. Phenom.* **8**, 161–164.
Kumar, S., Kumar Rai, A., Rai, S. B., Rai, D. K., Singh, A. N. & Singh, V. B. (2006). *J. Mol. Struct.* **791**, 23–29.
Kumar, S., Rai, A. K., Singh, V. B. & Rai, S. B. (2005). *Spectrochim. Acta A*, **61**, 2741–2746.
Leeuwen, R. van & Baerends, E. J. (1994). *Phys. Rev. A*, **49**, 2421–2431.
Lesarri, A., Cocinero, E. J., López, J. C. & Alonso, J. L. (2004). *Angew. Chem. Int. Ed.* **43**, 605–610.
Lesarri, A., Sánchez, R., Cocinero, E. J., López, J. C. & Alonso, J. L. (2005). *J. Am. Chem. Soc.* **127**, 12952–12956.
Linder, R., Nispel, M., HäBer, T. & Kleinermanns, K. (2005). *Chem. Phys. Lett.* **409**, 260–264.
Linder, R., Seefeld, K., Vavra, A. & Kleinermanns, K. (2008). *Chem. Phys. Lett.* **453**, 1–6.
McCarthy, I. E. & Weigold, E. (1991). *Rep. Prog. Phys.* **54**, 789.
Ning, C. G., Hajgató, B., Huang, Y. R., Zhang, S. F., Liu, K., Luo, Z. H., Knippenberg, S., Deng, J. K. & Deleuze, M. S. (2008). *Chem. Phys.* **343**, 19–30.
Plekan, O., Feyer, V., Richter, R., Coreno, M., Simone, M., Prince, K. C. & Carravetta, V. (2007). *J. Phys. Chem. A*, **111**, 10998–11005.
Powis, I., Rennie, E. E., Hergenhahn, U., Kugeler, O. & Bussy-Socrate, R. (2003). *J. Phys. Chem. A*, **107**, 25–34.
Sabino, A. S., De Sousa, G. P., Luz-Lima, C., Freire, P. T. C., Melo, F. E. A. & Mendes Filho, J. (2009). *Solid State Commun.* **149**, 1553–1556.
Saha, S., Wang, F., Falzon, C. T. & Brunger, M. J. (2005). *J. Chem. Phys.* **123**, 124315.
Schaftenaar, G. & Noordik, J. H. (2000). *J. Comput. Aided Mol. Des.* **14**, 123–134.
Schipper, P. R. T., Gritsenko, O. V., van Gisbergen, S. J. A. & Baerends, E. J. (2000). *J. Chem. Phys.* **112**, 1344–1352.
Selvam, L., Vasilyev, V. & Wang, F. (2009). *J. Phys. Chem. B*, **113**, 11496–11504.
Shirazian, S. & Gronert, S. (1997). *J. Mol. Struct. (Theochem)*, **397**, 107–112.
Slavícek, P., Winter, B., Faubel, M., Bradforth, S. E. & Jungwirth, P. (2009). *J. Am. Chem. Soc.* **131**, 6460–6467.
Stepanian, S. G., Reva, I. D., Radchenko, E. D. & Adamowicz, L. (1998). *J. Phys. Chem. A*, **102**, 4623–4629.
Stepanian, S. G., Reva, I. D., Radchenko, E. D. & Adamowicz, L. (1999). *J. Phys. Chem. A*, **103**, 4404–4412.
Stepanian, S. G., Reva, I. D., Radchenko, E. D., Rosado, M. T. S., Duarte, M. L. T. S., Fausto, R. & Adamowicz, L. (1998). *J. Phys. Chem. A*, **102**, 1041–1054.
Tulip, P. R. & Clark, S. J. (2004). *J. Chem. Phys.* **121**, 5201–5210.
Velde, G. t., Bickelhaupt, F. M., Baerends, E. J., Guerra, C. F., Gisbergen, S. J. A., v, Snijders, J. G. & Ziegler, T. (2001). *J. Comput. Chem.* **22**, 931–967.
Wang, F. (2003). *J. Phys. Chem. A*, **107**, 10199–10207.
Wang, F. (2007). *J. Phys. Chem. B*, **111**, 9628–9633.
Wang, F. & Pang, W. (2007). *Mol. Simul.* **33**, 1173–1185.