Intramolecular interactions of L-phenylalanine revealed by inner shell chemical shift

Aravindhan Ganesan and Feng Wang*

Centre for Molecular Simulation, Swinburne University of Technology, P. O. Box 218, Hawthorn, Melbourne, Victoria, Australia, 3122.

*Email: fwang@swin.edu.au

Tel: +61 3 9214-5065

Fax: +61 3 9214 5075

Abstract:

Intramolecular interactions of the functional groups, carboxylic acid, amino and phenyl in L-phenylalanine (Phe) have been revealed through inner shell chemical shift. The chemical shift and electronic structures are studied using its derivatives, 2-phenethylamine (PEA) and 3-phenylpropionic acid (PPA), through substitutions of the functional groups on the chiral carbon C_α, i.e., carboxylic acid (-COOH) and amino ($-NH_2$) groups. Inner shell spectra of L-phenylalanine are simulated using density functional theory (DFT) based B3LYP/TZVP and LB94/et-pVQZ models, which achieve excellent agreement with the most recently available synchrotron sourced X-ray photoemission spectroscopy of L-phenylalanine (Elettra, Italy). The present study reveals insight into behaviour of the peptide bond (CO-NH) through chemical shift of the C_1-C_α-C_β(-C_γ) chain and intramolecular interactions with phenyl. It is found that the chemical shift of carbonyl C_1(=O) site exhibits an apparently red-shift (smaller energy) when interacting with the phenyl aromatic group. Removal of the amino group ($-NH_2$) from L-phenylalanine (which forms PPA) brings this energy on C_1 close to that in L-alanine (δ<0.01 eV). Chemical environment of C_α and C_β exhibits more significant differences in L-alanine than in the aromatic species, indicating that the phenyl group indeed affects the peptide bond in the amino acid fragment. No direct evidences are found that the carbonyl acid and amino group interact with the phenyl ring through conventional hydrogen bonds.

Key Words: Chemical shift, functional groups, hydrogen bond, L-phenylalanine, 3-phenylpropionic acid, 2-phenethylamine, amino acids.

I. INTRODUCTION

Functionality and selectivity of biologically active molecules are strongly influenced by their electronic properties, shape, conformation, charge distribution as well as their interactions with environment. In biology, genetic codes among very different species differ only in a small percentage, usually about 5%. There are twenty naturally existing amino acids which make up to vastly different proteins by often small local modifications. Intrinsic properties of biomolecules are often hidden within their environment, such as macroscopic solvent effects, interactions with other molecules and constraints imposed by macromolecular backbones etc.[1] As a result, it is important to understand intrinsic molecular properties of these biomolecular building blocks of life in the gas phase, before we could reveal details of fundamental biochemical characteristics and interactions resulting from biological environment.

Until recently, practical barriers which prevent studies of building blocks of life from details, have been partly overcome if not removed, due to recent development in both experimental capability, such as synchrotron sourced X-ray spectroscopy[2-6] , and theoretical models and supercomputing resources [7-11]. Thus gas phase information will be useful as it provides deeper insight into such intrinsic molecular properties[7] so that interactions and behaviour of larger molecules can be understood and predicted. As a result, gas phase studies of small biomolecules and peptides have attracted a considerable research attention in recent years, since experimental evidences are able to verify and to stimulate development in theory. Therefore, one can achieve better understanding of short peptide chains and competition between secondary structures can be revealed with confidence.

All amino acids in the gas phase exhibit neutral structures with carboxyl acid (-COOH) and amino ($-NH_2$) terminal groups, rather than a zwitterionic structure as

they are in solid state and/or in aqueous phase[1]. In our previous study, the role of methyl in L-alanine was revealed against glycine[10]. Chemically, L-alanine can be considered that three functional groups, methyl ($-CH_3$), carboxyl acid (-COOH) and amino ($-NH_2$), join at the α-carbon ($*C_\alpha$). If one of the hydrogen atoms of the methyl moiety in L-alanine is replaced by a phenyl, L-alanine chemically becomes L-phenylalanine (Phe), an aromatic amino acid which is one of the most abundant aromatic acids in proteins.[12]

Aromatic amino acids exhibit hydrophobic characteristics which play a vital role in many of the biological processes in the living cells.[13] L-Phenylalanine has received great attention both theoretically and practically.[5, 14-17] Recent data mining projects have shown that the amide-aromatic interactions of phenylalanine are highly essential in stabilization of protein residues over large configurational spaces.[18] It is also observed that the orientation of functional groups connected at the α-carbon plays a vital role in overall structure of the molecule. A recent study recognised[16] that the most populated conformer of L-phenylalanine in gas phase is stabilized by hydrogen bond which is enhanced by the aromatic ring. In particular, a most recent synchrotron sourced soft X-ray photoemission (XPS)[5] experiment enables us to further explore information in the inner shell electronic structure of L-phenylalanine with confidence.

2-phenethylamine (PEA) [19-21] and 3-phenylpropionic acid (PPA) [22, 23] are not only derivatives of L-phenylalanine, but also important compounds themselves. 2-phenylethylamine (PEA) is a metabolite of the monoamine oxidase inhibitor (MAOI) antidepressant phenelzine. 3-Phenylpropionic acid (PPA) is an intermediate for the MAOI antipressant drug Indinavir[24] and it is also a key compound that reveals mechanism of β-oxidation for degradation of fatty acids[25]. Detailed studies of

interactions of the functional networks through model molecules will certainly lead to a much improved understanding of the local environment and its influence on the structural and dynamical behaviour of proteins.

Core ionization potentials (IP) are a property known to be very sensitive to chemical environment[26]. A study of Auger processes of bromine substitute in DNA indicated[8] that a major part of the biological enhancement is caused by inner-shell photo-ionization. Binding energies of inner-shell electrons depend on charge distributions in a molecule, and the ability of neighbouring atoms to screen positive charge introduced through ionization.[8, 27] Recent state-of-the-art third generation synchrotron facilities are able to probe electronic structures in the inner-shells of biological molecules[2, 3, 5, 6, 28-30] and to relate inner shell structures with hydrogen bonds[26]. As a result, studies of inner-shell will be useful to investigate fragment correlation[7-9, 29] and to reveal chemical shift of backbone carbon chain C_1-C_α-C_β in amino acids and/or peptides responses to substitutions on the α-carbon. In the present study, interactions of carboxyl acid, amino and phenyl in Phe are revealed through inner shell chemical shift and hydrogen bond affected by the functional groups in Phe and its derivatives PEA and PPA.

II. METHODS AND COMPUTATIONAL DETAILS

Geometrical parameters of the molecules in their ground electronic states, that is, L-phenylalanine (Phe), 3-phenylpropionic acid (PPA), 2-phenethylamine (PEA), together with L-alanine (Ala) and benzene were obtained using the density functional theory (DFT) B3LYP/TZVP model, which is implemented in the Gaussian 03 suite of computational chemistry programs.[31] The triple zeta valence polarized basis set of Godbout *et al*. [32] is employed, which not only produces good energy properties, but

also generate good quality molecular orbitals as demonstrated in our previous studies[33].

Corresponding stationary points to minimum or transition states have been checked by analytic calculations of harmonic vibrational frequencies at the same level of theory: absence of imaginary values for the minimum structures. No imaginary frequencies were obtained, indicating that the optimized structures are true minima. Vertical core ionization energies of the molecules were calculated using the LB94/et-pVQZ model[34, 35], which is embedded in the Amsterdam Density Functional Theory (ADF) chemistry package[36] without further modification or scaling.

III RESULTS AD DISCUSSIONS

III.1 Canonical conformer structures in gas phase

An interactive three dimensional (3D on a computer or online) structural diagram with relevant energies for the canonical structures of L-phenylalanine (Phe), 3-phenylpropionic acid (PPA), 2-phenethylamine (PEA), L-alanine and benzene in their ground electronic states (X^1A) are given in Figure 1. The 3D-PDF technique has been recently developed and detailed elsewhere[37]. The nomenclature of the alaninyl (-$CH_2CH(NH_2)COOH$) carbon backbone, C_β-C_α-C_1, is marked in this figure. Note that carbonyl C_1(=O) of -COOH in the present work labelled was as C_γ in reference 5. To activate the 3D structures on a computer (or on-line), double click the molecules to view the 3D structures. As shown in this figure, when a phenyl replaces one of the hydrogen atoms (H_β) at the C_β site of methyl group in L-alanine (Ala), which forms L-phenylalanine (Phe), the process associates with a large energy drop of -233.31 eV (using the B3LYP/TZVP model). When Phe loses an amino (NH_2), which produces 2-phenethylamine (PEA), an energy increase of 55.38 eV is obtained. However, if

Phe loses the carboxylic acid group to form 3-phenylpropionic acid (PPA) instead, the associated energy increases approximately three times, which gives an energy increase of 188.65 eV. This trend in energy is similar to the changes in dipole moments (marked on the structures).

Selected structural parameters from the optimized structures and their dipole moment are reported in Table 1. The canonical structures obtained using the B3LYP/TZVP model are in good agreement with other gas phase studies such as L-alanine[3, 38] and L-phenylalanine using the B3LYP/6-311++G**[16]. For Phe, there are no large discrepancies in parameters between the B3LYP and MP2 models, as seen in this table. Changes in the alaninyl network do not impose significant structural parameter changes. All bonds listed in the table, such as N-C_α, C_α-C_β, C_α-C_1, C_β-C_γ and C_1=O remain almost unchanged in the molecules. However, the C_1-O (H) bond shrinks in Phe and expands slightly in the PPA with respect to Ala. For example, the C_1-O(H) bond shrinks from 1.36 Å in Ala to 1.34 Å in Phe, whereas this bond length in PPA slightly increases, from 1.36 Å in Ala to 1.37 Å in PPA. In addition, the C_α-C_1 bond in PPA is apparently shorter (1.52 Å) than other molecules (1.54-1.55 Å) in this table. It is noted that the phenyl ring perimeter, R_6 (sum of all bond lengths consisting of the phenyl ring[39]) of Phe, PPA and PEA remains unchanged at the value of 8.36 Å, which is slightly longer than R_6 of an isolated benzene of 8.34 Å using the same model. It may indicate that the impact of phenyl ring to the carboxylic acid-amino (alaninyl) network is more apparent than the impact of the carboxylic acid-amino network to the phenyl ring, since the phenyl ring could serve as a buffer to resist the changes. Bond angles and dihedral angles show marginal deviations among the molecules. Comparing to Ala, apparent changes up to 5° (centred on the α-

carbon) increases in $\angle N\text{-}C_\alpha\text{-}C_\beta$ and decreases in $\angle N\text{-}C_\alpha\text{-}C_1$ angles are revealed in the present work. Other angles of the molecules exhibit only small relaxation.

Dipole moment of the model molecules reveals significant changes in their electron charge distribution. For example, dipole moment increases significantly from 1.30 D in L-Ala to 5.19 D in L-Phe (B3LYP/TZVP model). It is evident that removal of carboxylic acid (producing PEA) remarkably reduces the polarity of the product than removal of the amino group (producing PPA). Dipole moment of the former (PEA) is given by 1.26 D, a significant drop from 5.19 D in Phe, whereas the latter (PPA) is given by 4.10 D using the same model. Dipole moment of the four molecules indicates that all associated functional groups, phenyl, -COOH and $-NH_2$, affect the electron charge distribution of the molecules in their own ways.

III.2 Simulated and measured inner shell spectra

Inner shell spectra of the molecules clearly indicate the relationship between chemical shift and their chemical environment. Inner shell vertical ionization potentials (IP) of benzene and Ala (C-K) in their ground electronic states are given in Table 2, together with available theoretical and experimental results for comparison purposes. Agreement between theory and experiment for benzene is excellent --- the discrepancy between the calculated and the measured is within 0.7 eV, which is the same scale to the experimental resolution of even synchrotron sourced spectroscopy.[2, 4, 5] Although in the theory a number of approximations are employed, such as neglect of self-energy and relativistic effects, and the application of "meta-Koopman" theorem which approximates the vertical IPs as negative orbital energies, hereby ignores orbital relaxation, the errors in IP introduced by the present model are relatively small. Note that the inner shell of benzene split into four peaks with 1, 2, 2,

1 fold of degeneracy, which under the experimental resolution will produce a single peak at approximately 290 eV in the binding energy spectrum of benzene[30].

The simulated C-K spectra are given in Figure 2(a) for Ala and Figure 2(b) for Phe, which are compared with most recently available synchrotron sourced x-ray photoemission spectroscopy (XPS) in gas phase[2, 5]. In the simulated C-K spectra, full width at half maximum (FWHM) of 0.4 eV were employed in order to reproduce the experiment spectra as much as possible. A binding energy red shift (to the lower energy side) of 1.10 eV and 0.67 eV, respectively, has been applied to the simulated Ala and Phe spectra. This scheme is similar to the previously simulated inner shell spectra of adenine[8], in order to reduce certain systematic errors introduced in the model.

As seen in this figure, the agreement (after energy shift) with the experiment is excellent, in particular, the relevant C-K peak positions of $(C_\gamma) < C_\beta < C_\alpha < C_1$, for both Ala and Phe. The spectra exhibit competitive accuracy with more computational expensive models such as ΔDFT in StoBe[2, 5]. It is noted that the largest IP discrepancy between the present work and measurement in both Ala and Phe is that of C_1(=O). The discrepancy on C_1 may indicate that apart from effects such as self-energy, core orbitals relaxation and relativistic effects, Ala and Phe possess flexible single bonds (side chain) which lead to a number of conformers that may be well populated in experiment and the present model focuses only on one conformer (i.e., the canonical conformer). According to Feyer et al[2], the conformers may contribute as much as 10% towards the measurement, which accounts for approximately 0.5 eV for C-K spectral peaks and even larger for N-K and O-K spectral peaks. Figure 3(a) and (b) compare the simulated N-K and O-K spectra of L-Phen with a most recent synchrotron sourced experiment[5] in gas phase. The N-K and O-K spectra of Phe are

also well reproduced in the present work but engage with larger energy shifts than the C-K spectra of the same molecule.

Accurate calculations of inner shell ionization energies of bio-molecules represent a challenge. Of the available models applicable to larger bio-molecular systems, models such as ADC(4)[3], CV-B3LYP[40], StoBe[41], ΔSCF[42], ΔDFT[43] and the modified DFT-B3LYP model plus the self-interaction correction[26] etc, are currently among the most useful models for inner shell spectral simulations. However, all those models are significantly more computationally expensive than our DFT-LB94 model with a single calculation of the ground electronic state with competitive accuracy. For example, the XPS simulations for adenine[2, 8], guanine[4, 44] and purine[8, 43] indicate that our DFT-LB94 model exhibits competitive accuracy to the significantly more expensive ADC(4) [3, 4] and StoBe[8] models, in particular, for the C-K spectra. It is also demonstrated that the present model produces accurate inner shell IPs for cytidine[29] with competitive accuracy with a recently developed model of CV-B3LYP[40]. The present model can be more attractive than the modified DFT-B3LYP plus self interaction[26] incorporated with a semi-empirically introduced parameter of ζ for larger bio-molecules.

As previously pointed out by Uejio et al[45], a more realistic prediction of the IPs for flexible and conformer rich bio-molecules is not just considering the stable conformers with energy minima on the potential energy landscape applying the Boltzmann averaged spectrum of these dominant conformers, since merely analysing these conformers may not be adequate to accurately produce core-level spectra[45]. Rather, the application of molecular dynamic (MD) simulations, which is our next target, has the advantage of not merely sampling the lowest energy conformers, but also phase space the molecule would have to occupy as it changes conformers.

III. 3 Chemical shift and inner shell spectra of the model molecules

Chemical shift is very sensitive to local chemical environment of an atom and could serve as an indicator for the strength of a hydrogen bond[26]. Table 3 reports the inner shell vertical ionization potentials (IPs) of the aromatic species in their most stable configurations. For L-Phe, the present IP results are compared with recent results generated using the ADC(4) model and with the XPS experiment[5]. Excellent agreement with both theory and experiment is achieved, in particular, in the interesting C-K spectrum. The IPs of the aromatic molecules always follow a trend of C(Phe)>C(PPA)≥C(PEA), except for the lowest IP C-site (phenyl) where the C sites of PPA is smaller than that of PEA. It indicates that the carboxyl acid moiety causes larger chemical shift than the amino group.

As observed previously, amino acid side chain reduces the high symmetry of benzene so that the phenyl C-K IPs split in Phe, which is also seen in PPA when the amino group, NH_2, of the side chain is replaced by an H atom. However, the phenyl C-K IPs in PEA where the carboxyl acid group is removed, exhibit accident 1, 2 and 3 folds degeneracy in the phenyl C-K IPs (last column of Table 3). It is noted an apparent energy drop at C_α and C_γ sites is observed, to respond to the removal of functional groups in the amino acid chain. For example, there is an energy drop of approximately 1 eV at the C_α sites in PPA and PEA with respect to Phe. However, removal of the functional groups in the amino acid fragment of Phe, results in energy shift at C_β site in opposite directions: removal of -NH_2 causes a small energy drop of 0.08 eV in PPA, whereas removal of -COOH causes a energy rise of 0.49 eV in PEA. The latter leads to the C_β position below the energy of C_γ of a phenyl site in PEA.

Simulated C-K spectra of the aromatic molecules, together with benzene and Ala, which are calculated using the LB94/et-pVQZ model, are given in Figure 4. It is obvious that the C1s energy patterns of L-Phe and PPA are very similar to that of Ala, except that in the L-alanine spectrum the phenyl carbon peak does not exist. In this energy region of 289.5-290.5 eV, interactions between the alaninyl and phenyl groups reduce the high symmetry of benzene (D_{6h}), causing the degenerated phenyl core states to split, as indicated by Wang[7]. That is, the symmetric phenyl spectral peak in benzene is asymmetric in Phe, PPA and PEA, due to the closely located side chain C_β peak. The C-K peak positions shift to accommodate the side chain alternation in the model molecules but the energy order of the carbon peaks remains unchanged as: C_γ (phenyl carbon), $C_\beta < C_\alpha < C_1$. The spectral patterns of Phe and PPA are similar to that of L-Ala, except for lack of the phenyl band around 289.8 eV in Ala. However, the chemical environment of C_β-C_α-C_1 in Phe exhibits subtle differences to those in Ala. For example, the C_α and C_β peaks exhibit blue shift (larger energy) but the C_1 peak shifts to lower energy side. Chemical shift $\Delta\varepsilon$ (C_α-C_1) in Phe of 1.5 eV is smaller than that of 2.2 eV in Ala, reflecting apparent effects on the C_β-C_α-C_1 side chain.

C-K spectra of the model molecules in Figure 4 reveal the relationship of the amino acid side chain and phenyl ring. The former leads to small phenyl C-K energy splitting. The phenyl ring acts like a buffer to resist changes caused by the functional groups in PPA and PEA. Chemical environment of the C_α sites in PPA and PEA is closer to each other, since this carbon in PPA and PEA is not a chiral carbon. Hirshfeld charges of the molecules which are listed in Table 4 also reflect such trends. As seen in other bio-molecules such as adenine[8] and guanine[41], all N and O sites of the compounds possess negative charges. However, unlike adenine and

guanine, the charge of the C sites can be both positive as well as negative, which depends on the role of the carbon in the molecules. In amino acid pair, Ala and Phe, C_β site is the only carbon site which exhibit negative Hirshfeld charges. If either –COOH or $–NH_2$ group is removed, the C_α carbons in PPA and PEA are no longer chiral carbons and become negatively charged to balance the positive charges. However, the role of phenyl carbons is different in benzene and in Phe, PPA and PEA. In benzene, all carbons are negatively charged to balance the positively charged hydrogen sites; whereas in Phe, PPA and PEA, the phenyl carbons are positively charged so that to balance the negative charges on the –COOH and $–NH_2$ groups.

An orbital based inner shell energy diagram of the molecules is given in Figure 5. Reactions to intramolecular interactions on the carbon sites are observed. For example, C_α and C_β, with respect to L-Ala, and C_γ with respect to benzene, shift up in Phe, but down in PPA and down again in PEA. The C_1 site, exhibits however, an opposite trend: it shifts down in Phe but up in PPA. Removal of the amino group, -NH_2, from the C_α site (produces PPA) causes a significant C_α energy drop of -0.87 eV, whereas removal of this amino group only causes a small secondary energy drop of -0.07 eV on the C_β site. On the other hand, removal of carboxyl acid from the C_α site of Phe causes an even larger energy drop of -1.21 eV, and an energy drop of -0.49 eV on the C_β site. In this figure, the IPs of C_α and C_β do not mix up with the phenyl carbon sites in Phe and PPA, but a significant energy drop of C_β in PEA places the C_β of PEA well below the energy of the first phenyl carbon, C_γ. As a result, it likely causes confusion when assigning experimental PEA spectra for C_β and C_γ. Significant changes to pattern of the energy spectra occur among L-Phe, PPA and PEA. For example, the energy gap between C_α and C_β is similar in Ala and Phe, both are ca. 1.42 eV but it is significantly reduced in PPA and PEA to about 0.50 eV.

III.4 Near-atom interactions and H-bond

Apart from energy and charge, near-atom interactions due to conformational variations in 3D structures of the molecules are important, as the H····A distances vary hereby creating certain H-bonds or eliminating other H-bonds. A number of studies[16, 46, 47] on Phe suggested that H-bonds are responsible for the interactions in the side chain[16, 46, 47] as well as interactions between the side chain and the phenyl ring[45, 46]. For example, Lee et al.[47] suggested that in the three lowest energy conformers of Phe, "an amino hydrogen atom interacts with the phenyl ring in what appear to be a π-hydrogen bond, thereby forming a complete sequence of hydrogen bonds, COOH→ NH_2→π". Moreover, in Ref. 46, it is suggested that electrostatic interactions between a carboxyl oxygen and an aromatic hydrogen appear in a number of Phe conformers including the lowest and second lowest energy conformers. In the most recent comprehensive study for the conformational landscape of Phe[16], however, it does not suggest such interactions with the phenyl ring existing in the ten lowest energy conformers.

In the present study, all possible H····A distances of the molecules where A represents for O or N atoms, are examined and reported in Table 5. It is clearly shown that the phenyl ring changes the orientation of the amino acid side chain in 3D space and therefore, through space H-bond interactions as well as through bond interactions are re-organized in the aromatic amino acid. An only unchanged H····A distance within the 2.80 Å cutoff criterion[16] is the H_α····N distance, in Ala, Phe and PEA. This H_α····N "bond" is not a strict H-bond, as both N and H_α connect with the same atom C_α in a tetrahedral structure. None of the possible H-bonds (associated with different atoms within 2.80 Å) in Ala maintains as H-bonds in L-Phe, PPA and PEA. Interestingly, the possible H-bonds presented in Phe as well as in PPA are all formed

by the H····A pairs which are not possible H-bonds in Ala. Therefore, it is evident that the H-bond interactions in Phe and PPA are from a completely different set of H-bond interactions from Ala, due to their 3D structures in space.

The H····A pairs given in this table also indicates that in the most stable conformer of Phe, it is unlikely that the hydrogen atoms in $-NH_2$ could interact with the C_γ-C bond in the phenyl ring as suggested in Ref 47. Rather, apart from the only H_α····N "hydrogen bond" which inherit from Ala and the newly formed OH····N bond, Phe (and PPA) forms two strong H-bonds, one is (C=)O····H_β and the other is (C=)O····H_{ring} on the phenyl ring. The (C=) O····H_{ring} hydrogen-bond has been also suggested by Ref. 46 as electrostatic interaction between a carboxyl oxygen and an aromatic hydrogen. The (C=)O····H_β can be examined in a proper 3D environment, as provided in the 3D-PDF (Figure 1).

IV Conclusions

Interactions and impact between amino acid network and phenyl in L-phenylalanine (Phe) have been examined with respect to its model molecules, 3-phenylpropionic acid (PPA) and 2-phenethylamine (PEA), through their inner shell chemical shift, ionization spectra and hydrogen bond network. The interactions in L-phenylalanine are dominated by three important functional groups of phenyl, carboxyl acid and amino group. Present study further reveals that the H-bond and intra-molecular interaction network of L-alanine and L-phenylalanine are significantly different.

The binding energy spectra show very different chemical shifts in C_α, C_β and C_1 sites in 3-phenylpropionic acid (PPA) and 2-phenethylamine (PEA) in the absence of carboxyl acid and amino group, with respect to L-phenylalanine. The inner shell

spectra highlight differences in electronic environment of the three aromatic species, which provides a unique insight into the behaviour of functional groups of L-phenylalanine. Strong through bond and through space interactions presented in L-phenylalanine can be more comprehensively understood when combining information in valence space, which is currently under investigation.

Acknowledgements

FW thanks the Vice-Chancellor's Research Award at Swinburne University of Technology, which makes the Ph.D. scholarship of AG possible, and Dr. C. Falzon for providing initial information for this project. The authors acknowledge Dr. K. Prince of Elettra Sincrotrone (Italy) for providing their XPS experimental digital traces for L-alanine and L-phenylalanine. Finally, we acknowledge National Computational Infrastructure (NCI) at Australian National University for the award under the Merit Allocation Scheme and Swinburne University Supercomputing Facilities.

References

[1] M. S. de Vries and P. Hobza, Annual Review of Physical Chemistry **58** (1), 585 (2007).

[2] V. Feyer, O. Plekan, R. Richter, M. Coreno, K. C. Prince, and V. Carravetta, The Journal of Physical Chemistry A **112** (34), 7806 (2008).

[3] 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, Chemical Physics **347** (1-3), 360 (2008).

[4] O. Plekan, V. Feyer, R. Richter, M. Coreno, Vall-Ilosera, K. C. Prince, A. B. Trofimov, I. L. Zaytseva, Moskovskaya, E. V. Gromov, and J. Schirmer, (submitted, 2009).

[5] W. Zhang, V. Carravetta, O. Plekan, V. Feyer, R. Richter, M. Coreno, and K. C. Prince, (submitted, 2009); K. Prince (private communications, 2009).

[6] K. Kaznacheyev, A. Osanna, C. Jacobsen, O. Plashkevych, O. Vahtras, Agren, V. Carravetta, and A. P. Hitchcock, The Journal of Physical Chemistry A **106** (13), 3153 (2002).

[7] F. Wang, Journal of Physics: Conference Series **141**, 012019 (2008).

[8] S. Saha, F. Wang, J. B. MacNaughton, A. Moewes, and D. P. Chong, Journal of Synchrotron Radiation **15** (2), 151 (2008).

[9] F. Wang, Q. Zhu, and E. Ivanova, Journal of Synchrotron Radiation **15** (6), 624 (2008).

[10] T. F. Chantal, W. Feng, and P. Wenning, Journal of Physical Chemistry B **110**, 9713 (2006).

[11] L. Triguero, O. Plashkevych, L. G. M. Pettersson, and H. Ågren, Journal of Electron Spectroscopy and Related Phenomena **104** (1-3), 195 (1999).

[12] M. M. Ilczyszyn, T. Lis, M. Wierzejewska, and M. Zatajska, Journal of Molecular Structure **919** (1-3), 303 (2009).

[13] J. J. Robinet, C. Baciu, K.-B. Cho, and J. W. Gauld, The Journal of Physical Chemistry A **111** (10), 1981 (2007).

[14] L. M. Ghiringhelli and L. Delle Site, Journal of the American Chemical Society **130** (8), 2634 (2008).

[15] A. Kaczor, I. D. Reva, L. M. Proniewicz, and R. Fausto, The Journal of Physical Chemistry A **110** (7), 2360 (2006).

[16] Z. Huang, W. Yu, and Z. Lin, Journal of Molecular Structure: THEOCHEM **758** (2-3), 195 (2006).

[17] T. Hashimoto, Y. Takasu, Y. Yamada, and T. Ebata, Chemical Physics Letters **421** (1-3), 227 (2006).

[18] G. Duan and V. H. Smith, International Journal of Quantum Chemistry **90** (2), 669 (2002).

[19] J. J. Urban, C. W. Cronin, R. R. Roberts, and G. R. Famini, Journal of the American Chemical Society **119** (50), 12292 (1997).

[20] J. Yao, H. S. Im, M. Foltin, and E. R. Bernstein, The Journal of Physical Chemistry A **104** (26), 6197 (2000).

[21] P. R. Richardson, S. P. Bates, and A. C. Jones, The Journal of Physical Chemistry A **108** (7), 1233 (2004).

[22] R. T. Kroemer, K. R. Liedl, J. A. Dickinson, E. G. Robertson, J. P. Simons, D. R. Borst, and D. W. Pratt, Journal of the American Chemical Society **120** (48), 12573 (1998).

[23] J. L. Castro, M. R. L. Ramírez, J. F. Arenas, and J. C. Otero, Journal of Molecular Structure **744-747**, 887 (2005).

[24] J. G. de Vries, G. Roelfes, and R. Green, Tetrahedron Letters **39** (45), 8329 (1998).

[25] L.-F. Mao, C. Chu, and H. Schulz, Biochemistry **33** (11), 3320 (1994).

[26] G. Tu, Y. Tu, O. Vahtras, and H. Agren, Chemical Physics Letters **468** (4-6), 294 (2009).

[27] M. Abu-samha, K. J. Borve, J. Harnes, and H. Bergersen, The Journal of Physical Chemistry A **111** (37), 8903 (2007).

[28] J. Magulick, M. M. Beerbom, B. Lagel, and R. Schlaf, The Journal of Physical Chemistry B **110** (6), 2692 (2006).

[29] S. S. Alexander Thompson, Feng Wang, Takashi Tsuchimochi, Ayako Nakata, Yutaka Imamura, Hiromi Nakai, Bulletin of the Chemical Society of Japan **82** (2), 187 (2009).

[30] I. Powis, E. E. Rennie, U. Hergenhahn, O. Kugeler, and R. Bussy-Socrate, The Journal of Physical Chemistry A **107** (1), 25 (2003).

[31] G. W. T. M. J. Frisch, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G.

Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT (2009).

[32] Nathalie Godbout, Dennis R. Salahub, Jan Andzelm, and E. Wimmer, Canadian Journal of Chemistry **70** (2), 560 (1992).

[33] F. Wang, The Journal of Physical Chemistry A **107** (47), 10199 (2003).

[34] E. Van Lenthe and E. J. Baerends, Journal of Computational Chemistry **24** (9), 1142 (2003).

[35] D. P. Chong, E. Van Lenthe, S. Van Gisbergen, and E. J. Baerends, Journal of Computational Chemistry **25** (8), 1030 (2004).

[36] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler, Journal of Computational Chemistry **22** (9), 931 (2001).

[37] L. Selvam, V. Vasilyev, and F. Wang, The Journal of Physical Chemistry B (accepted, 2009).

[38] T. F. Chantal, W. Feng, and P. Wenning, Lecture Notes in Computer Science, Eds. V. Alexandrov, D. van Albada, P. Sloot and J. Donggarra, LNCS, Springer, 82 (2006).

[39] F. Wang, M. T. Downton, and N. Kidwani, Journal of Theoretical & Computational Chemistry **4**, 247 (2005).

[40] A. Nakata, Y. Imamura, T. Otsuka, and H. Nakai, The Journal of Chemical Physics **124** (9), 094105 (2006).

[41] K. Hermann, L. G. M. Pettersson, M. E. Casida, C. Daul, A. Goursot, A. Koester, E. Proynov, A. St-Amant, D. R. Salahub., V. Carravetta, H. Duarte, C. Friedrich, N. Godbout, J. Guan, C. Jamorski, M. Leboeuf, M. Leetmaa, M.

Nyberg, S. Patchkovskii, L. Pedocchi, F. Sim, L. Triguero, and A. Vela., StoBe-deMon version 2.2 (2006).

[42] T. Karlsen, K. J. Borve, L. J. Sthre, K. Wiesner, M. Bassler, and S. Svensson, Journal of the American Chemical Society **124** (26), 7866 (2002).

[43] Y. Takahata, A. K. Okamoto, and D. P. Chong, International Journal of Quantum Chemistry **106**, 2581 (2006).

[44] Q. Zhu, F. Wang, and E. Ivanova, Journal of Synchrotron Radiation (2009, in press).

[45] J. S. Uejio, C. P. Schwartz, R. J. Saykally, and D. Prendergast, Chemical Physics Letters **467** (1-3), 195 (2008).

[46] L. C. Snoek, E. G. Robertson, R. T. Kroemer, and J. P. Simons, Chemical Physics Letters **321** (1-2), 49 (2000).

[47] K. T. Lee, J. Sung, K. J. Lee, Y. D. Park, and S. K. Kim, Angewandte Chemie **114**, 4288 (2002).

Table 1 Ground electronic state geometries of L-phenylalanine (Phe), 3-phenylpropionic acid (PPA) and 2-phenethylamine (PEA) obtained using the B3LYP/TZVP and MP2/TZVP (Phe only) models. Results are compared with values taken from the literature.

	Phe			***PPA***	***PEA***		***Ala***[d]
Parameters	**B3LYP**	**MP2**	**Other Work**[e]	**B3LYP**	**B3LYP**	**Other Work**	**B3LYP**
C-O(H)	1.34	1.34	1.34	1.37			1.36
C=O	1.20	1.21	1.20	1.20			1.21
C_α-N	1.47	1.47	1.47		1.46		1.46
C_α-C_β (Å)	1.55	1.54	1.55	1.54	1.55	1.54[c]	1.53
C_β-C_γ (Å)	1.51	1.51	1.51	1.51	1.51	1.46[c]	
C_α-$C_{(1)}$ (Å)	1.55	1.54	1.55	1.52			1.54
R_6(Å)	8.36	8.36	8.38	8.36	8.36		
N-C_α-C_1/°	109.4	108.9	108.9				113.5
O=$C_{(1)}$-O(H)/°	123.0	123.8	123.3	119.7			122.3
(H)O-$C_{(1)}$-C_α/°	113.7	113.4	113.4	115.5			112.0
O=$C_{(1)}$-C_α/°	123.1	122.6	123.2	124.7			125.5
N-C_α-C_β/°	115.8	114.7	116.3		116.7	115.8[c]	111.1
C_α-C_β-C_γ/°	114.2	111.6	114.0	113.6	113.6	111.8[c]	
N-C_α-C_β-C_γ/°	52.4	51.7	52.4		63.1	61.1[c]	
$C_{(1)}$-C_α-C_β-C_γ/°	-73.9	-72.1	-73.7	-75.9			
O=$C_{(1)}$-O- C_α/°	-177.7	-177.9	-177.7	-177.4			-179.2
O=$C_{(1)}$-O-H/°	179.0	179.7	178.0	177.9			178.4
O=$C_{(1)}$-C_α-N/°	-169.8	171.9	-165.3				-168.1
μ (Debye)	5.19	5.44	5.44	4.10	1.26		1.30

[a] B3LYP/6-311++G** calculations. [17]
[b] B3LYP/6-311++G** calculations. [16]
[c] MP2/6-311G** calculations. [20]
[d] B3LYP/TZVP calculations.[10]
[e] B3LYP/6-311++G** calculations.[16]

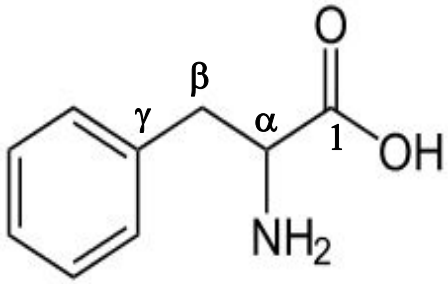
Table 2. Comparison of C1s core ionization potentials (IPs) in the ground electronic states of L-alanine and benzene calculated using the LB94/et-pVQZ model with available experiments (eV).

Site	Benzene	Expt.[30]	L-Ala	Expt.[2]	Expt.[29]
C_1			293.41	295.3	295.0
Cα			291.18	292.3	292.2
C_β			289.76	291.2	291.0
C-ben	289.502	290.2			
C-ben	289.498				
C-ben	289.498				
C-ben	289.490				
C-ben	289.490				
C-ben	289.486				

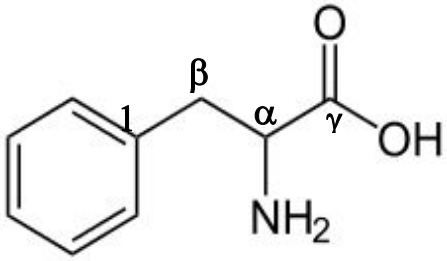
Table 3. Comparison of C1s, N1s and O1s ionization potentials (IPs) in the ground electronic states of L-phenylalanine, PPA and PEA obtained using the LB94/et-pVQZ model with available theory and experiment (eV).

Site	L-Phe			PPA	PEA
	This work	Theory [a 5]	Expt. [a 5]	This work	This work
C_1	293.02	293.61	294.85	293.42	-
C_α	291.49	291.46	291.90	290.62	290.28
C_β	290.08	290.29		290.00	289.59
C_γ (ring)	290.02	289.91	290.30	289.84	289.73
C-(ring)	289.73	289.47		289.53	289.44
C-(ring)	289.71	289.50		289.54	289.48
C-(ring)	289.65	289.57		289.49	289.44
C-(ring)	289.63	289.43		289.48	289.48
C-(ring)	289.56	289.55		289.40	289.44
N	403.70	404.95	405.70	-	402.49
O(=C)	534.36	536.50	538.05	534.69	-
O(-OH)	535.86	538.13	539.87	536.41	-

[a]ΔDFT calculations using BP86 for conformer 1 as the most stable configuration (Ref. 5). The nomenclature of the carbon chain in Phen is given by C_γ-C_β-C_α-C_1(=O) which is labelled by (O=)C_1-C_β-C_α-C_γ in Ref. 5.



L-Phenylalanine used in this work



L-Phenylalanine used in ref. 5[a]

Table 4 Hirshfeld charges of the carbon chains among the model molecules, calculated using the LB94/et-pVQZ model.

	Benzene	L-Ala	L-Phe	PPA	PEA
C_1	-	0.21	0.21	0.22	-
C_α	-	0.02	0.03	-0.06	-0.02
C_β	-	-0.10	-0.07	-0.06	-0.07
C_γ	-0.04	-	0.01	0.01	0.01
N	-	-0.24	-0.21		-0.25
O(=C)	-	-0.30	-0.21	-0.27	-
O(-OH)	-	-0.20	-0.29	-0.29	-

Table 5. H····N and H····O distances of the model molecules. The H····A distances are within the 2.80 Å cutoff distance for near-atom interaction (H-bond) are underlined (Å).

H-Bond	Alanine	Phe	PPA	PEA
N....H_α	<u>2.08</u>	<u>2.08</u>	-	<u>2.08</u>
O-H....N	4.34	<u>1.89</u>	-	-
H-O....H_α	<u>2.49</u>	2.95	2.86	-
H-O....$H_{\beta1}$	<u>2.74</u>	3.96	4.03	-
H-O....Hβ2	3.97	4.51	4.56	-
H-O....H_1-N	3.83	3.43	-	-
H-O....H_2-N	3.98	3.08	-	-
H-O....H_{ring}	-	3.78	4.27	-
C=O....H_1-N	2.89	4.38	-	-
C=O....H_2-N	<u>2.68</u>	4.03	-	-
C=O....$H_{\beta1}$	3.69	<u>2.57</u>	<u>2.58</u>	-
C=O....$H_{\beta2}$	4.27	3.89	3.89	-
C=O....H_{ring}	-	<u>2.68</u>	<u>2.72</u>	-
C=O....H_α	3.19	<u>2.78</u>	2.91	-

Figure Captions:

Figure 1: Chemical structures and nomenclature of L-phenylalanine (Phe), 3-phenylpropionic acid (PPA) and 2-phenethylamine (PEA) and their relevant energies. Double click on a molecular structure online or a computer to activate the 3D-PDF structures (You need Acrobat Adobe 8.0 or above to manipulate the 3D-PDF strcutrues).

Figure 2 (a): Comparison of the simulated and the measured C-K spectra of L-alanine. The simulated spectrum is obtained using the LB94/et-pVQZ model with an FWHM = 0.57 eV. The simulated spectrum was shifted by 1.10 eV to eliminate systematic errors.

Figure 2(b): Comparison of the simulated and the measured C-K spectra of L-phenyalanine. The simulated spectrum is obtained using the LB94/et-pVQZ model with an FWHM = 0.40 eV. The simulated spectrum was shifted by 0.67 eV to eliminate systematic errors.

Figure 3 (a): Comparison of the simulated and the measured N-K spectra of L-phenyalanine. The simulated spectrum is obtained using the LB94/et-pVQZ model with an FWHM = 0.40 eV. The simulated spectrum was shifted by 1.35 eV to eliminate systematic errors.

Figure 3(b): Comparison of the simulated and the measured O-K spectra of L-phenyalanine. The simulated spectrum is obtained using the LB94/et-pVQZ model with an FWHM = 0.40 eV. The simulated spectrum was shifted by 3.61 eV to eliminate systematic errors.

Figure 4: Simulated C-K spectra of the model molecules using the LB94/et-pVQZ model with the FWHM=0.4 eV in the simulations.

Figure 5: Energy diagram of benzene, L-alanine, L-phenylalanine (Phe), 3-phenylpropionic acid (PPA) and 2-phenethylamine (PEA), in the energy range of 289-295 eV. Chemical shift variation of the C site is revealed in this diagram as a function of the functional groups.

Figure 1

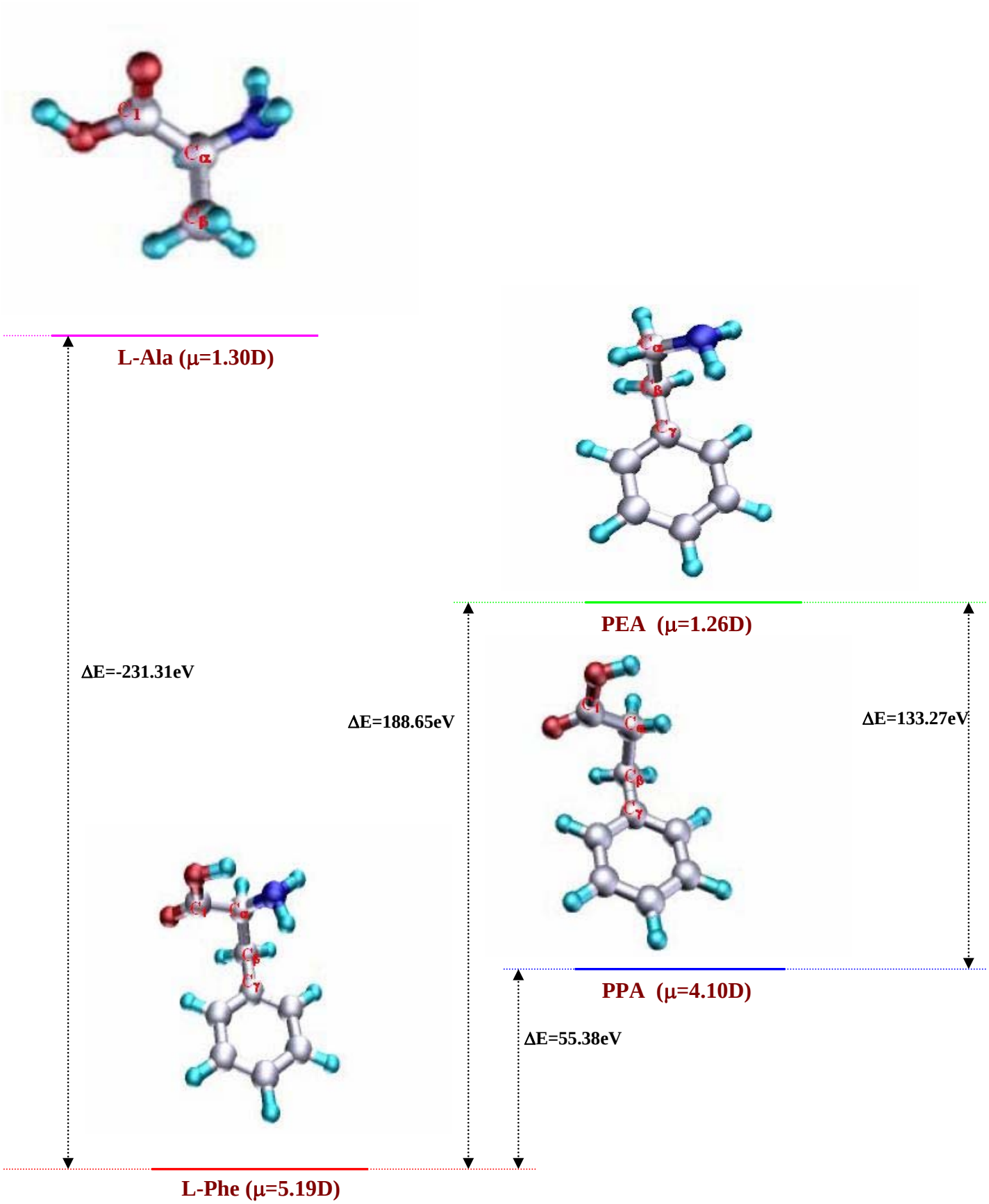

L-Ala (μ=1.30D)
PEA (μ=1.26D)
PPA (μ=4.10D)
L-Phe (μ=5.19D)
ΔE=-231.31eV
ΔE=188.65eV
ΔE=133.27eV
ΔE=55.38eV

Figure 2 (a) Ala C-K

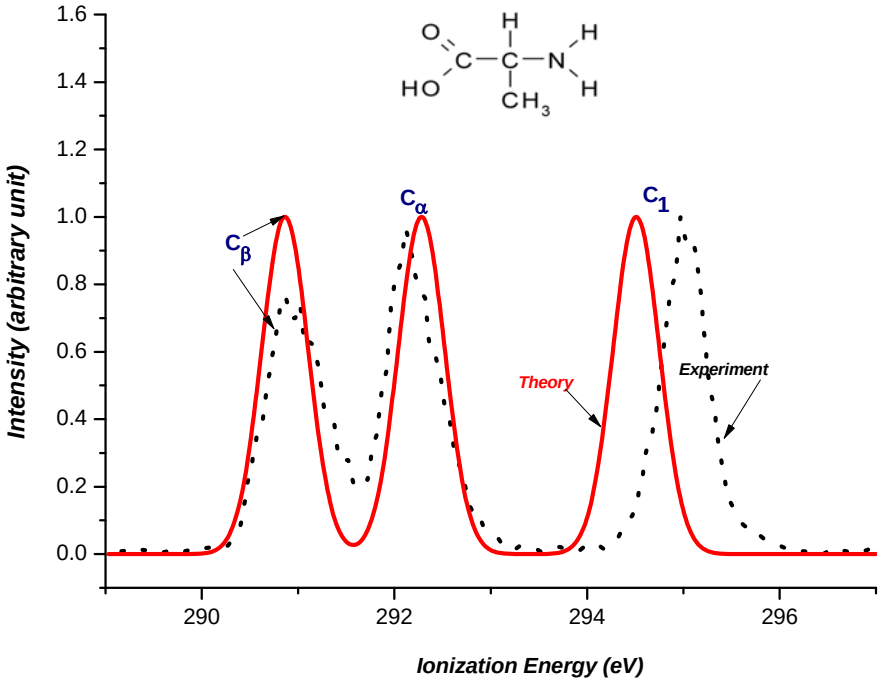


Figure 2 (b) Phe C-K

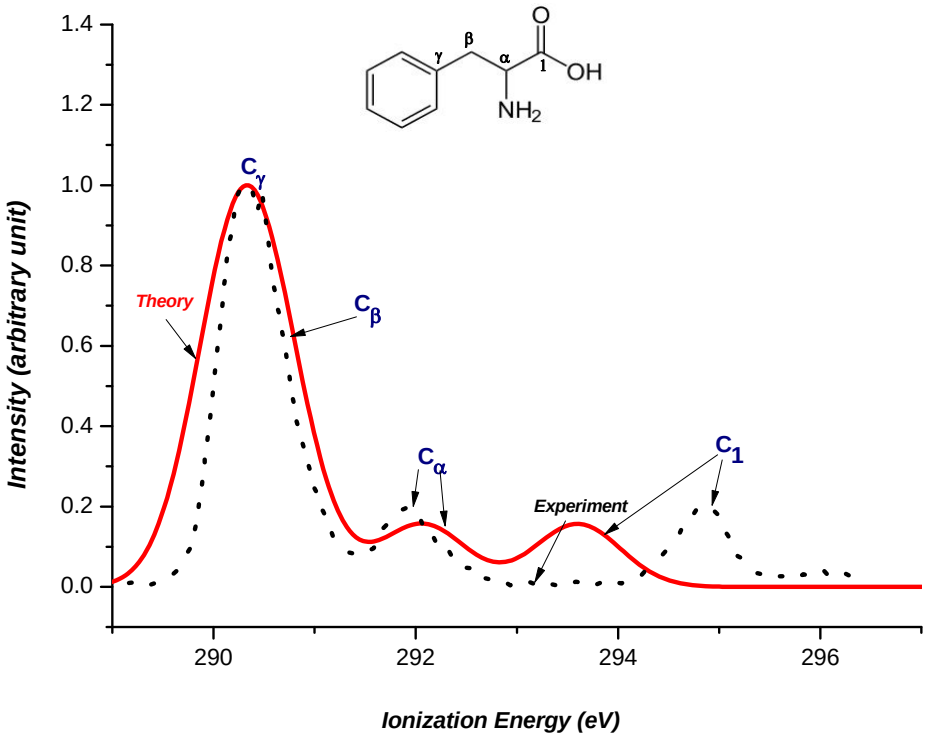


Figure 3(a) Phe N-K

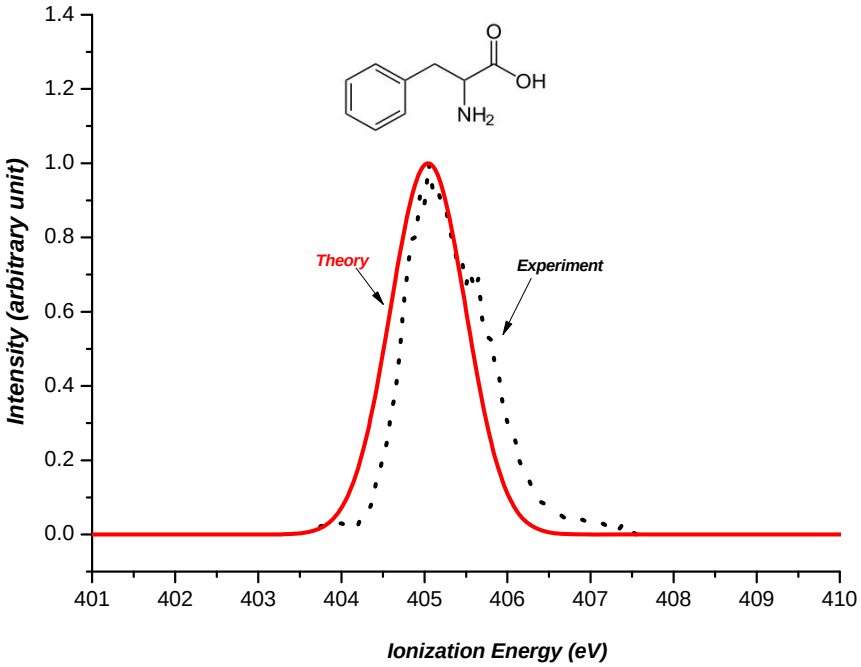


Figure 3(b) Phe O-K

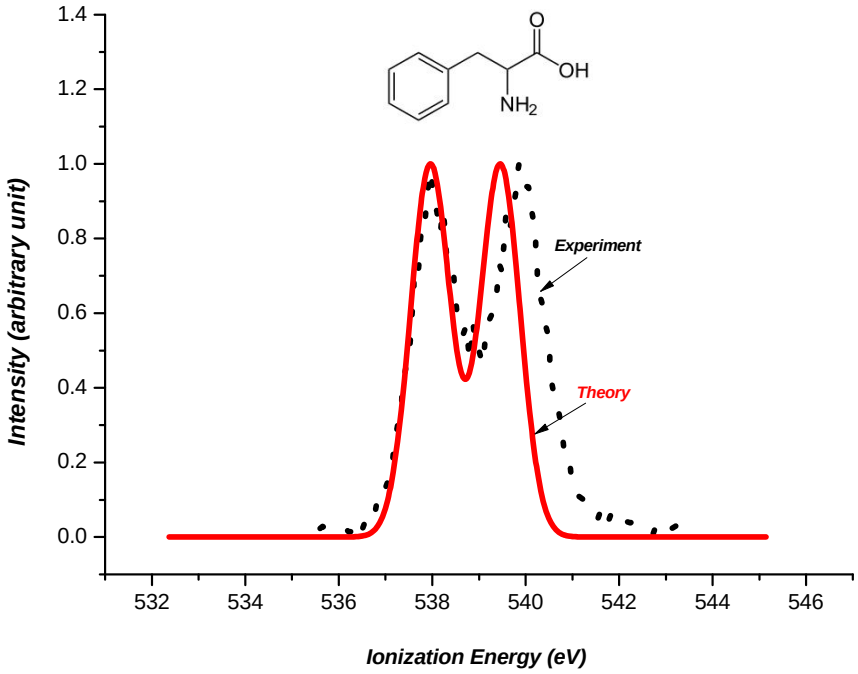


Figure 4

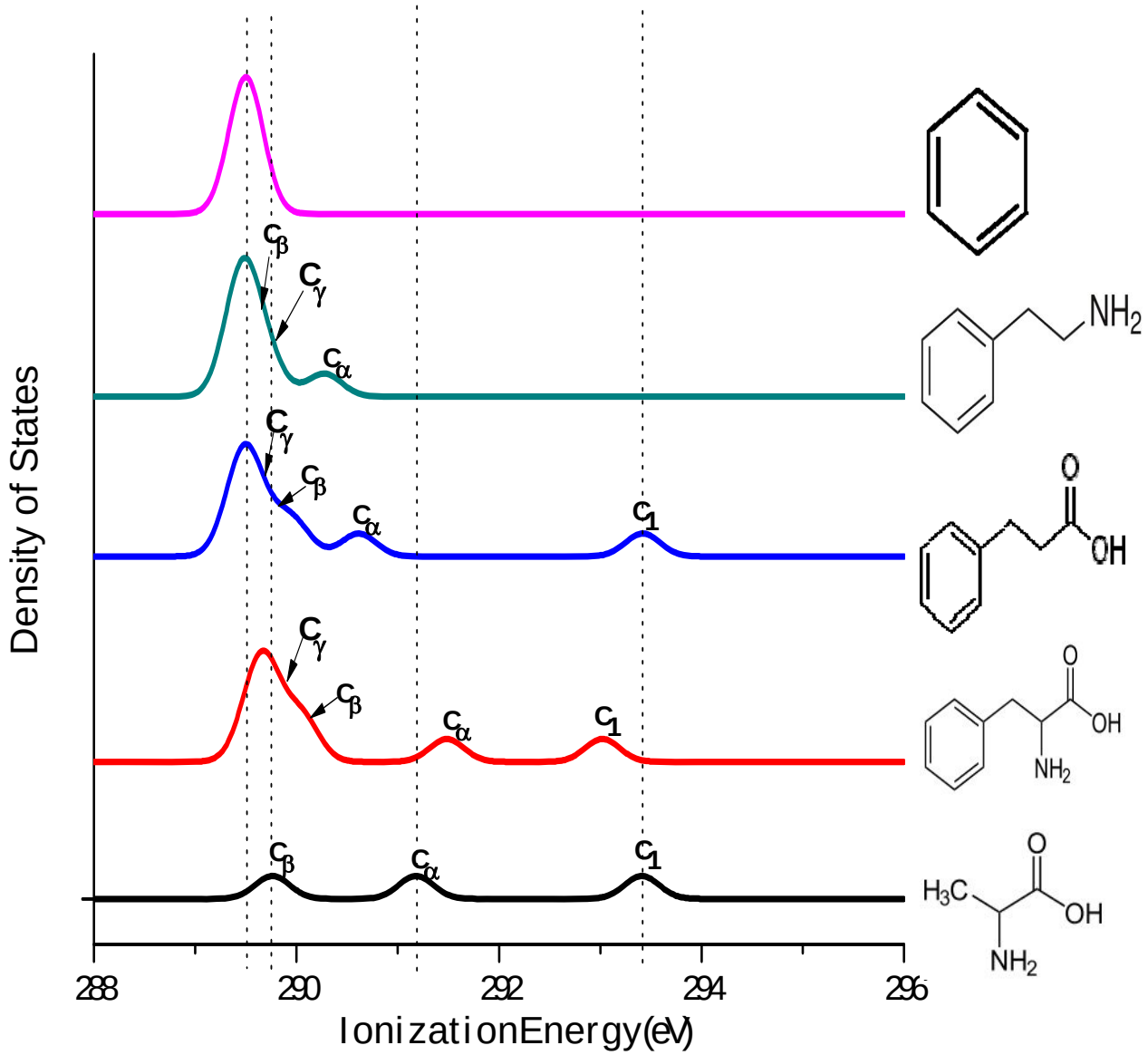
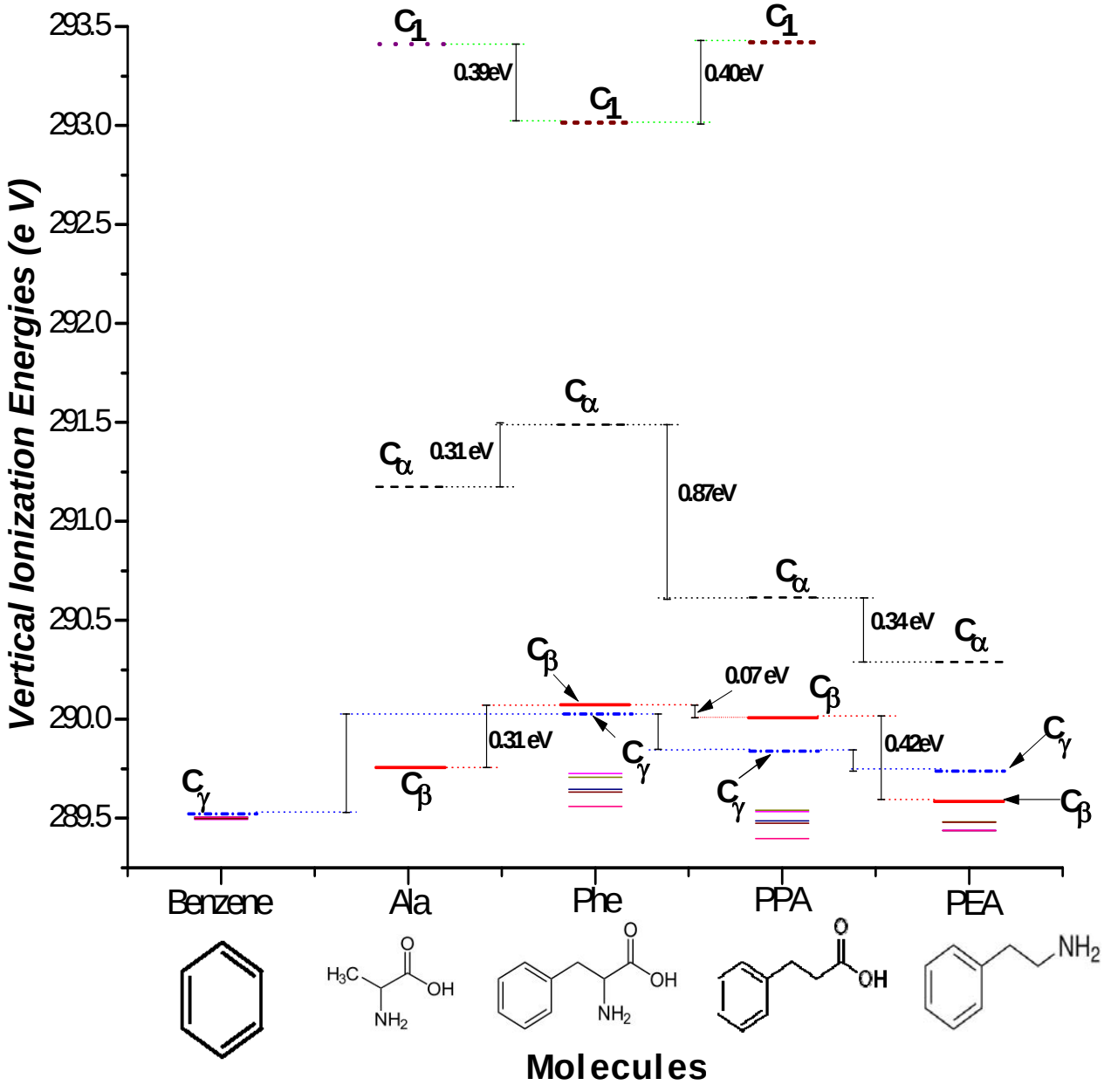

Density of States
288
290
292
294
296
Ionization Energy (eV)
C_β
C_γ
C_α
C_1
NH_2
O
OH
H_3C

Figure 5


Vertical Ionization Energies (e V)
293.5
293.0
292.5
292.0
291.5
291.0
290.5
290.0
289.5
C1
0.39eV
C1
C1
0.40eV
Cα
Cα
0.31 eV
0.87eV
Cα
0.34 eV
Cα
Cβ
0.07 eV
Cβ
0.31 eV
Cγ
0.42eV
Cγ
Cγ
Cγ
Cβ
Cβ
Benzene
Ala
Phe
PPA
PEA
Molecules