PAPER • **OPEN ACCESS**

A combined density functional theory and x-ray photoelectron spectroscopy study of the aromatic amino acids

To cite this article: Anna Regoutz *et al* 2020 *Electron. Struct.* **2** 044005

View the article online for updates and enhancements.

This content was downloaded from IP address 155.198.10.231 on 15/03/2021 at 11:58

Electronic Structure

OPEN ACCESS



RECEIVED
29 October 2020

REVISED
8 December 2020

ACCEPTED FOR PUBLICATION
23 December 2020

PUBLISHED
20 January 2021

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



PAPER

A combined density functional theory and x-ray photoelectron spectroscopy study of the aromatic amino acids

Anna Regoutz[1,*], **Marta S Wolinska**[2], **Nathalie K Fernando**[1] and **Laura E Ratcliff**[2,*]

[1] Department of Chemistry, University College London, 20 Gordon Street, London, WC1H 0AJ, United Kingdom
[2] Department of Materials, Imperial College London, London SW7 2AZ, United Kingdom
[*] Authors to whom any correspondence should be addressed.

E-mail: a.regoutz@ucl.ac.uk and laura.ratcliff08@imperial.ac.uk

Keywords: density functional theory, x-ray photoelectron spectroscopy, amino acids, core states, aromatic molecules

Supplementary material for this article is available online

Abstract

Amino acids are essential to all life. However, our understanding of some aspects of their intrinsic structure, molecular chemistry, and electronic structure is still limited. In particular the nature of amino acids in their crystalline form, often essential to biological and medical processes, faces a lack of knowledge both from experimental and theoretical approaches. An important experimental technique that has provided a multitude of crucial insights into the chemistry and electronic structure of materials is x-ray photoelectron spectroscopy. While the interpretation of spectra of simple bulk inorganic materials is often routine, interpreting core level spectra of complex molecular systems is complicated to impossible without the help of theory. We have previously demonstrated the ability of density functional theory to calculate binding energies of simple amino acids, using ΔSCF implemented in a systematic basis set for both gas phase (multiwavelets) and solid state (plane waves) calculations. In this study, we use the same approach to successfully predict and rationalise the experimental core level spectra of phenylalanine (Phe), tyrosine (Tyr), tryptophan (Trp), and histidine (His) and gain an in-depth understanding of their chemistry and electronic structure within the broader context of more than 20 related molecular systems. The insights gained from this study provide significant information on the nature of the aromatic amino acids and their conjugated side chains.

1. Introduction

Amino acids form the basis of peptides and proteins, which are fundamental building blocks of life, and they are of great scientific interest for a multitude of reasons, first and foremost due to their role in biology and related use in pharmacology and medicine. Their systematic nature also makes them perfect test systems to understand important aspects of the behaviour of molecular systems, including local and long-range structure and interactions, polymorphism, the three-dimensional arrangement of proteins, and ionic behaviour and its tunability by the environment. While the motivation to study amino acids is clear, experimental strategies are generally limited to structural techniques such as x-ray diffraction (XRD). A complementary technique, which can provide an additional level of information on chemical states and electronic structure not accessible to XRD, is x-ray photoelectron spectroscopy (XPS). Recently, we have started to explore the application of XPS to amino acids in their crystalline, powder form in combination with theoretical calculations based on density functional theory (DFT) [1, 2]. Our first study established a combined experiment-theory approach to predict and interpret primarily the C 1*s* core level spectra of the simple amino acids glycine (Gly), alanine (Ala), and serine (Ser) [3]. Here, we expand and improve our previous approach to amino acids with aromatic side chains, including phenylalanine (Phe), tyrosine (Tyr), tryptophan (Trp), and histidine (His). Figure 1 shows a schematic of their atomic structures, including alanine (Ala) which is used as a reference throughout and which we have reported previously [3].

© 2021 The Author(s). Published by IOP Publishing Ltd

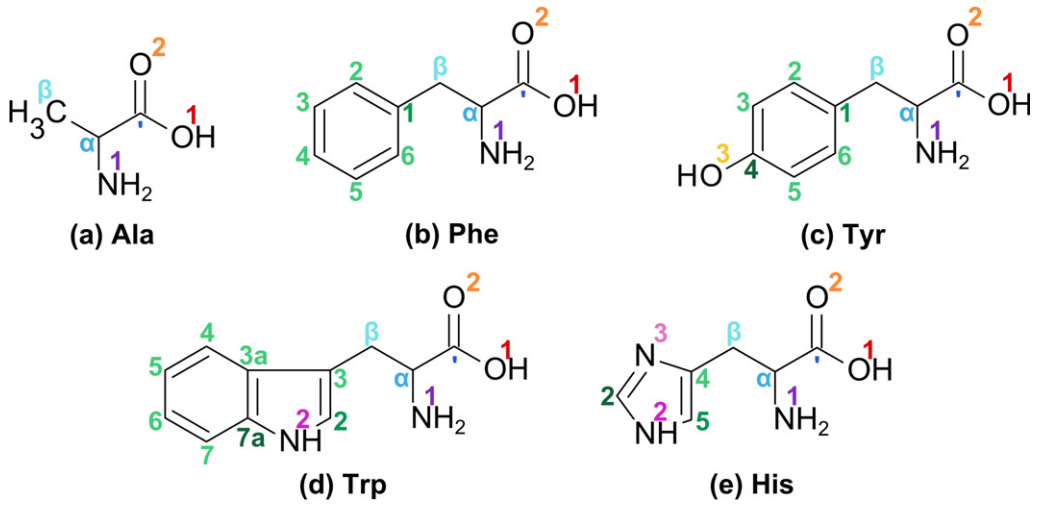


Figure 1. Schematic of (a) Ala, (b) Phe, (c) Tyr, (d) Trp and (e) His showing the atomic structures and atom labels, which will be used in the following.

As for XPS studies on amino acids in general, very few studies exist on the aromatic subgroup. A small number of experiments have been performed on Phe, Tyr and His adsorbed on single crystal substrates, including Au, Ag, Cu and TiO_2 [4–7]. While gas phase experiments are often used to study amino acids, this is difficult to achieve for the aromatic subgroup as they generally have high melting points (and consequently low vapour pressures) as well as low thermal stability [8]. A very limited number of studies on solid powders has been reported, which often suffer from low experimental resolution complicating peak assignments [9, 10]. The 2013 work by Stevens *et al* provides the most systematic and detailed study of solid phase amino acids to date, in which only His of the aromatic subgroup is included [11]. Beyond XPS, x-ray absorption and emission spectroscopy as well as electron energy loss spectroscopy have been employed to understand the chemistry and structure of the aromatic amino acids [12–14]. Due to the complexity of aromatic amino acids, the use of theory to guide the interpretation of spectra is essential. This is particularly true in the solid state, where intermolecular interactions can have an important effect, posing a further challenge to peak assignment. Nonetheless, from a theoretical point of view, only a handful of examples of core binding energy (BE) calculations of aromatic amino acids exist, which are all limited to the gas phase [8, 15–18]. Furthermore, to the best of our knowledge the core state BEs of His have not previously been calculated.

In this work, the aromatic amino acids Phe, Tyr, Trp and His are explored using both experiment and theory. The subgroup classification of amino acids usually includes Phe, Tyr and Trp in the aromatic group with Tyr also sometimes grouped with the polar amino acids. Due to the basic properties of His it is often classified as a polar amino acid. For completeness, we include all amino acids containing aromatic side chains here, independent of their polar nature. XPS experiments in the solid phase are compared to theoretical calculations based on DFT using the ΔSCF (self-consistent field) approach as implemented in systematic basis sets. Due to the complexity of the observed core level spectra and the apparent strong influence of not only nearest, but also next-nearest and even further removed neighbouring atoms, a molecular subspecies approach was followed to aid in the rationalisation and explanation of observed BE shifts, particularly for the C and N 1*s* core states. This is shown to be an extremely useful approach to gain a full and detailed understanding of the chemical and electronic structure of these important biological building blocks.

2. Methods

2.1. Theoretical approach

DFT was used to calculate the core state BEs of Ala, Phe, Tyr, Trp, and His. The primary motivation is the calculation of solid state BEs to aid the interpretation of experimental spectra. However, the solid state BEs are influenced by a combination of factors including the presence of different functional groups, the molecular and crystal structure, and intermolecular interactions. Theory is essential to disentangle these competing effects. Initial crystal structures for Ala, Phe, Tyr, Trp and His were obtained from references [19–23], respectively. In order to assess the influence of the molecular structure, different gas phase conformers were tested for each molecule and selected as follows. Four low energy conformers for Phe, Tyr and Trp, and three low energy conformers for His were taken from the literature [24, 25]. Following geometry optimisation and BE calculations,

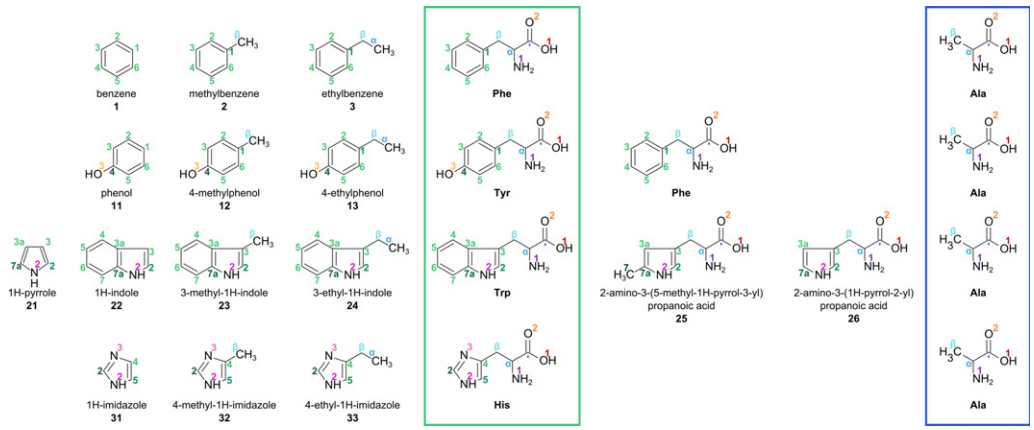


Figure 2. Molecular structure of the amino acids Ala, Phe, Tyr, Trp and His, as well as the molecular subspecies calculated, including atom labels used throughout this work. The molecules included are (**1**) benzene, (**2**) methylbenzene, (**3**) ethylbenzene, (**11**) phenol, (**12**) 4-methylphenol, (**13**) 4-ethylphenol, (**21**) 1H-pyrrole, (**22**) 1H-indole, (**23**) 3-methyl-1H-indole, (**24**) 3-ethyl-1H-indole, (**25**) 2-amino-3-(5-methyl-1H-pyrrol-3-yl)propanoic acid, (**26**) 2-amino-3-(1H-pyrrol-2-yl)propanoic acid, (**31**) 1H-imidazole, (**32**) 4-methyl-1H-imidazole, and (**33**) 4-ethyl-1H-imidazole.

the two conformers of each amino acid with the most distinct BEs were retained for further investigation. For Ala, only the lowest energy conformer used in reference [3] was considered. The structures of each conformer are presented in the supplementary information (https://stacks.iop.org/EST/2/044005/mmedia). The gas phase conformers were compared to the molecule extracted directly from the bulk, which was allowed to relax away from the zwitterionic state into its neutral form. BEs were calculated using the ΔSCF approach, however, in order to distinguish between initial and final state effects, gas phase BEs were additionally calculated at the level of Koopmans'. To determine the contribution from intermolecular interactions, the BEs of both the gas and solid phase are compared. Finally, in order to assess the impact of different functional groups, a systematic series of subspecies molecules was investigated, as depicted in figure 2, which are derived from the aromatic amino acids studied here.

In order to aid interpretation of experimental spectra, the relative BE positions of contributing chemical environments are needed. Absolute BEs are not necessary for this approach, and DFT is more reliable for relative than absolute BEs. However, some recent work exists where DFT has been shown to accurately reproduce absolute BEs [26, 27]. When comparing calculations across the molecules, it is important to note that while BEs calculated for molecules in the gas phase can be directly compared between molecules, this is not the case for solid state calculations, since the core hole calculations are performed in charged supercells. Although schemes exist to account for the use of periodic boundary conditions (e.g. reference [27]), this can introduce an additional source of uncertainty. Therefore, since it is not essential for the current work, BEs of the amino acids in the solid state are not directly compared.

2.1.1. Computational details

Gas phase geometry optimisations were performed using BigDFT [28, 29], in open boundary conditions, with a wavelet grid spacing of 0.185 Å, coarse and fine radius multipliers of 5 and 8, respectively, and HGH-GTH pseudopotentials (PSPs) [30, 31]. Gas phase BE calculations were performed at the level of both Koopmans' and ΔSCF using the MADNESS molecular DFT code [32] with open boundary conditions. A mixed all-electron (AE)/PSP approach was used [33], wherein the atom of interest was treated at the AE level, with remaining atoms treated at the PSP level, as described in reference [3]. Ground state calculations used a wavelet threshold of 10^{-4} followed by 10^{-6} (wavelet order $k = 6$ and $k = 8$), while core hole calculations directly used a wavelet threshold of 10^{-6} ($k = 8$). A convergence criterion of 10^{-3} was used for both the density and Kohn–Sham wavefunction residuals. Following reference [3] the ground state wavefunctions were used as an input guess for the core hole calculations, localisation was imposed on the wavefunctions for the ground state while core hole calculations used canonical orbitals, and the B-spline projection based derivative operator was used (except for the calculation of the kinetic energy operator) [34]. Calculations employed the same PSPs as BigDFT.

Solid state geometry optimisations and BE calculations using the ΔSCF approach were performed with the CASTEP plane-wave DFT code [35]. Core hole PSPs were used to represent the core-excited atom, following the same procedure and with the same norm-conserving on-the-fly generated PSPs as reference [3]. Calculations were performed with a cut-off energy of 900 eV, and Monkhorst–Pack [36] k-point grids of $2 \times 1 \times 2$, $2 \times 2 \times 1$, $2 \times 1 \times 2$, and $2 \times 2 \times 2$ for Ala, Phe, Tyr and His, respectively, with Trp calculations

Table 1. Experimental ('Exp') and PBE/PBE0 calculated relative C, O and N 1*s* core state BEs for solid state amino acids.

	Ala			Phe		Tyr			Trp		His		
	PBE	PBE0	Exp	PBE	Exp	PBE	PBE0	Exp	PBE	Exp	PBE	PBE0	Exp
C′	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C^{α}	−1.7	−2.1	−2.0	−1.4	−1.9	−1.5	−2.1	−2.5	−1.3	−2.1	−1.8	−2.1	−1.9
C^{β}	−3.1	−3.5	−3.3	−2.5	−2.6	−2.7	−3.4	−3.6	−2.3	−3.1	−2.7	−3.2	−3.2
C_1	—	—	—	−2.9	−3.1	−3.6	−4.3	−4.3	—	—	—	—	—
C_2	—	—	—	−3.2	−3.7	−3.7	−4.4	−4.3	−2.7	−3.7	−2.1	−2.5	−2.4
C_3	—	—	—	−3.1	−3.7	−3.8	−4.5	−4.3	−3.3	−4.3	—	—	—
C_{3a}	—	—	—	—	—	—	—	—	−3.2	−4.3	—	—	—
C_4	—	—	—	−3.2	−3.7	−2.1	−2.8	−3.1	−3.4	−4.3	−2.8	−3.2	−3.2
C_5	—	—	—	−3.2	−3.7	−3.8	−4.5	−4.3	−3.4	−4.3	−2.8	−3.3	−3.2
C_6	—	—	—	−3.1	−3.7	−3.6	−4.3	−4.3	−3.3	−4.3	—	—	—
C_7	—	—	—	—	—	—	—	—	−3.2	−4.3	—	—	—
C_{7a}	—	—	—	—	—	—	—	—	−2.3	−3.1	—	—	—
O^1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
O^2	0.2	0.2	0.0	0.0	0.0	0.1	0.1	0.0	−0.1	0.0	0.0	0.0	0.0
O^3	—	—	—	—	—	0.8	0.8	1.2	—	—	—	—	
N^1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
N^2	—	—	—	—	—	—	—	—	−1.2	−1.5	−0.6	−0.7	−1.0
N^3	—	—	—	—	—	—	—	—	—	—	−2.2	−2.3	−2.7

performed at the Γ-point only. Geometry optimisations used the semi-empirical dispersion correction scheme of Grimme [37].

Gas phase BE calculations were performed using the semi-local Perdew–Burke–Ernzerhof (PBE) exchange correlation functional only [38], while solid state BE calculations were performed using both PBE and the hybrid PBE0 functional [39], except for Phe and Trp where PBE0 calculations were prohibitively expensive due to their large unit cells containing 184 and 432 atoms, respectively. All BEs were calculated in the vertical approximation.

All geometry optimisations used the PBE functional and a force tolerance of 0.02 eV $Å^{-1}$. For solid state geometry optimisations the cell was also allowed to relax. For molecules extracted from the optimised crystals, only the H atoms were relaxed, with all other atoms frozen. In order to prevent collapse back to the zwitterionic state, an initial perturbation was applied to one of the H atoms. For all other gas phase calculations, all atoms were allowed to relax. All calculations were spin restricted and relativistic effects were neglected, since although these can have a significant effect when calculating absolute BEs (see e.g. reference [40]), they are less significant when considering relative BEs, as in this work. The same computational parameters were used for both gas phase amino acids and molecular subspecies calculations. Molecule and crystal structures were visualized using VESTA [41].

2.2. Experimental approach

Powders of the L-stereoisomers of all investigated amino acids were purchased from Sigma-Aldrich (Ala ⩾ 99%, Phe ⩾ 98%, Tyr ⩾ 98%, Trp ⩾ 98%, His ⩾ 99%). Core level spectra were recorded on a Thermo Scientific K-Alpha+ XPS system with a monochromated, microfocused Al Kα x-ray source ($h\nu = 1486.7$ eV), which was operated a 6 mA emission current and 12 kV anode bias. The base pressure was 2×10^{-9} mbar. All core level spectra were collected at a pass energy of 20 eV using an x-ray spot size of 400 μm. Samples were mounted on conducting carbon tape and a flood gun was employed to prevent sample charging. As amino acids are prone to suffer from radiation damage, samples were rastered and data collected at four points across the samples, which were then averaged to achieve the necessary signal statistics for peak fitting. All data were analysed using the Avantage software package. Differences in peak positions across the different measurement points were less than 50 meV for all core levels. For peak fit analysis, Shirley-type backgrounds and Voigt functions were used with both the full width at half maximum (FWHM) and Lorentzian/Gaussian (L/G) ratios refined.

3. Results

3.1. Calculated solid state binding energies

In line with our previous work [3], the use of PBE with semi-empirical dispersion corrections for the solid state geometry optimisations resulted in a good description of the crystal structure. Relaxed lattice parameters and angles, which are reported in the supplementary information alongside the relaxed crystal structures, are in good agreement with the experimental values, with maximum discrepancies of 3.0% and 1.2%, respectively.

Calculated BEs for the solid state amino acids are presented in table 1. Due to the relatively large unit cell sizes of the aromatic amino acids, it is highly desirable to perform calculations using semi-local functionals such as PBE, rather than hybrid functionals such as PBE0. Indeed, for Phe and Trp PBE0 BE calculations were prohibitively expensive. For the amino acids where PBE0 calculations were possible (Ala, Tyr, and His), significant quantitative differences can be seen between the two functionals for C 1*s*, up to 0.7 eV in the most severe cases. However, qualitatively the differences are less significant, and it is primarily the BE of C′ relative to the other states which is most strongly affected. Importantly, the order of BEs remains constant to within 0.1 eV, so that for the purposes of aiding in peak assignment in experimental spectra it is not necessary to go beyond PBE. Furthermore, the differences for O and N 1*s* core states are negligible. Therefore, the calculations presented in the following sections were performed using PBE only.

While the calculated BE positions describe the experimental core level spectra very well, which will be discussed in detail in section 3.3, it is not easy to intuitively rationalise the order and relative positions of the different constituents, in particular for the case of C 1*s* with its many chemical states. Therefore, a molecular subspecies approach was chosen to systematically explore core level energy changes with the removal or introduction of part of the amino acids and their functional groups.

3.2. Molecular subspecies series

Twenty-two additional small molecular systems were explored theoretically to aid our understanding of the core level spectra observed for the aromatic amino acids. Figure 2 gives an overview of the main set of molecular subspecies calculated and their relationship to the aromatic amino acids and Ala. Figure 3 provides an overview of the C 1*s* BEs of the molecular series, while the tables of the corresponding BEs are also given in the supplementary information. A set of additional subspecies was explored to understand specific questions arising around nitrogen groups and aromatic systems, which is shown in the supplementary information. In the following subsections the results and main conclusions for each of the aromatic amino acids are discussed.

3.2.1. Phe

In parallel to Phe being the simplest of the amino acids explored here, it also reduces to the simplest submolecule, benzene (**1**). As expected, all C atoms for benzene have the same BEs as each other in both the Koopmans' and the ΔSCF approaches. Moving to methylbenzene (**2**) a clear difference between C_1 and the remaining C atoms of the aromatic ring is noticeable, in line with previous calculations [42]. This is clearly illustrated by the differences between the ground state electronic densities of (**2**) and (**1**), which are depicted in the supplementary information, where the addition of the CH_3 group changes the density around C_1. There are also non-negligible changes in the density around all other aromatic C atoms C_{arom}. Combined with changes in the atomic structure between (**2**) and (**1**), and which are not accounted for in the visualisation of the densities, this explains why the Koopmans' BEs of all C atoms change between the two molecules. Comparing to experimental gas phase measurements by Ohta *et al* [43], we observe that while the relative BEs agree reasonably well with experiment, their peak assignments are more in line with the Koopmans' values. In particular, C_1 has the highest BE, while C^{β} is at the lowest BE.

While C^{β} in (**2**) and ethylbenzene (**3**) and C^{α} in (**3**) occur at the lowest BEs, this changes completely using the ΔSCF approach, where C^{α} and C^{β} move to the higher BE side of all other C atoms. In addition, a clear chemical shift between the CH_2 and CH_3 groups of the side chain for (**3**) is also apparent. In order to understand to what extent the shift in C^{β} for ΔSCF is affected by the aromaticity of (**2**), we also compare with methylcyclohexane (**44**), for which results are given in the supplementary information. In particular, the ΔSCF results for (**44**) only show a small spread, but otherwise both C^{β} and C_1 are at very similar energies to the remaining C atoms, in contrast to (**2**). In other words, the conjugated system is much more sensitive to the addition of the CH_3 group when final state effects are taken into account. When examining the density difference between (**3**) and (**2**), there is a small change in the density around C_1, which in turn gives rise to a small change in the Koopmans' BEs. All other C atoms in the ring, however, remain unaffected by the addition of the CH_3 group, so that the corresponding BEs of the C_{arom} atoms do not change between (**2**) and (**3**).

For Phe itself the C_{arom} atoms including C_1 behave similarly to (**1**)–(**3**), with a clear spreading in BE of C_1–C_6. With the addition of the carboxylic COO^- group the separation between C_{arom} and C^{α} and C^{β} increases significantly and C^{α} and C^{β} switch places. The ΔSCF results for the gas phase molecules show similar variations between conformers and are in agreement with previous calculations from Zhang *et al* [8] who included four different conformers. When comparing the ΔSCF gas phase conformers with the solid Phe a clear bunching up of BEs is observed, while the relative BE order of the different C environments remains the same. The significant change in BE of C^{α} and C′ can be explained by the change from $COOH/NH_2$ to the zwitterionic COO^-/NH_3^+ environments and the resulting intermolecular interactions. The main observable difference between the Koopmans' and ΔSCF results for Phe lies in the differentiation of C^{β} from C_{arom}. While they are

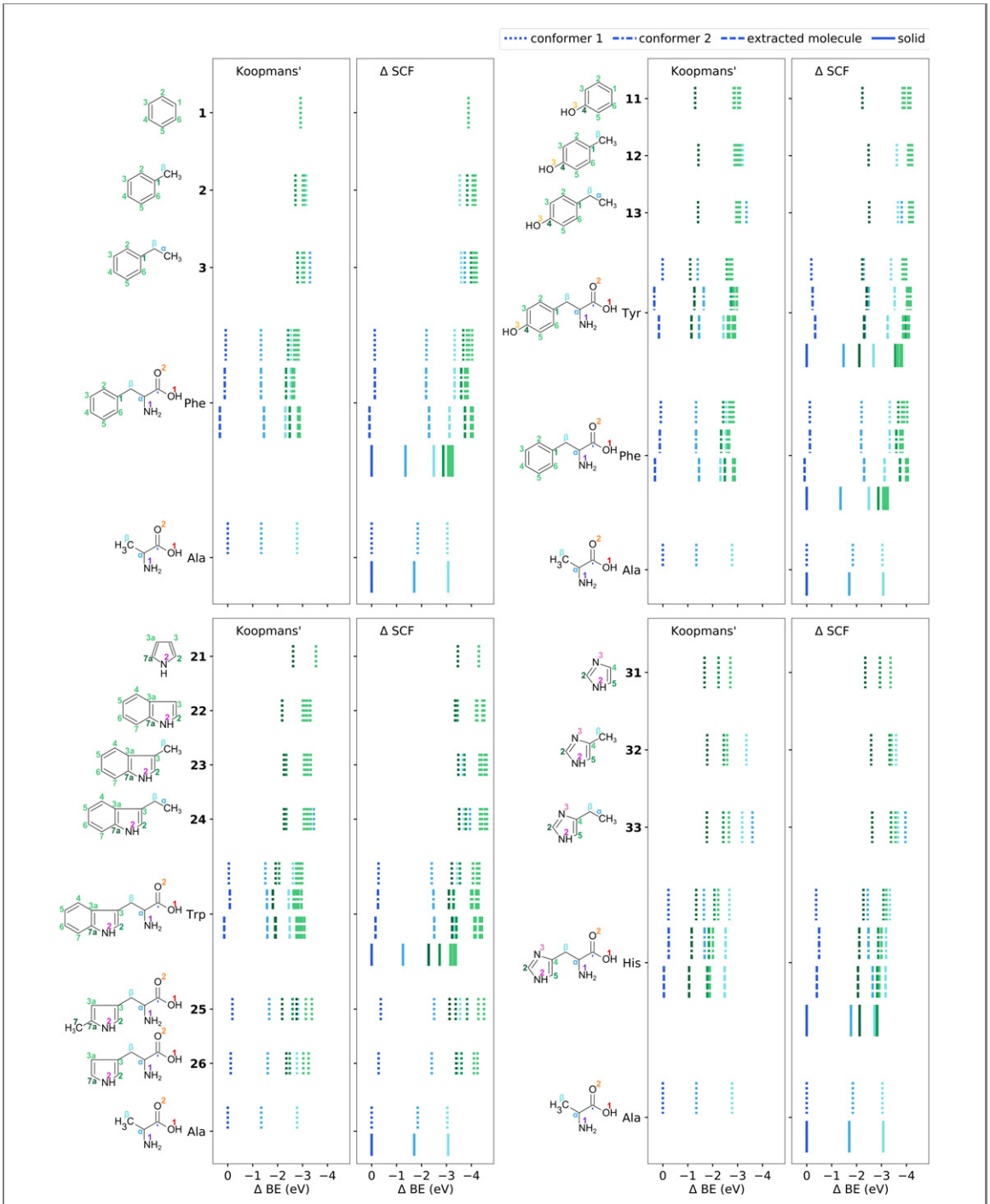


Figure 3. PBE-calculated C 1*s* BEs for the amino acids and the series of subspecies molecules. Gas phase BEs are relative to Ala C′, while solid state BEs are relative to C′ of that amino acid.

very close in energy or even overlap for some conformers, C^{β} moves to significantly higher BEs in ΔSCF due to final state effects, which is consistent with the behaviour of C^{β} in (**2**) and (**3**).

3.2.2. Tyr

Across the series from phenol (**11**) to 4-methylphenol (**12**) and 4-ethylphenol (**13**) a common feature is the spreading out of C_{arom} BEs due to the presence of the hydroxyl group. Similarly to the equivalent series for Phe, there is a significant change in the electronic density (shown in the supplementary information) on all C_{arom} going from (**11**) to (**12**), with corresponding changes in the BEs. However, the changes in density between (**12**) and (**13**) are again primarily localized on C_1, with the remaining C_{arom} unaffected by the addition of the CH_3 group. In parallel to the spreading out of the C_{arom} BEs, a large gap also opens up between C_4 and the remaining C atoms. Comparing (**11**) with cyclohexanol (**47**), for which results are presented in the supporting information, this gap is much larger in (**11**) than in (**47**) for both Koopmans' and ΔSCF, demonstrating the strong influence of the aromaticity and the importance of final state effects. Both Koopmans' and ΔSCF results

for (**11**) agree well with experimental gas phase results from Ohta *et al* [43]. One interesting point to note about (**11**) is the difference in BEs between C_3 and C_5, and C_2 and C_6, which in contrast have the same BEs in both (**2**) and aniline (**46**) and the same is true for the equivalent non-conjugated molecules. However, due to the presence of the hydroxyl group, neither (**11**) nor (**47**) have symmetric structures, and the small asymmetry of the BEs can be attributed to this asymmetry of the atomic structures.

As was the case for the subspecies molecules for Phe, C^{α} and C^{β} in (**12**) and (**13**) occur at the lowest BEs in the Koopmans' approach, but swap when ΔSCF is used. Compared to Phe, C^{β} in Tyr shifts to even higher BE relative to C_{arom} in the ΔSCF approach. This is a direct result of the addition of the hydroxyl group onto the aromatic ring and showcases the strong long-range intramolecular interactions taking place. Of course C_4 is now also clearly separated from the rest of the aromatic ring and located at a BE intermediate between C^{α} and C^{β}.

As with Phe, there is also variation between Tyr conformers, where the results are again in line with calculations from Zhang *et al* [8]. A significant change in the BE separation of C^{α} and C_4 occurs when moving from the gas phase calculations to the solid state case. While in the gas phase their binding energies are almost identical across all Tyr molecules considered, they separate significantly in the solid. This is due to the hydroxyl group taking part in intermolecular hydrogen bonding as can be clearly seen from the crystal structures shown in the supplementary information.

3.2.3. Trp

1H-pyrrole (**21**) nicely exemplifies the symmetric nature of the ring with C_2/C_{7a} and C_3/C_{3a} grouping together for both Koopmans' and ΔSCF. In 1H-indole (**22**) C_2 and C_{7a} remain at significantly higher BEs than all other C atoms. Comparing the Koopmans' and ΔSCF results for 3-methyl-1H-indole (**23**) and 3-ethyl-1H-indole (**24**) a considerable change in BE for C^{α} and C^{β} is observed as in the previous cases discussed. A systematic difference in the ΔSCF BEs of C_2 and C_{7a} is noted across all molecules in the series except (**21**), even if the six-membered ring is removed as is the case in 2-amino-3-(5-methyl-1H-pyrrol-3-yl)propanoic acid (**25**) and 2-amino-3-(1H-pyrrol-2-yl)propanoic acid (**26**).

As with Phe and Tyr, and again in agreement with Zhang *et al* [8], there is noticeable variation between the Trp conformers. While the Koopmans' BEs for Trp are in line with chemical intuition, the ΔSCF BEs are harder to explain. In particular, contrary to the expectation that aromatic and aliphatic C atoms should have similar BEs, C^{β} is noticeably higher in BE than the C_{arom} which do not neighbour an N atom. Indeed, the BE of C^{β} is especially sensitive to final state effects, as evidenced by the difference between Koopmans' and ΔSCF values. This is also the case for both Phe and Tyr, and by comparing (**2**) and (**44**) was attributed to the conjugated nature of the ring. Similarly, the density comparisons discussed in relation to Phe and Tyr demonstrated that the functionalisation of an aromatic ring can impact on the density and thus the BEs of *all* atoms in the ring, not just the nearest neighbour. This explains for example why it is not just the BE of C^{α} which is affected by the addition of the amino group when going from (**24**) to Trp.

Furthermore, in Trp the BE of C^{β} is surprisingly close to that of both C_2 and C_{7a} in the gas phase and the same as C_{7a} in the solid state, which cannot be explained by arguments based purely on electronegativity. On the contrary, since they each neighbour an N atom, one would expect the BE of C^{α} to be close to that of C_2 and C_{7a}, which is not the case in either Trp, (**25**), or (**26**). In addition to next-nearest neighbour effects, this can also be explained by the protonation state of the N atoms. In order to provide further insights on the influence of different protonation states of N on C 1*s* BEs, we also considered an additional set of subspecies molecules containing nitrogen, for which results are given in the supplementary information. Taking for example the series of ethylamine (**41**) to diethylamine (**42**) to triethylamine (**43**), one can see a clear trend in the ΔSCF BEs, where the higher the protonation state of the N atom, the higher the BE of C^{α}. This trend is in agreement with C^{α} having a higher BE than C_2 and C_{7a}. Finally, we note that the BEs of C^{β} in the alkylamine series are also affected by the change in N protonation state, providing further support for the importance of next-nearest neighbour effects, although the magnitude of variations is much smaller than for C^{α}.

To further test the influence of aromaticity, the BEs for (**46**) and cyclohexanamine (**45**) were calculated, for which results are given in the supplementary information. Both Koopmans' and ΔSCF results for (**46**) are in good agreement with experimental gas phase results from Ohta *et al* [43]. Consistent with (**11**) and (**47**), a larger gap between C_1 and the remaining C atoms is observed for (**46**) than (**45**), while there is also a larger spread of the C atoms in the ring in (**46**) compared to (**45**). A clear overall trend is observed upon the addition of a functional group to a ring, whether conjugated or non-conjugated, where an increasing split between the C atom the group binds to and the remaining C atoms is observed in line with the increasing electronegativity in going from C to N to O in the functional groups CH_3, NH_2, and OH. Comparing the effect on the conjugated versus non-conjugated rings, this difference is always bigger for the conjugated ring.

3.2.4. His

In 1H-imidazole (**31**) the three C atoms all have considerably different BEs, including a clear distinction in C BE depending on the protonation of the neighbouring N atom in line with the previous observations for molecules (**41**)–(**43**). The addition of the methyl and ethyl side chains in 4-methyl-1H-imidazole (**32**) and 4-ethyl-1H-imidazole (**33**), respectively, reduces the difference in BE between C_4 and C_5. Going from Koopmans' to ΔSCF a significant change in the BEs of C^{β} and C^{α} relative to the three C atoms in the aromatic ring, C_2, C_4, and C_5, is observed. The relative differences between C_2, C_4, and C_5 remain very similar between the two approaches.

Across all ΔSCF gas phase calculations of His, C_4, C_5 and C^{β} are very close in BE. This is comparable to the observations made for C_2, C_{7a} and C^{β} in Trp. Another similarity between Trp and His is that C^{α} is the most sensitive to changes in conformer and gas/solid phases, and its BE changes significantly between calculations. In the solid phase C^{α} is even higher in BE than C_2, which is not the case in either the Koopmans' or ΔSCF gas phase calculations, and this is most certainly not immediately intuitive. However, based on the results presented so far, this is a consequence of a complex interplay between the protonation of the N atoms, the influence of the aromatic ring, and the intermolecular interactions of both the NH_3^+ and NH groups in His. As will be discussed in more detail in the following section, previous experimental work by Stevens *et al* assigned the chemical states present closely to the results we find for the Koopmans' approach [11].

To summarise the observations made to this point, the molecular subspecies approach is invaluable to rationalise and discuss the complex relative BE changes observed in the amino acids. A fascinating, if somewhat subjective, result from the combination of experiment and theory and the exploration of the molecular subspecies is that chemical intuition and the experience of a spectroscopist usually reflects the results given by Koopmans' theorem. The additional rearrangement of BE positions observed in ΔSCF is often surprising, resulting in our hypothesis that human brains are not best placed to compute final state effects ad hoc without the aid of DFT.

3.3. Core level spectra of the amino acids

Where experimental core level spectra exist in the literature, they are very similar to the data presented here, albeit often with lower energy resolution [8, 9, 11, 44]. The main difference is often found in the peak fits, including the number of peaks fitted and their relative BEs and intensities. The peak fits presented here are based on robust, physically justifiable line shapes, including FWHM and L/G ratio, with the number of peaks informed from theory where needed due to overlap. In figure 4 a Shirley-type background has been subtracted to aid comparison with theory, while the relative BEs are presented in table 1 alongside the calculated values. Absolute BEs resulting from the peak fits are given in the supplementary information. It should be noted that adventitious carbon at around 285 eV is present in all samples as is expected for XPS of ex-situ prepared powders, leading to a slight deviation from expected relative intensities.

3.3.1. C 1s

Considering first the C 1*s* BEs, it is clear that PBE0-calculated values agree more closely with experiment. The main discrepancy is that for PBE calculations the BE of C′ is much closer to C^{α} than for PBE0. In the worst case, Tyr, the difference between C′ and C^{α} is 1 eV smaller than for the experimental BEs, while the difference for PBE0-calculated BEs is much closer to experiment. This is clearly evident in figure 4. However, as previously discussed, the relative BEs of all C atoms other than C′ are in very similar positions relative to each other for both PBE and PBE0. As a result, where the calculated BEs are aligned with respect to the lowest BE peak as in figure 4, the only visible difference between PBE and PBE0 is in the position of C′. This is reflected in the mean absolute error (MAE) of the BEs between experiment and theory—taking C′ as a reference the MAE is at 0.2 eV or less for PBE0, while in the worst case for PBE, Trp, this is much higher at 0.9 eV. If, however, the BEs are aligned with respect to the lowest BE peak, the PBE MAEs are similar to PBE0.

His is the only amino acid included here for which high resolution solid state spectra have previously been reported [11]. As the work by Stevens *et al* includes detailed information on the peak fits and resulting peak positions, this can be directly compared with the present results. The peak assignments made in the Stevens work agree well with what the Koopmans' level of theory predicts for gas phase His, with C′, C^{α} and C^{β} in good agreement with the present results. The main difference lies in the assignment of the subpeaks of the aromatic C atoms C_2, C_4 and C_5. C_4 and C_5 are assigned an intermediate BE between C^{β} and C^{α} in the Stevens work, but based on the solid state ΔSCF theory results presented here, it is clear that both overlap with C^{β}. And while C_2 is assigned the second highest BE in the previous work, it becomes clear that it actually lies below C^{α}. The peak fits presented in figure 4 take into account the theoretical results and a good agreement between the two is found.

In addition to the main photoionisation features all C 1*s* spectra include $\pi - \pi^*$ shake-up satellites at 6–7 eV above the main photoionisation peak at lowest BE with relative intensities of $\leqslant$3% compared to the aromatic

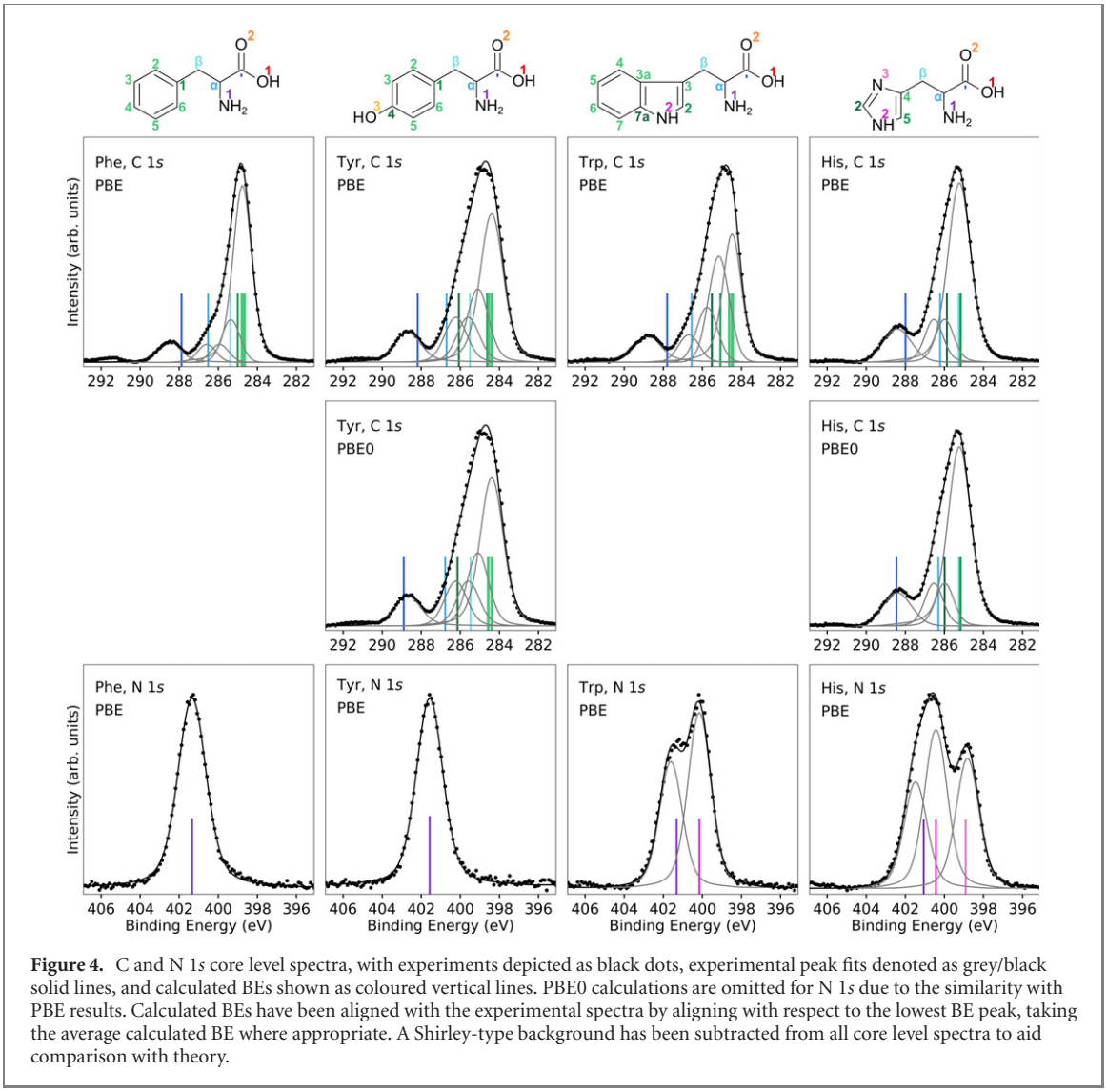


Figure 4. C and N 1*s* core level spectra, with experiments depicted as black dots, experimental peak fits denoted as grey/black solid lines, and calculated BEs shown as coloured vertical lines. PBE0 calculations are omitted for N 1*s* due to the similarity with PBE results. Calculated BEs have been aligned with the experimental spectra by aligning with respect to the lowest BE peak, taking the average calculated BE where appropriate. A Shirley-type background has been subtracted from all core level spectra to aid comparison with theory.

contribution of the C 1*s* core level. This is in good agreement with observations made for many conjugated systems, including early studies of Phe, Tyr and Trp by Clark *et al* [44]. The calculation of satellite features is challenging and they are not included in the theoretical calculations presented here, although we note that approaches based on both DFT and time-dependent DFT have been successfully employed for large molecules [45, 46].

3.3.2. N 1s

In contrast to C 1*s*, where a considerable difference in PBE vs PBE0-calculated BE values is observed, the N 1*s* BEs are not strongly affected by the functional. The calculated BEs are closer together than the experimental BEs, however the MAEs are in line with those for C 1*s*. Only Trp and His have more than one N atom and therefore only these two will be discussed in detail in this section. The BEs for the molecular subspecies series as well as gas phase amino acids are given in the supplementary information.

For Trp, a big change in the difference between the BEs for N^1 and N^2 is observed when going from Koopmans' to ΔSCF for the gas phase calculations, but in both cases N^2 is at a higher BE, in agreement with calculations from Zhang *et al* [8]. The order of the calculated BEs in (**25**) and (**26**) is also consistent with gas phase Trp. The calculations by Zhang *et al* also show a strong variation between conformers, particularly for N^1 which varies by up to 0.7 eV, which they attribute to differences in the nature of the internal hydrogen bonding present in a given conformer. In the solid phase the BE order of N^1 and N^2 flips compared to the gas phase, which is attributed to the presence of the zwitterion state in the solid phase and the resulting intermolecular interactions. To understand the differences in the BEs observed for N atoms with varying protonation further, subspecies molecules (**41**)–(**43**) were calculated. In the Koopmans' approach the BEs are in the order $N^3 > N^2 > N^1$, while this is reversed in ΔSCF. Both the ordering and values from the ΔSCF approach agree very well with gas phase measurements from Cavell and Allison [47]. Therefore, the observed flipping of N^1

and N^2 is most likely not solely caused by intermolecular interactions but also originates from intrinsic final state effects. The molecules (**45**) and (**46**) once again reinforce the observed influence of aromatic systems on the BEs. In particular, the aromatic aniline (**46**) molecule is more affected by ΔSCF, with a relative change of 0.3 eV compared to Koopmans'. Furthermore, there is a large difference between the BEs of (**45**) and (**46**) −0.7 eV for Koopmans' and 0.9 eV for ΔSCF, where again the ΔSCF results are in good agreement with the difference of 0.6 eV measured by Cavell and Allison.

The calculated N 1*s* BEs for Trp agree well with the experimentally observed values. In the experimental N 1*s* spectrum of Trp a higher intensity of the peak assigned to N^2 relative to N^1 is observed. This deviation from the 1:1 ratio of the two N components has been reported previously [44], and is most likely caused by a partial deprotonation of the NH_3^+ group at the surface of the powder sample.

The N 1*s* BEs of His show a similar sensitivity to a range of factors as for Trp. Looking at the gas phase conformers, N^2 shows a consistently higher BE for (**31**)–(**33**) and all His conformers, for both Koopmans' and ΔSCF results. However, the ordering of N^1 and N^3 changes between different conformers. Quantitatively, the BEs also vary significantly between Koopmans' and ΔSCF, with N^3 typically being affected most strongly, although there do not appear to be any general trends. This again highlights the importance of taking final state effects into account. Furthermore, the trend in energies cannot be explained purely by considering protonation states, but is likely influenced by both aromaticity and interactions between the two N atoms in the ring. As with Trp, the solid state BEs are qualitatively different from the gas phase conformers, with N^1 now having the highest BE and N^3 having the lowest BE. The fact that N^1 has the highest BE agrees with the behaviour in Trp, and as for Trp the change between gas and solid state BEs is likely due to a combination of the zwitterionic nature of the amino acid in the solid state as well as the related intermolecular interactions.

Two previous experimental studies have reported N 1*s* spectra for His. Feyer *et al* show N 1*s* core level spectra comparable to those reported here, but are not able to resolve N^1 and N^2 in their analysis [48]. Stevens *et al* report BE values of 398.8 eV (N^3), 400.4 eV (N^2), and 401.4 eV (N^1) for His, which are in good agreement with our measurements and peak assignments, and both agree well with the calculated values.

3.3.3. O 1s

To complete the set of core states present in the aromatic amino acids, the O 1*s* spectra are presented in the supplementary information. As with N 1*s* BEs, the calculated values are not affected by the functional, and the MAE between theory and experiment is also in line with N 1*s*. However, overall, these spectra do not provide much additional information beyond what has been discussed based on the C and N 1*s* results and only Tyr has more than one oxygen environment present in the solid state. In addition, O 1*s* has an intrinsically high lifetime width and small magnitude of chemical shifts, which in combination with the presence of surface states, limits its usefulness for the study of amino acids.

4. Conclusion

This work presents the first detailed, systematic exploration of the core state energies of the four aromatic amino acids combining both high resolution XPS and state-of-the-art DFT. A ΔSCF approach, which we have successfully developed and applied to simpler amino acids previously, is extended to amino acids with aromatic side chains and proves robust in predicting the core levels observed in XPS and all contributing local chemical environments. More than 20 additional molecular subspecies are calculated to aid in the discussion and interpretation of the amino acid core states and underpin the assignments made in experimental spectra. This approach provides further understanding and rationalisation of the often complicated and surprizing changes in binding energies observed in the calculations for the solid state amino acids. This work substantially improves our understanding of the aromatic amino acids and gives crucial insights into their intra- and intermolecular structure. Furthermore, it reemphasises the need to combine theory with experiment in order to obtain an accurate and robust picture of the local chemistry and electronic structure and forms the basis for future work on conjugated molecular systems in general.

Acknowledgments

AR acknowledges support from the Analytical Chemistry Trust Fund for her CAMS-UK Fellowship. NKF acknowledges support from the Engineering and Physical Sciences Research Council (EP/L015277/1). LER acknowledges support from an EPSRC Early Career Research Fellowship (EP/P033253/1) and the Thomas Young Centre under Grant No. TYC-101. Calculations were performed on the Imperial College High Performance Computing Service and the ARCHER UK National Supercomputing Service.

ORCID iDs

Anna Regoutz https://orcid.org/0000-0002-3747-3763
Nathalie K Fernando https://orcid.org/0000-0001-7814-1741
Laura E Ratcliff https://orcid.org/0000-0002-9760-5465

References

[1] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864–71
[2] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133–8
[3] Pi J M, Stella M, Fernando N K, Lam A Y, Regoutz A and Ratcliff L E 2020 *J. Phys. Chem. Lett.* **11** 2256–62
[4] Thomas A G, Flavell W R, Chatwin C P, Kumarasinghe A R, Rayner S M, Kirkham P F, Tsoutsou D, Johal T K and Patel S 2007 *Surf. Sci.* **601** 3828–32
[5] Feyer V, Plekan O, Richter R, Coreno M, Prince K C and Carravetta V 2008 *J. Phys. Chem.* A **112** 7806–15
[6] Feyer V, Plekan O, Tsud N, Chaḃ V, Matolín V and Prince K C 2010 *Langmuir* **26** 8606–13
[7] Reichert J *et al* 2010 *ACS Nano* **4** 1218–26
[8] Zhang W, Carravetta V, Plekan O, Feyer V, Richter R, Coreno M and Prince K C 2009 *J. Chem. Phys.* **131** 035103
[9] Zubavichus Y, Zharnikov M, Shaporenko A, Fuchs O, Weinhardt L, Heske C, Umbach E, Denlinger J D and Grunze M 2004 *J. Phys. Chem.* A **108** 4557–65
[10] Cardenas J F and Gröbner G 2006 *J. Electron Spectrosc. Relat. Phenom.* **152** 87–90
[11] Stevens J S, De Luca A C, Pelendritis M, Terenghi G, Downes S and Schroeder S L M 2013 *Surf. Interface Anal.* **45** 1238–46
[12] Boese J, Osanna A, Jacobsen C and Kirz J 1997 *J. Electron Spectrosc. Relat. Phenom.* **85** 9–15
[13] Cooper G, Gordon M, Tulumello D, Turci C, Kaznatcheev K and Hitchcock A P 2004 *J. Electron Spectrosc. Relat. Phenom.* **137–140** 795–9
[14] Meyer F *et al* 2017 *J. Phys. Chem.* B **121** 6549–56
[15] Wang F 2008 *J. Phys.: Conf. Ser.* **141** 012019
[16] Ganesan A and Wang F 2009 *J. Chem. Phys.* **131** 044321
[17] Ganesan A, Mohammadi N and Wang F 2014 *RSC Adv.* **4** 8617
[18] Wang F and Ganesan A 2014 *RSC Adv.* **4** 60597–608
[19] Destro R, Marsh R E and Bianchi R 1988 *J. Phys. Chem.* **92** 966–73
[20] Ihlefeldt F S, Pettersen F B, vonBonin A, Zawadzka M and Görbitz C H 2014 *Angew. Chem., Int. Ed.* **53** 13600–4
[21] Mostad A, Nissen H M, Rømming C, Olli M and Pilotti Å 1972 *Acta Chem. Scand.* **26** 3819–33
[22] Görbitz C H, Törnroos K W and Day G M 2012 *Acta Crystallogr.* B **68** 549–57
[23] Fronczek F R 2016 *CSD Communication* CCDC 1446832
[24] Purushotham U, Vijay D and Narahari Sastry G 2012 *J. Comput. Chem.* **33** 44–59
[25] Ropo M, Schneider M, Baldauf C and Blum V 2016 *Sci. Data* **3** 160009
[26] Kahk J M and Lischner J 2019 *Phys. Rev. Mater.* **3** 100801
[27] Ozaki T and Lee C C 2017 *Phys. Rev. Lett.* **118** 026401
[28] Genovese L *et al* 2008 *J. Chem. Phys.* **129** 014109
[29] Ratcliff L E *et al* 2020 *J. Chem. Phys.* **152** 194110
[30] Goedecker S, Teter M and Hutter J 1996 *Phys. Rev.* B **54** 1703–10
[31] Hartwigsen C, Goedecker S and Hutter J 1998 *Phys. Rev.* B **58** 3641–62
[32] Harrison R J *et al* 2016 *SIAM J. Sci. Comput.* **38** S123–42
[33] Ratcliff L E, Thornton W S, Mayagoitia Á V and Romero N A 2019 *J. Phys. Chem.* A **123** 4465–74
[34] Anderson J, Harrison R J, Sekino H, Sundahl B, Beylkin G, Fann G I, Jensen S R and Sagert I 2019 *J. Comput. Phys.* X **4** 100033
[35] Clark S J, Segall M D, Pickard C J, Hasnip P J, Probert M I, Refson K and Payne M C 2005 *Z. Kristall.* **220** 567–70
[36] Monkhorst H J and Pack J D 1976 *Phys. Rev.* B **13** 5188–92
[37] Grimme S 2006 *J. Comput. Chem.* **27** 1787–99
[38] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865–8
[39] Adamo C and Barone V 1999 *J. Chem. Phys.* **110** 6158–70
[40] Besley N A, Gilbert A T B and Gill P M W 2009 *J. Chem. Phys.* **130** 124308
[41] Momma K and Izumi F 2008 *J. Appl. Crystallogr.* **41** 653–8
[42] Myrseth V, Børve K J and Thomas T D 2007 *J. Org. Chem.* **72** 5715–23
[43] Ohta T, Fujikawa T and Kuroda H 1975 *Bull. Chem. Soc. Japan* **48** 2017–24
[44] Clark D T, Peeling J and Colling L 1976 *Biochim. Biophys. Acta* **453** 533–45
[45] Brena B *et al* 2004 *Phys. Rev.* B **70** 1–6
[46] Gao B, Wu Z and Luo Y 2008 *J. Chem. Phys.* **128** 234704
[47] Cavell R G and Allison D A 1977 *J. Am. Chem. Soc.* **99** 4203–4
[48] Feyer V, Plekan O, Skala T, Chaḃ V, Matolín V and Prince K C 2008 *J. Phys. Chem.* B **112** 13655–60