

PAPER

View Article Online

View Journal | View Issue

Optical backbone-sidechain charge transfer transitions in proteins sensitive to secondary structure and modifications†

I. Mandal,  S. Paul and R. Venkatramani  *

Received 2nd October 2017, Accepted 30th October 2017

DOI: 10.1039/c7fd00203c

The absorption of light by proteins can induce charge transfer (CT) transitions in the UV-visible range of the electromagnetic spectrum. Metal–ligand complexes or active site prosthetic groups which absorb in the visible region exhibit prominent CT transitions. Furthermore, the protein backbone also exhibits CT transitions in the far UV range. In this manuscript, we present a detailed computational study of new near UV-visible CT transitions that involve amino acids with charged side chains. Specifically, using time dependent density functional theory calculations, we examine the absorption spectra of naturally charged amino acids (Lys, Glu, Arg, Asp and His), extracted from solution phase protein structures generated by classical molecular dynamics simulations, and phosphorylated amino acids (Tyr, Thr and Ser) from experimentally determined protein structures. We show that amino acids with charged sidechains present a directed electronic donor–bridge–acceptor paradigm, with the lowest energy optical excitations demonstrating peptide backbone-sidechain charge separations. The UV-visible spectral range of the backbone-sidechain CT transitions is determined by the chemical nature of the donor, bridge and acceptor groups within each amino acid, amino acid conformation and the protein secondary structure where the amino acids are located. Photoinduced CT occurs in opposite directions for the anionic and cationic amino acids along the ground state dipole moment vector for the chromophores. We find that photoinduced charge separation is more facile for the anionic amino acids (Asp, Glu, pSer, pThr and pTyr) relative to that for the cationic amino acids (Lys, Arg and Hsp). Our results provide a foundation for the development of spectroscopic markers based on the recently proposed Protein Charge Transfer Spectra (ProCharTS) which are relevant for the study of DNA-binding or intrinsically disordered proteins that are rich in charged amino acids.

Introduction

The absorption spectra of proteins in the ultraviolet and visible (UV-vis) region of the electromagnetic spectrum are primarily characterized in terms of aromatic

Department of Chemical Sciences, Tata Institute of Fundamental Research, Dr. Homi Bhabha Road, Colaba, Mumbai 400005, India. E-mail: ravi.venkatramani@tifr.res.in

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7fd00203c

amino acid sidechain excitations (Tyr, Trp and Phe) between 250–300 nm^{1–3} and excitations of peptide bonds around 190–220 nm.^{4–7} Such excitations are local in nature, consisting of $n \rightarrow \pi^*$ transitions near 220 nm and $\pi \rightarrow \pi^*$ transitions near 190 nm for the peptide bond and $\pi \rightarrow \pi^*$ transitions near 270 nm for the aromatic amino acids. Additionally, light absorption in proteins leads to charge transfer (CT) transitions in the peptide backbone below 180 nm, which show signatures in both the electronic absorption and circular dichroism spectra of proteins.^{8,9} Other prominent CT transitions in proteins involve metal–ligand complexes or active site prosthetic groups which absorb in the visible region.^{2,10,11} CT transitions, wherein photoexcitation directly transfers electrons between covalently or non-covalently linked donor (D) and acceptor (A) groups, are sensitive to the chemical nature of the D and A groups and the bridging medium (B) separating them. For instance, CT transitions in metal–ligand complexes are sensitive markers of the metal-coordination geometry and can be used to track chemical reactions that are catalysed by such centres in enzymes.¹⁰ Computational studies have proposed that the electronic spectra of proteins below 180 nm are sensitive to the conformation and secondary structure of the polypeptide chain due to the presence of peptide backbone CT transitions.^{8,9,12,13}

Recently, we and our collaborators have reported new photoinduced CT transitions in proteins arising from Lys and Glu amino acids in the near UV-vis region (250–800 nm) of the electromagnetic spectrum.¹⁴ Through computational and experimental studies of α_3C , a synthetic protein that is rich in charged amino acids, we demonstrated that interactions between charged Lys amino and Glu carboxylate side-chains can produce distinct spectral signatures well beyond 250 nm in charged amino acid rich proteins even if they lack aromatic amino acids. In the past decade and a half, there have been reports of non-aromatic near-UV/visible absorption (emission) extending up to 450 nm (600 nm) in high concentration protein solutions and peptide aggregates.^{15–17} However, our studies indicate that even at low concentrations, monomeric proteins that are rich in charged amino acids should also exhibit significant absorption in the near UV-vis region. Indeed, monomeric α_3C (5–105 μM) solutions demonstrated distinctive absorption features spanning 250–800 nm ($\epsilon = 7338 \pm 191 \text{ M}^{-1} \text{ cm}^{-1}$, $964 \pm 129 \text{ M}^{-1} \text{ cm}^{-1}$, and $501 \pm 66 \text{ M}^{-1} \text{ cm}^{-1}$ at 250, 450 and 800 nm, respectively). The discovery of new photoinduced CT transitions for Lys and Glu amino acids adds a new spectral window (250–800 nm) which we call ProCharTS (Protein Charge Transfer Spectra), for label free tracking of biomolecular structure and dynamics.

Our previous study¹⁴ also showed that ProCharTS signatures are also prominently present in natural proteins that are rich in charged amino acids containing aromatic chromophores. While ProCharTS features extend up to 800 nm, well beyond the absorption range of aromatic amino acids (up to 320 nm), the presence of ProCharTS has implications even for spectral ranges below 300 nm. Near-UV absorbance at 280 nm, which is typically only assumed to be from aromatic amino acids, and which can be used to quantify protein solutions (*e.g.* for estimating concentrations), should account for significant ProCharTS absorption when the protein is rich in charged amino acids. For instance, the α_3C protein which lacks aromatic amino acids still exhibits significant ProCharTS absorption ($\epsilon = 4531 \pm 133 \text{ M}^{-1} \text{ cm}^{-1}$) at 280 nm. A better understanding of the ProCharTS spectral features below 300 nm, which is the range where monomer charged amino acids are expected to contribute, can thus significantly improve the

quantification of protein solutions. Finally, the broad ProCharTS profile will always have some overlap with the emission spectra of natural fluorescent chromophores such as Trp or synthetic dyes. This overlap may produce multi-exponential fluorescence lifetime decay kinetics which can be interpreted in terms of the dynamics and interactions of charged amino acids within protein folds. Charged amino acids such as Lys and Arg are integral constituents of DNA and RNA binding proteins such as histones, spliceosomal proteins and transcription factors.¹⁸ Furthermore, intrinsically disordered proteins, which are implicated in both cell regulatory events as well as pathological disorders, are also rich in charged amino acids.¹⁹ In order to develop the potential of ProCharTS to examine such proteins of biological and biomedical relevance, it is imperative to understand the distinct spectral ranges and features of photoinduced transitions in natural charged amino acids.

In our previous study,¹⁴ we uncovered two distinct photoinduced CT transitions involving Lys and Glu amino acids within the helical protein α_3 C: (1) Peptide Backbone-Sidechain (PBS) and (2) Sidechain-Sidechain (SS) CT transitions. PBS-CT transitions lead to charge separation between the charged carboxylate/amino sidechain groups of Glu/Lys and the peptide backbone. In contrast, SS-CT transitions occur between the Glu carboxylate and the Lys amino sidechain groups. We systematically showed that the broad 250–800 nm ProCharTS spectrum of α_3 C could be explained by considering the sensitivity of PBS-CT and SS-CT transitions to the inter-residue contacts imposed by protein tertiary folds and their dynamics, solvation of the chromophores, and the protonation states of the amino acids. Based on these results, we hypothesize that charged amino acids, in isolation or in collective assemblies, form electronic D–B–A molecular complexes that exhibit photoinduced D to A CT transitions. Here, we validate our hypothesis by conducting a systematic investigation of the charged chromophores that are available in protein sidechains (Fig. 1).

In this manuscript, we present a detailed computational study and comparison of the absorption spectra of all amino acids that are charged under physiological conditions: Lys, Glu, Arg, Asp and His (Fig. 1). While the absorption spectra for the Lys and Glu amino acids located within α -helices were explored in our previous study, here we present new calculations for these amino acids within other secondary structure motifs (β -sheet and random coil). Furthermore, we also examine the absorption spectra of phosphorylated forms of the amino acids Ser, Thr and Tyr (pSer, pThr and pTyr), wherein negatively charged phosphate groups are post-translationally attached to these otherwise neutral amino acids (Fig. 1). We carry out Time Dependent Density Functional Theory (TDDFT) based spectra calculations on the charged amino acids extracted from a microsecond (μ s) long classical molecular dynamics (MD) trajectory of the protein ubiquitin and from experimentally derived protein data bank structures of phosphorylated proteins. Our studies are aimed at demonstrating the presence of facile PBS-CT transitions in the absorption spectra of charged amino acid chromophores and comparing the properties of these transitions as a function of the following amino acid properties: (1) chemical structure, (2) conformation and (3) location in distinct secondary structural elements.

Our results have a direct bearing on the design of optical probes based on ProCharTS for monitoring the structure, dynamics and function of proteins that are rich in charged amino acids or that undergo post-translational modifications.

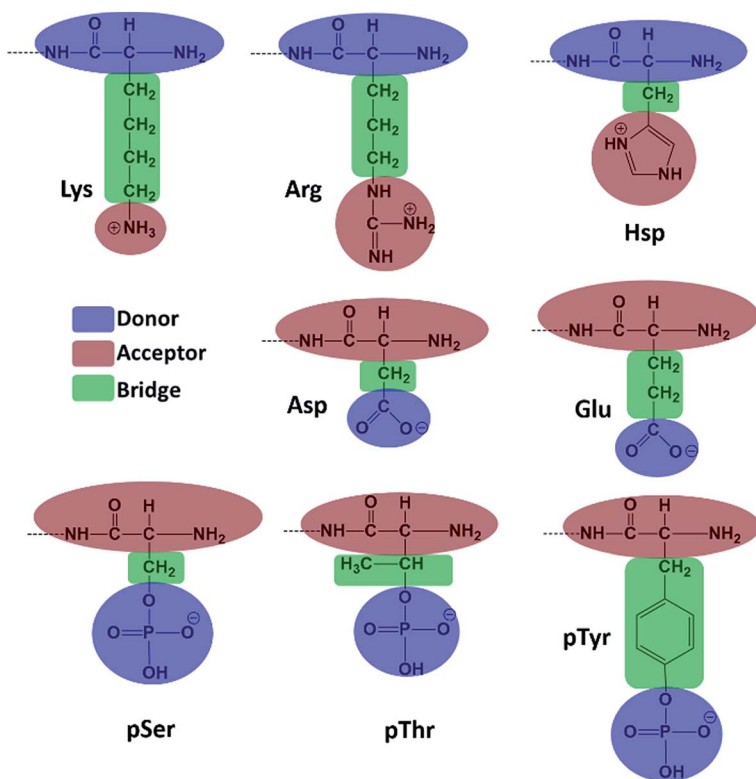


Fig. 1 Charged amino acids considered in this study. The colored segments indicate the electronic donor (D)–bridge (B)–acceptor (A) segments for the PBS-CT transitions expected in these systems.

Methods

The goal of the present manuscript is to examine the nature of PBS-CT transitions in the absorption spectra of charged and post-translationally modified amino acids located within protein folds. To this end, we carried out TDDFT calculations on amino acid conformations sampled from a classical atomistic MD trajectory of fully solvated ubiquitin, a globular protein with a β -grasp fold. Furthermore, we extracted multiple post-translationally modified amino acid conformers from the protein data bank. We then carried out a detailed analysis of the ground state electronic properties and UV-vis absorption spectra for each charged amino acid chromophore ensemble. Our methodology is described in the following subsections.

Molecular dynamics simulations of ubiquitin

Ubiquitin is a small 76 residue globular protein with a β -grasp topology consisting of 5 β -strands, a single α -helix, and a 3–10 helix (Fig. 2). In a biological context, the protein serves as a post-translational molecular tag for the regulation of cellular processes. The protein contains 7 Lys residues which play a critical role in the

function of the protein. The sequence of ubiquitin consists of 23 amino acids (7 Lys, 6 Glu, 5 Asp, 4 Arg and 1 His) whose sidechains exist in charged forms at physiologically relevant pH values. Fig. 2A shows the location of these charged amino acids in the ubiquitin sequence and their structural arrangement over the ubiquitin fold. Initial coordinates for ubiquitin were taken from a crystal structure in the protein data bank (PDB code: 1UBQ). We added hydrogen atoms and solvated the protein in a cubic box of explicit water molecules with 10 Å solvent padding. The solvated models were then subjected to an equilibration protocol and subsequent equilibrium MD production runs using GROMACS version 5.0²⁰ and the CHARMM36 force field with the TIP3P water model for the solvent.²¹ Our simulations employed periodic boundary conditions (PBC) with Particle Mesh Ewald (PME) calculations of long range electrostatic interactions and a cut-off of 1.0 nm for short range van der Waals and electrostatic interactions. The equilibration protocol consisted of a minimization step and thermal equilibration at

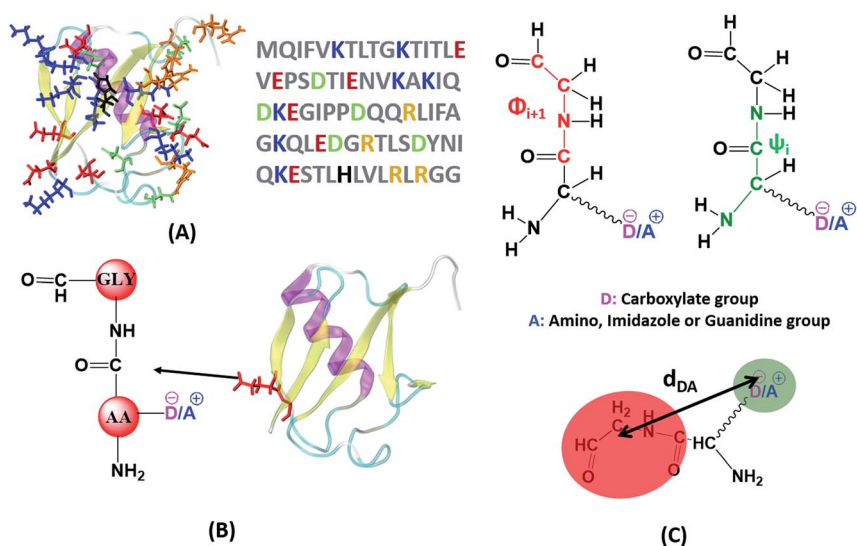


Fig. 2 (A) The structure (PDB code: 1UBQ) and amino acid sequence of ubiquitin. Ubiquitin has 23 charged amino acids whose position in the sequence and structure are indicated by colored single letter codes and colored licorice representations, respectively. (B) A schematic representation of the procedure that was used to create charged amino acid monomer models for our electronic structure calculations. We extracted a dimer (AA–XX) containing one of the 23 charged amino acids (AA = Lys/Glu/Asp/Arg/His), mutated the neighbouring amino acid (XX) to Gly and capped the amino and carbonyl ends with hydrogens. (C) We use two geometric parameters to describe the variation in conformation of the charged amino acid monomer models sampled from the MD trajectory of ubiquitin: (top) Ramachandran torsion angles (Φ_{i+1} and Ψ_i) and (bottom) the donor–acceptor distance (d_{DA}). Here, Ψ_i is the Ramachandran dihedral angle for the charged amino acid at the i^{th} position in the sequence, calculated using the C–C $_{\alpha}$ –N–C atoms (highlighted in green), and Φ_{i+1} is the Ramachandran dihedral angle for the neighbouring residue (represented by Gly) at the $i + 1^{\text{th}}$ position, calculated using the N–C $_{\alpha}$ –C–N atoms (highlighted in red) of the dimer fragment. The distance d_{DA} is measured between the geometric centers of the extended backbone of the dimer fragment (red shaded region) and the sidechain headgroup of the charged amino acid (green shaded region).

300 K for 100 ps. Both of these steps were carried out initially with the protein's heavy atoms fixed, later constraining them harmonically around their structural positions with a spring constant of $25 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The solvated protein system was then equilibrated further under constant temperature and pressure (NPT) conditions at 300 K and 1 atm. We carried out successive NPT equilibration runs (150 ps each), during which harmonic constraints on the heavy atoms were gradually reduced from $25\text{--}5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. Finally, after an unconstrained 1 ns NPT equilibration run, an NVT production trajectory of $1 \mu\text{s}$ was generated. The μs long simulation of the ubiquitin structure in explicit solvent provided us with a solvated form for the protein with inter-residue sidechain interactions matching that of NMR derived solution structures (Fig. S1 of the ESI†). We uniformly sampled 100 charged amino acid conformations from the protein snapshots in the $1 \mu\text{s}$ trajectory for further analysis and spectra calculations. The sampling scheme is described in detail in the next subsection.

TDDFT calculations for amino acids with charged sidechains

In order to calculate the absorption spectra of Lys, Glu, Arg, Asp, His and doubly protonated His (Hsp), we extracted chromophore fragments from uniformly sampled protein snapshots from the $1 \mu\text{s}$ MD run of ubiquitin. The sampling covered multiple conformations (14–100) along the MD trajectory (protein snapshots captured every 70–10 ns) for all possible instances of each charged amino acid in the protein sequence. For instance, there are 7 Lys in the ubiquitin sequence (Fig. 2A). We therefore uniformly sampled 15 conformations for 2 of the 7 Lys and 14 conformations for the remaining 5 Lys to arrive at 100 monomer conformations. In contrast, since there is only one His in the ubiquitin sequence, we sampled 100 His monomer conformations from the MD trajectory. Through this sampling scheme we captured both the thermal conformational fluctuations for each charged amino acid monomer as well as the variations in conformation of the charged amino acids due to their placement in the protein fold.

For each sampled amino acid monomer conformation, we created a model with an extended backbone which mimics the polypeptide chain (Fig. 2B). Specifically, we extracted a dimer fragment containing the charged amino acid of interest (Lys/Arg/His/Glu/Asp) and sidechain atoms of the residue adjacent to the charged amino acid were deleted in order to mutate the residue to Gly. This procedure created an extended backbone monomer model. Then the dangling bonds on the backbone were capped with hydrogens to create a modified C terminus (CHO group) and N terminus (NH_2 group) using the psfgen module in VMD.²² This capping strategy has been previously used by us before,¹⁴ and we discuss its suitability along with control calculations in the last subsection of the “Methods” section. We generated 100 conformations for each charged amino acid (Lys/Glu/Asp/Arg/His) by sampling multiple MD conformations for all 23 charged amino acids in ubiquitin. We considered two protonation states for His, with the neutral form protonated at the epsilon nitrogen (His) and the doubly protonated form with protons at both the epsilon and delta nitrogens (Hsp) of the imidazole ring. For the 100 conformations of each charged amino acid, we calculated the absorption spectra using TDDFT with a CAM-B3LYP²³ functional and a 6–31++G(d) basis set on all atoms in the Gaussian 09 program.²⁴ The first 100 lowest energy electronic transitions were calculated in all TDDFT calculations.

Difference electron density plots were calculated using the Multiwfn 3.3.8 software²⁵ and visualised (isovalue set to 0.0004) in GaussView 5.0.²⁶

Our previous studies on Lys/Glu indicated that the photoinduced PBS-CT transitions should be sensitive to the geometric (structural) parameters of the amino acids which influence the coupling between electronic donor-acceptor groups or the delocalization of the electronic donor and acceptor states. Two such geometric parameters (Fig. 2C) are the backbone-charged sidechain headgroup distance (d_{DA}) and the backbone torsion/Ramchandran angles for our monomer fragment models (Fig. 2C). Fig. S2–S5 of the ESI† show the variation in the geometric parameters (d_{DA} and torsion angles) defined in Fig. 2C over the 100 amino acid conformations sampled from the ubiquitin MD trajectory. In later sections we examine the sensitivity of the excitation energies and oscillator strengths for the PBS-CT transitions in charged amino acid monomer models to the geometric parameters (Fig. 2C) and the chemical nature of the electronic donor-bridge-acceptor complex.

TDDFT calculations for post-translationally modified amino acids

The post-translationally modified amino acid fragments of Ser, Thr and Tyr were extracted from 10 different experimental structures of proteins containing those residues from the protein data bank (<https://www.rcsb.org/>).²⁷ Post-translationally modified amino acid fragments include mono-anionic (singly protonated) phosphorylated Ser, Thr and Tyr residues. Although di-anionic (unprotonated) phosphate group forms exist under physiological conditions, we did not consider them here in our spectra calculations. The spectra of the mono-anionic forms can be more directly compared with the spectra of the singly charged amino acids (Lys, Glu, Arg, Asp and His). The pSer residues were extracted from 5 experimental structures (PDB codes: 4JFZ, 2FWN, 4KK4, 5TR5 and 2LO6), the pThr residues were extracted from 5 experimental structures (PDB codes: 2ERK, 3Q52, 3A62, 2JQL and 3OB1), and the pTyr residues were extracted from 6 experimental structures (PDB codes: 2ERK, 3OBI, 2DVJ, 2IVT, 1IR3, 1TRN and 1VX3). Specifically, we extracted a dimer fragment containing the phosphorylated version of the amino acid of interest (pThr/pSer/pTyr) and the sidechain atoms of the adjacent residue were deleted in order to mutate the residue to Gly. This procedure created an extended backbone monomer model with dangling bonds capped with hydrogens, as described in the previous subsection. TDDFT absorption spectra were computed (first 100 lowest energy electronic transitions), as described in the previous section, for 10 conformations of each phosphorylated amino acid.

Characterization of transitions

We computed two measures in order to characterize the transitions as charge transfer (CT) transitions or non-CT transitions. One measure is the average hole-electron separation distance:^{25,28}

$$\Delta r = \frac{\sum_{ij} (K_i^j)^2 |\langle \varphi_j | r | \varphi_i \rangle - \langle \varphi_i | r | \varphi_j \rangle|}{\sum_{ij} (K_i^j)^2} \quad (1)$$

where ϕ is the molecular orbital (MO) and the indexes i and j go over all occupied and vacant MOs, respectively. Here $K_i^j = P_i^j + Q_i^j$, where P_i^j and Q_i^j are the excitation ($i \rightarrow j$) and de-excitation ($j \leftarrow i$) configuration coefficients. The second measure^{25,29} is the distance between the centroid of the hole and electron distribution (D_{CT}), defined as

$$D_{CT} = \sqrt{(D_{CT,X})^2 + (D_{CT,Y})^2 + (D_{CT,Z})^2} \quad (2)$$

with

$$D_{CT,\alpha} = \left| \int \alpha \rho^{\text{electron}}(r) dr - \int \alpha \rho^{\text{hole}}(r) dr \right| \quad (3)$$

where the index α represents the Cartesian components (X, Y, Z) and $\rho^{\text{electron/hole}}$ is the electron/hole density distribution. The two measures were calculated using Multiwfn 3.3.8²⁵ and then the following conditions were used to categorize the transitions: (i) CT transition when $\Delta r > 2 \text{ \AA}$ and $D_{CT} > 1 \text{ \AA}$, (ii) non-CT transition when $\Delta r < 2 \text{ \AA}$ or if $\Delta r > 2 \text{ \AA}$ and $D_{CT} < 1 \text{ \AA}$. Furthermore, transitions for which Δr was within 5% of the threshold ($2 \text{ \AA}$) value were classified as borderline transitions. In the manuscript we only show the classification of transitions as either CT or non-CT transitions.

Critical discussion of modelling assumption and caveats

We have previously explored the solvation effects and inter-residue sidechain interactions for Lys and Glu chromophores using full QM treatments on systems of varying size and sampling conformations with different charge states.¹⁴ Our previous study showed that the essential features of photoinduced CT transitions for Lys/Glu were present in the monomer spectra of these amino acids. Thus, in the present study, we examine only the monomer charged amino acid spectra with the goal of comparing the features of photoinduced CT transitions for anionic (Glu, Asp, pThr, pSer and pTyr) and cationic (Lys, Arg and His) amino acids. Since His can occur in both neutral (singly protonated) and charged (doubly protonated) states under physiological conditions, here we computed the electronic structures of both the singly protonated neutral (His) and doubly protonated cationic (Hsp) forms.

In the present study, we carry out spectra calculations on a large number of monomer amino acid fragments using the TDDFT method, which has been shown by us and others to be the method of choice for molecular systems with a good balance of cost and accuracy.^{14,15,30–32} Furthermore, in our TDDFT calculations, we employ the CAM-B3LYP range corrected hybrid exchange-correlation (XC) functional which provides a reasonable description of CT states.^{23,33,34}

Our goal in the present study is to examine the spectral features of amino acid chromophores embedded within protein folds. To this end, we constructed amino acid fragment models with extended backbones which mimic the polypeptide chain. This was accomplished by extracting a dimer fragment (AA-XX), where AA was one of the 23 charged amino acids in the ubiquitin sequence and XX was its neighboring residue. We mutated XX to Gly and capped the amino and carboxylate ends with hydrogens. We have critically examined the capping strategies for amino acids used in our electronic structure calculations by carrying out control calculations on extended peptide backbones and through utilizing

alternative capping models using methyl groups.¹⁴ Based on these calculations, we concluded that extending the polypeptide backbone by adding an extra Gly unit to the C-terminal end of amino acid fragments provided robust converged spectra above 250 nm. Our control calculations (using Glu as a reference) show that further extending the backbone by adding Gly units or using methyl capping groups only alters the backbone transitions around 200 nm. However, backbone transitions and signature CT transitions involving charged amino acid sidechains/backbones near and above 250 nm are not altered by either extending the backbone or changing the capping groups. Here, we present new control calculations on Lys/Arg chromophores (Fig. S6 of ESI†) with more extended peptide backbones. These calculations show that using a more extended polypeptide backbone slightly increases the range of Lys and Arg excitation wavelengths towards lower energies by 30 and 10 nm, respectively. However, relative trends in the ranges for the cationic systems (Lys *vs.* Arg) or between cationic and anionic systems (Lys/Arg data from Fig S6† *vs.* Glu data from ref. 14) remain unaltered. Furthermore, the nature of the transitions is also preserved, with the lowest energy transitions demonstrating distinct PBS-CT character. Thus, the present capping scheme provides converged trends with respect to the observations reported and discussed in this study. We note that the capping scheme used here does produce a small number of fictitious low intensity non-CT transitions above 300 nm which are localized on our capping groups (see Fig. 6).¹⁴ However, we were able to cleanly identify and separate these non-CT transitions above 300 nm from our characteristic CT transitions of interest. Thus our results and conclusions on PBS-CT transitions are free of any artefacts introduced by the capping scheme.

Results and discussion

Ground state electronic structures of charged amino acids present a directed D–B–A paradigm

The absorption of light by a molecular scaffold can create charge separation, provided the scaffold possesses distinct electron donor (D) and electron acceptor (A) moieties. Such groups may be coupled to each other within the molecular scaffold *via* a bridging medium (B) with less propensity (relative to D/A) to accept/donate electrons.^{35,36} Fig. 3 shows the ground state energy gap distributions and representative localizations of the HOMO and LUMO for cationic Arg/Lys/Hsp, neutral His, and anionic Glu/Asp/pSer/pThr/pTyr. For the cationic species we find the HOMO to be localized on the peptide backbone, thus marking this group as the electronic D unit. In contrast, the LUMO is localized on the sidechain guanidine, amino and imidazole moieties of Arg, Lys and Hsp, respectively, thus marking these groups as electronic A units. The alkyl sidechain which possesses σ -character connects these D/A groups to form the B unit for the cationic amino acids. In contrast, for the anionic species the peptide backbone reverses its role to form the electronic A unit with the LUMO localized there. Here, the sidechain carboxylate moieties of Glu/Asp and phosphate moieties of the phosphorylated residues form the electronic D unit, with the HOMO localized on these groups. Except for the case of pTyr, where the sidechain B unit has a distinctive π -character due to the presence of the aromatic ring, the B unit for all other anionic amino acids is an alkyl chain with σ -character. Given that for both anionic and

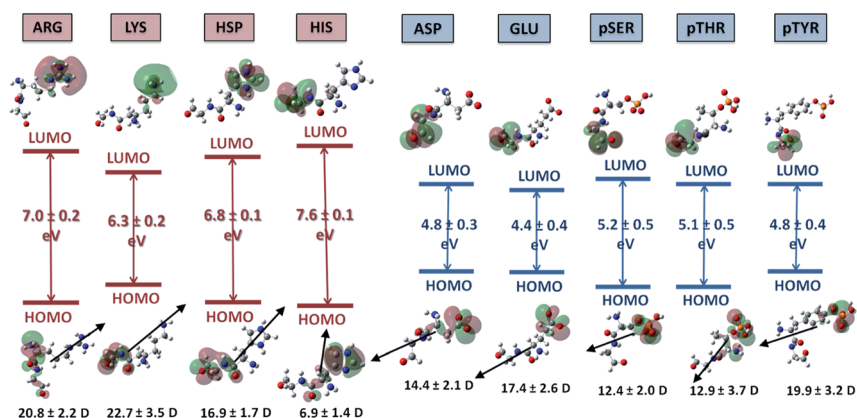


Fig. 3 Average ground state HOMO–LUMO gaps and dipole moment values along with their standard deviation from 100 conformations of all of the charged amino acids. The representative molecular structures at the top and bottom of the energy gap diagrams are identical, showing HOMO (bottom structures) and LUMO (top structures) contours plotted with an isovalue of 0.02. Ground state dipole vectors (arrows going from the negative end to the positive end on the bottom structures) for each system are also shown.

cationic systems, the D and A groups are populated with filled and empty frontier orbitals, respectively, we anticipate the lowest energy photoinduced electronic transitions in these ensembles to be PBS-CT transitions which transfer charge from D to A groups. The anionic systems exhibit considerably smaller (by 1–3 eV) HOMO–LUMO energy gaps relative to those of the cationic systems, thus indicating much more facile charge transfer and lower energy PBS-CT transitions for the former relative to the latter. Since the anionic (excluding pTyr) and cationic systems possess the same peptide backbone and a σ -bridge, the differences in ground state energy gaps can be attributed to the stronger electronic character of the anionic carboxylate and phosphate A groups relative to the guanidine, amino and imidazole D groups of the cationic systems.

For both anionic and cationic systems, the charged state of the sidechains polarizes the filled and empty orbitals in opposite directions. For the neutral His system, the frontier filled and empty orbitals are distributed indiscriminately over the peptide backbone and the imidazole ring. In the representative structure shown in Fig. 3, the HOMO is located mostly on the imidazole ring while the LUMO is localized on the peptide backbone. In contrast, for the doubly protonated HSP chromophore, the excess positive charge on the imidazole ring polarizes the molecular orbitals to shift the empty orbitals to the imidazole ring, thus transforming it to an electronic A group. The filled orbitals are in turn shifted to the peptide backbone, thus transforming it to an electronic D group. The dipole moment of HSP is also more well defined (about 10 Debye larger) relative to that of neutral His. Both cationic and anionic systems exhibit well defined dipole moments pointing from their D to A groups indicating the direction of photo-induced CT (*vide infra*). However, we find no systematic correlation between the dipole moment magnitude and the energy gap which could serve as a ground state proxy for predicting PBS-CT excitation energies. The dipole moment of cationic systems is larger than that for anionic systems, despite the latter

possessing significantly smaller energy gaps than the former. Within the cationic species, Lys has the largest dipole moment and the lowest energy gap. However, the dipole moment of Arg is slightly greater than that of Hsp, while both amino acids display similar energy gaps. Overall, the average energy gaps (E_{gap}) and the ground state dipole moment magnitudes (μ_{gs}) are both well resolved for the cationic species and follow the trends $E_{\text{gap}}(\text{Lys}) < E_{\text{gap}}(\text{Arg}) \sim E_{\text{gap}}(\text{Hsp})$ and $\mu_{\text{gs}}(\text{Lys}) > \mu_{\text{gs}}(\text{Arg}) > \mu_{\text{gs}}(\text{Hsp})$. Within the anionic species, pTyr possesses the largest dipole moment but its energy gap is comparable to that of Asp whose dipole moment is significantly smaller. The variance in the energy gaps is large (0.3–0.5 eV) for the anionic species, and differences in their average energy gaps cannot be resolved. The dipole moment magnitudes for the anionic species follow the trend $\mu_{\text{gs}}(\text{Glu}) \sim \mu_{\text{gs}}(\text{pTyr}) > \mu_{\text{gs}}(\text{pAsp}) \sim \mu_{\text{gs}}(\text{pSer}) \sim \mu_{\text{gs}}(\text{pThr})$.

To summarize this section, our ground state analysis of both cationic and anionic species reveals the significant polarization of orbitals due to excess charge in the side-chains, which enables us to clearly identify the electronic D–B–A groups for each chromophore. The ground state analysis reveals that the electronic D and A states are populated by frontier filled and empty orbitals, respectively, thus suggesting that the lowest energy transitions in these systems would be PBS-CT transitions. Furthermore, all of the charged amino acids examined here possess well defined dipole moments vectors pointing from D to A indicating the direction of photoinduced charge transfer.

The lowest energy transitions of anionic amino acids are PBS-CT transitions directed from the amino acid sidechains to the peptide backbone

We computed the optical absorption spectra (first 100 transitions) for all of the anionic amino acids using TDDFT electronic structure calculations. The top row of panels in Fig. 4A shows the computed spectra above 200 nm for 100 conformations of Asp and Glu sampled from the ubiquitin MD trajectory. Fig. 4B shows the spectra above 200 nm for 10 conformations of the three phosphorylated residues, pSer, pThr and pTyr, extracted from experimentally determined protein structures (see the Methods section). The insets in the top panels of Fig. 4A and B show the lowest energy transitions for each species in terms of difference density plots which present the photoexcited hole (pink) and electron (blue) densities on each amino acid fragment. In the lower panels of Fig. 4A and B, we present a classification of the transitions into CT or non-CT transitions based on two measurements of charge separation (see the Methods section). These panels show the percentage of CT and non-CT transitions within 5 nm windows over the whole absorption spectral range for each of the charged chromophores.

The computed spectrum for Glu in ubiquitin extends up to 483 nm, which is about ~ 34 nm higher relative to the computed spectral range for Glu residues from $\alpha_3\text{C}$ obtained in our previous study.¹⁴ We attribute this decrease in the PBS-CT excitation energies to the difference in secondary structure folds of ubiquitin and $\alpha_3\text{C}$ (*vide infra*). The difference density plots for the lowest energy transition for Glu show well separated photoexcited electron and hole densities, which reside on the peptide backbone and Glu carboxylate groups, respectively. This transition is therefore a PBS-CT transition from D (carboxylate) to A (peptide backbone), in accordance with our ground state analysis (previous subsection). The difference density plots for the lowest energy transitions of Asp mirror that of

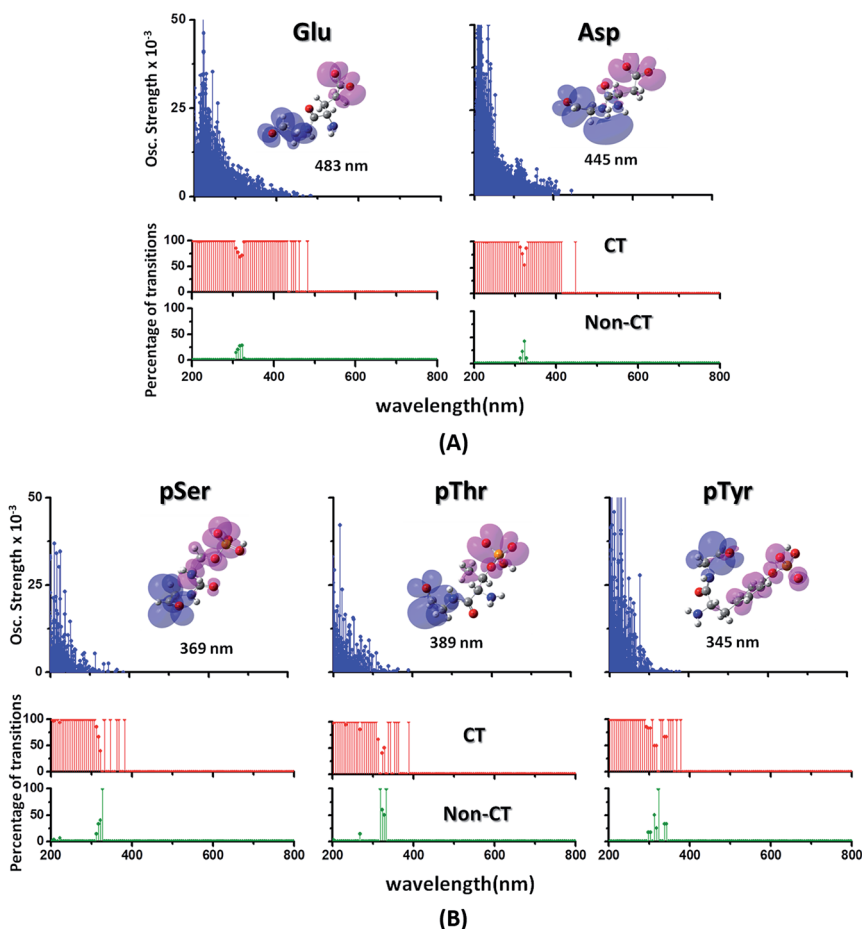


Fig. 4 Simulated absorption spectra (wavelength vs. oscillator strength) for the anionic amino acid monomers. The top panels show spectra for (A) 100 conformations of the Glu and Asp amino acids sampled from a 1 μ s MD trajectory of ubiquitin and for (B) 10 different conformations of pSer, pThr and pTyr taken from crystal and NMR structures from the Protein Data Bank. Each amino acid fragment (chemical structures shown in Fig. 1) has an extended backbone capped with hydrogen atoms (see the Methods section). The lower panels in (A) and (B) show the assignment of transitions in the spectra to CT vs. non-CT transitions across the absorption profile. Difference density plots (the pink lobes represent hole density and the blue lobes represent electron density after photoexcitation) for the lowest energy transitions for each spectrum are shown in the insets of the top panels in (A) and (B).

Glu, showing a PBS-CT transition from D (carboxylate) to A (peptide backbone). Glu and Asp share the same electronic D and A states but differ in the length (1 methylene unit for Asp vs. 2 for Glu) of the sidechain B group. Intuitively, the shorter bridge length in Asp should lead to a stronger coupling of D and A states, thus leading to a more facile photoinduced CT relative to Glu. However, Fig. 4 shows that the spectral range of Asp (extending up to 445 nm) is in fact slightly lower than that observed for Glu. The shorter bridge length for Asp neither lowers the excitation energy of the PBS-CT transitions nor does it increase the oscillator strength for the low energy transitions relative to that of Glu. These results are

consistent with the ground state calculations (Fig. 3) which show that the average HOMO–LUMO gaps for Asp and Glu are comparable and that the lowest energy gap value for Asp is about 0.3 eV higher than that for Glu. The lower panels in Fig. 4 for Asp and Glu indicate that the spectra above 200 nm for these charged species are dominated by CT transitions. The few low intensity non-CT transitions above 300 nm are spurious transitions localized on the capping group,¹⁴ and we do not ascribe any physical significance to them.

The computed spectra (Fig. 4) for the phosphorylated residues pSer, pThr and pTyr all lie below 400 nm, with the lowest energy transitions in these systems appearing at 369 nm, 389 nm and 345 nm, respectively. The difference density plots for the lowest energy transitions in all three systems show well separated photo-excited hole and electron densities on the phosphate and peptide backbone groups, thus indicating that these are PBS-CT transitions. The significantly higher excitation energies (0.4–0.8 eV) required for PBS-CT in phosphorylated amino acids relative to Asp/Glu loosely correlate with our ground state analysis (Fig. 3) which shows that the lowest energy gaps for the phosphorylated species are 0.1–0.9 eV higher than for Asp/Glu. The pSer and Asp chromophores share the same A (peptide backbone) and B (methylene) bridge, but differ in the nature of the D (carboxylate for Asp *vs.* phosphate for pSer) group. Thus, a comparison of the Asp and pSer spectra suggests that the phosphate group is a weaker electron donor than the carboxylate group. The phosphorylated amino acids share the same D and A group but differ in their B unit. Both pSer and pThr have similar B units which possess σ character. Accordingly, the ground state analysis reveals very similar energy gaps for pSer and pThr (Fig. 3), which correlates well with the similar spectral features observed for the two systems (Fig. 4). The pTyr chromophore, on the other hand, is distinct from the other anionic species in that the B unit possesses π character. In fact, the hole density in Fig. 4 shows significant delocalization on the phosphate group and the Tyr aromatic ring to form an extended D state. While the presence of a π bridge does not significantly change the average ground state energy gap (Fig. 3) relative to that of the other phosphorylated systems, the lowest energy gap for pTyr is still 0.4–0.5 eV lower than that for pSer and pThr (Fig. 4). However, in contrast with the ground state energy gap observations, the lowest energy PBS-CT transition for pTyr is 0.2–0.4 eV higher than that observed for pSer and pThr. This case illustrates the complexities in comparing PBS-CT excitation energies for systems with distinct B units. The pTyr spectra shows more intense transitions in the 200–300 nm window relative to those of the other phosphorylated species. The lower panels in Fig. 4B show that the entire spectral range above 200 nm for the phosphorylated species is dominated by CT transitions.

For a D–B–A complex, the lower bound of the TDDFT excitation energies for CT transitions (E_{CT}) is given by the relation:³⁷

$$E_{\text{CT}} = (\text{IP} - \text{EA}) - \frac{1}{R} \quad (4)$$

where IP and EA represent the ionization potential of the donor and electron affinity of the acceptor, respectively, and the last term represents the electrostatic binding energy of the electron–hole pair created by the excitation and separated by a distance R . An orbital decomposition of the lowest energy PBS-CT transitions for Glu and Asp (Fig. S7 of ESI[†]) reveals that these transitions are predominantly (>96%) composed of the HOMO–LUMO D–A pairs. Thus, for these systems, we

can approximate eqn (4) for the lower bound of the PBS-CT excitation energies in terms of the HOMO–LUMO energy gap (ΔE) and the charge separation index³⁴ D_{CT} (eqn (2)) as:

$$E_{CT} = \Delta E - \frac{1}{D_{CT}} \quad (5)$$

Fig. 5 shows that the excitation energies for the lowest energy PBS-CT transitions from 100 conformations of Glu and Asp correlate strongly with the HOMO–LUMO energy gaps and moderately with the hole–electron binding energy. For the phosphorylated systems, we observe a greater degree of mixing of the higher lying orbitals (*e.g.* LUMO+1) in the composition of the PBS-CT transitions. We thus expect the correlation of excitation energies with HOMO–LUMO gaps to be weaker for these systems.

The lowest energy transitions of cationic amino acids are PBS-CT transitions directed from the peptide backbone to the amino acid sidechains

We next examine the computed optical absorption spectra (100 lowest energy transitions) for the three cationic amino acids. The top panels in Fig. 6 show the TDDFT spectra beyond 200 nm for 100 conformations of the Lys, Arg and Hsp monomers sampled from the 1 μ s MD trajectory of ubiquitin. As with the anionic species (previous subsection), we display difference density plots for the lowest energy transitions (insets) and a classification of the transitions as CT and non-CT transitions (lower panels).

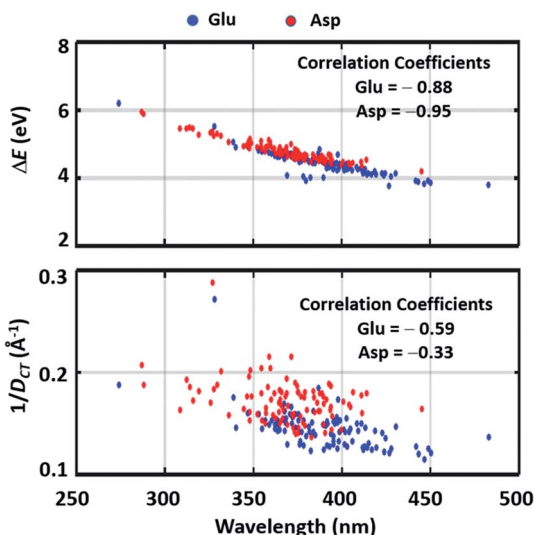


Fig. 5 (Top panel) Correlation between the HOMO–LUMO energy gap and the excitation wavelength for the lowest energy transitions of Asp and Glu. (Bottom panel) Correlation between the inverse charge separation distance (eqn (2)) and the excitation wavelength for the lowest energy transitions of Asp and Glu. The data was generated from 100 MD conformations of Asp and Glu.

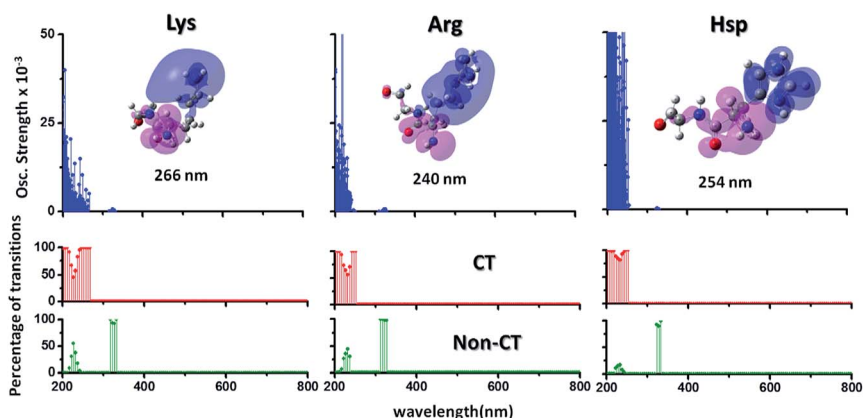


Fig. 6 Simulated absorption spectra (wavelength vs. oscillator strength) for the cationic amino acid monomers. The top panels show spectra for 100 conformations of the Lys, Arg and Hsp amino acids sampled from a 1 μ s MD trajectory of ubiquitin. Each amino acid fragment (chemical structures shown in Fig. 1) has an extended backbone capped with hydrogen atoms (see the Methods section). The lower panel for each amino acid fragment shows the assignment of transitions in the spectra to CT vs. non-CT transitions across the absorption profile. Difference density plots (the pink lobes represent hole density and the blue lobes represent electron density after photoexcitation) for the lowest energy transitions for each spectrum are shown in the insets of the top panel.

The spectra for all three cationic species lie below 300 nm. The Lys spectra extends up to 266 nm which matches the range obtained by us earlier for amino acid conformers in α_3 C.¹⁴ The spectral range of the Arg and Hsp chromophores is similar to that of Lys, extending up to 240 nm and 254 nm, respectively. The difference density plots reveal clear backbone to sidechain PBS-CT transitions for the lowest energy excitations of the Lys, Arg and Hsp ensembles. For all three chromophores, the lowest photoexcited transition shows the hole density to be mostly localized on the peptide backbone, with some delocalization on the alkyl sidechain. On the other hand, the electron density for the lowest energy transition is mostly localized on the sidechain amino (Lys), guanidine (Arg) and imidazole (Hsp) groups, with some delocalization on the alkyl chain. Thus, for the lowest energy PBS-CT transitions in the cationic systems, we see more extended D/A units and less well-defined B units. In contrast, the anionic systems (Fig. 4) show more localized and well defined D, B and A units. Overall, the lowest PBS-CT excitation energy of the conformational ensemble of cationic amino acids is $\sim$ 1.0–2.5 eV higher than that of the anionic amino acids. This correlates well with our ground state analysis (Fig. 3) which shows that the HOMO–LUMO energy gaps for the anionic systems are up to 2.5 eV lower than those for the cationic systems.

The spectra for Arg and Hsp show significantly higher oscillator strengths between 200–240 nm relative to Lys. This is likely due to the more extended conjugation in the Arg guanidine and Hsp imidazole moieties relative to the Lys amino group. On the other hand, the excess positive charge on the guanidine and imidazole moieties is more delocalized and negatively impacts the polarization and stabilization of the ground state orbitals. For instance, in the ground state picture (Fig. 3), the HOMO–LUMO gaps for Arg and Hsp are 0.5–0.7 eV higher

than that for Lys. Furthermore, a decomposition of the orbitals for the lowest energy PBS-CT transitions (Fig. S8 of ESI†) of Lys, Arg and Hsp confirms a greater mixing of higher-lying filled and empty orbitals for the latter two systems. In fact, even the Lys system shows a greater degree of mixing of higher lying orbitals (LUMO+1) in the composition of the lowest energy PBS-CT transition relative to that of the anionic Glu and Asp chromophores. Furthermore, the spectral range for the PBS-CT transitions of the Arg and Hsp conformational ensembles overlaps with that of the localized backbone transitions (210–230 nm). In our calculations, we find that for 50% of Arg conformations and for 6% of His conformations, PBS-CT transitions are not the lowest energy excitations. Taken together, the above observations suggest that for cationic systems, the correlation between the excitation energy and the HOMO–LUMO gap should be significantly weaker relative to that for the anionic systems. Indeed, our analysis reveals significantly weak correlation between the ground state energy gap and $1/D_{CT}$ with the excitation wavelength for the lowest energy transitions extracted from 100 conformations of Lys, Arg and Hsp (Fig. S9 of ESI†).

Sensitivity of CT transitions to protein secondary structure

In this section, we present evidence for the sensitivity of the lowest energy PBS-CT transitions of charged amino acid residues to the protein secondary structure. The charged amino acid chromophores studied here were sampled from a 1 μ s MD trajectory of ubiquitin, a globular protein with 5 β -strands wrapped around a single α -helix. The charged amino acid chromophores sampled in our study therefore come from different secondary structure elements including β -strands, α -helices, and loop regions of the protein. Fig. 7 shows the variation in the lowest energy PBS-CT excitation wavelength as a function of the Ramachandran torsion angles (Φ_{i+1} and Ψ_i) defined in Fig. 2C for conformations of the 7 Lys (left panel) and 6 Glu (right panel) amino acids in ubiquitin sampled from the MD trajectory. The correspondence of these torsion angles with the Ramachandran angles (Φ_i , Ψ_i) for the Lys and Glu residues is shown in Fig. S2 and S4 of ESI.†

The Lys conformational ensemble in Fig. 7 shows a single diffuse cluster in the β -sheet region of the plot (labelled as β_{Lys}), 2 localized clusters in the α -helical region (labelled α_{1Lys} and α_{2Lys}), and one cluster (F_{Lys}) in the forbidden region ($\Phi_{i+1} = 60$ to 90 degrees, $\Psi_i = 120$ to 150 degrees) for Lys. By examining the excitation wavelength of the lowest energy PBS-CT transitions (colour code shown in the left panel of Fig. 7) of each conformation as a function of the torsion angles (Φ_{i+1} and Ψ_i), we find distinct wavelength ranges for each conformational cluster. For instance, conformations in the α_{1Lys} cluster show relatively high energy PBS-CT transitions below ~ 245 nm. On the other hand, the lowest wavelength transitions (~ 245 – 265 nm) come from conformations in the β_{Lys} cluster. Conformations in the α_{2Lys} and F_{Lys} clusters show intermediate wavelength range PBS-CT transitions between 245 – 255 nm. For the Glu conformational ensemble, we identify 2 distinct clusters of conformations in the β -sheet region (β_{1Glu} and β_{2Glu}), one cluster in the α -helical region (α_{Glu}), and 2 clusters in the forbidden region of the plot: $F1_{Glu}$ ($\Phi_{i+1} = -90$ to -60 degrees, $\Psi_i = 0$ to 30 degrees) and $F2_{Glu}$ ($\Phi_{i+1} = 60$ to 120 degrees, $\Psi_i = -60$ to 0 degrees). Upon examining the variation in the lowest energy excitation wavelengths of the PBS-CT transitions over the conformational ensemble (colour code shown in the right panel of Fig. 7),

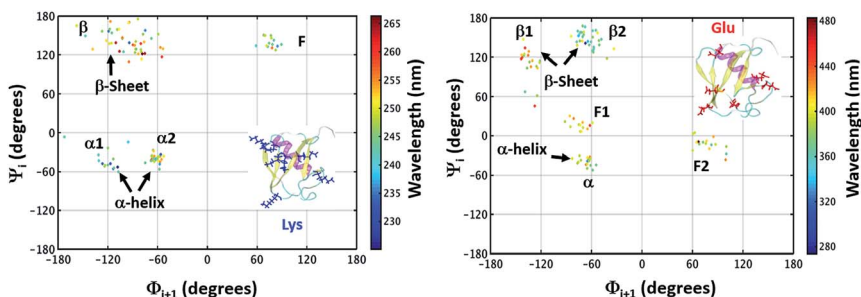


Fig. 7 Ramachandran dihedral angle (ϕ , ψ) maps for the Lys (left panel) and Glu (right panel) residues. Each point on the map represents a residue conformation sampled from 100 snapshots of a 1 μ s ubiquitin trajectory (see the Methods section). Each point is colour coded according to the excitation wavelength of the lowest energy PBS-CT transition for that conformation. The wavelength range is indicated by a colour scale bar to the right of each panel. Ubiquitin has 7 Lys and 6 Glu residues and we sample all of them in multiple conformations in our analysis. The distribution of Lys and Glu residues over the ubiquitin β -grasp fold sampled in our calculations is shown in the inset of each panel.

we again find a sensitivity of the excitation energies to the clustering of conformations. Conformations in $\beta_{1\text{Glu}}$ and $\beta_{2\text{Glu}}$ show the lowest (~ 280 – 400 nm) and highest (~ 360 – 480 nm) ranges of excitation wavelengths for the lowest energy PBS-CT transitions, respectively. On the other hand, conformations within α_{Glu} , $F_{1\text{Glu}}$ and $F_{2\text{Glu}}$ mostly exhibit transitions in the intermediate wavelength (~ 360 – 440 nm) range. Interestingly, one of the Glu conformations with the lowest excitation energy PBS-CT transition (~ 480 nm) lies in the $F_{2\text{Glu}}$ cluster.

To summarize, the data in this subsection show that the excitation wavelengths of the lowest energy PBS-CT transitions in charged amino acid monomers are sensitive to the secondary structure fold where the individual charged amino acids are located. These results motivate more detailed investigations of the spectra of interacting charged amino acid conformations sampled from different secondary structure elements in protein folds with the goal of connecting ProCharTS spectral features to the underlying protein structure and dynamics.

Conclusions

In this manuscript, we provide a detailed computational study and comparison of new optical (UV-vis) CT transitions in charged amino acid monomers. Previously, we had shown that such CT transitions from Lys and Glu amino acids could produce significant absorption in proteins that are rich in these charged amino acids well beyond 250 nm.¹⁴ Here we computationally examine the UV-vis spectra of six other charged amino acid chromophores, Arg, Hsp, Asp and phosphorylated sidechains of Ser, Thr and Tyr (pSer, pThr, and pTyr), and compare them with those of Lys and Glu. For all eight amino acids we carried out electronic structure and TDDFT calculations on multiple conformations sampled from a classical MD trajectory and structures from the protein data bank. Our analysis reveals that the excess charge on the sidechains of charged amino acids polarizes the frontier orbitals to create well defined electronic D–B–A complexes in these systems. TDDFT calculations show that the UV-vis spectra of all charged amino acid

chromophores is dominated by CT transitions and that the lowest energy transitions are PBS-CT transitions. The direction of PBS-CT transitions in the anionic systems is opposite (sidechain to peptide backbone) to that in the cationic systems (peptide backbone to sidechain) and is well described in terms of the ground state dipole moments of these chromophores. We find that the lowest energy PBS-CT in the anionic Glu and Asp transitions predominantly involve the HOMO and LUMO and consequently the red edge (lowest excitation wavelength) of the PBS-CT transitions in these systems correlates strongly with the ground state HOMO–LUMO gap. In contrast, for the cationic (Lys, Arg and Hsp) and other anionic (pSer, pThr and pTyr) species, we find that higher lying filled and empty orbitals contribute to the lowest energy PBS-CT transitions. Thus, for these systems the red-edge does not correlate significantly with the ground state HOMO–LUMO gap. Finally, we presented evidence for the sensitivity of the UV-vis spectral range associated with PBS-CT transitions to the protein secondary structure fold. These results advance our understanding of the range of CT transitions present in non-aromatic amino acids. The PBS-CT transitions studied here are sensitive to sidechain interactions and the protein fold¹⁴ and produce novel spectra termed ProCharTS which can be used to probe the structure and dynamics of proteins that are rich in charged amino acids. The systematic investigation of charged amino acids presented here builds a foundation for understanding distinct features in the ProCharTS absorption spectra of proteins.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank the Tata Institute of Fundamental Research (Department of Atomic Energy, India) for their financial support.

References

- 1 M. C. Otey and J. P. Greenstein, Studies on polycysteine peptides and proteins. II. Apparent dissociation constants, and ultraviolet and infrared absorption spectra of isomeric cystinylcystine peptides, *Arch. Biochem. Biophys.*, 1954, **53**, 501–513.
- 2 C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part 2: Techniques for the study of biological structure and function*, W. H. Freeman and Company, New York, 1980.
- 3 G. D. Fasman, *Practical Handbook of Biochemistry and Molecular Biology*, CRC Press, New York, 1992.
- 4 H. D. Hunt and W. T. Simpson, Spectra of Simple Amides in the Vacuum Ultraviolet, *J. Am. Chem. Soc.*, 1953, **75**, 4540–4543.
- 5 K. Rosenheck and P. Doty, The far ultraviolet absorption spectra of polypeptide and protein solutions and their dependence on conformation, *Proc. Natl. Acad. Sci. U. S. A.*, 1961, **47**, 1775–1785.
- 6 D. B. Wetlaufer, Ultraviolet spectra Of Proteins and Amino Acids, in *Advances in Protein Chemistry*, 1963, vol. 17, pp 303–390.

- 7 J. S. Ham and J. R. Platt, Far U.V. spectra of peptides, *J. Chem. Phys.*, 1952, **20**, 335–336.
- 8 M. T. Oakley and J. D. Hirst, Charge-transfer transitions in protein circular dichroism calculations, *J. Am. Chem. Soc.*, 2006, **128**, 12414–12415.
- 9 L. Serrano-Andres and M. P. Fulscher, Theoretical study of the electronic spectroscopy of peptides. III. Charge-transfer transitions in polypeptides, *J. Am. Chem. Soc.*, 1998, **120**, 10912–10920.
- 10 E. I. Solomon, D. E. Heppner, E. M. Johnston, J. W. Ginsbach, J. Cirera, M. Qayyum, M. T. Kieber-Emmons, C. H. Kjaergaard, R. G. Hadt and L. Tian, Copper Active Sites in Biology, *Chem. Rev.*, 2014, **114**, 3659–3853.
- 11 D. L. Ma, M. D. Wang, C. F. Liu, X. M. Miao, T. S. Kang and C. H. Leung, Metal complexes for the detection of disease-related protein biomarkers, *Coord. Chem. Rev.*, 2016, **324**, 90–105.
- 12 L. Serrano-Andres and M. P. Fulscher, Charge transfer transitions in neutral and ionic polypeptides: A theoretical study, *J. Phys. Chem. B*, 2001, **105**, 9323–9330.
- 13 A. T. B. Gilbert and J. D. Hirst, Charge-transfer transitions in protein circular dichroism spectra, *J. Mol. Struct.: THEOCHEM*, 2004, **675**, 53–60.
- 14 S. Prasad, I. Mandal, S. Singh, A. Paul, B. Mandal, R. Venkatramani and R. Swaminathan, Near UV-Visible electronic absorption originating from charged amino acids in a monomeric protein, *Chem. Sci.*, 2017, **8**, 5416–5433.
- 15 D. Pinotsi, L. Grisanti, P. Mahou, R. Gebauer, C. F. Kaminski, A. Hassanali and G. S. Kaminski Schierle, Proton Transfer and Structure-Specific Fluorescence in Hydrogen Bond-Rich Protein Structures, *J. Am. Chem. Soc.*, 2016, **138**, 3046–3057.
- 16 P. Guptasarma, Solution-state characteristics of the ultraviolet A-induced visible fluorescence from proteins, *Arch. Biochem. Biophys.*, 2008, **478**, 127–129.
- 17 A. Shukla, S. Mukherjee, S. Sharma, V. Agrawal, K. V. Radha Kishan and P. A. Guptasarma, novel UV laser-induced visible blue radiation from protein crystals and aggregates: scattering artifacts or fluorescence transitions of peptide electrons delocalized through hydrogen bonding?, *Arch. Biochem. Biophys.*, 2004, **428**, 144–153.
- 18 S. Jones, P. van Heyningen, H. M. Berman and J. M. Thornton, Protein–DNA interactions: A structural analysis, *J. Mol. Biol.*, 1999, **287**, 877–896.
- 19 R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, Classification of intrinsically disordered regions and proteins, *Chem. Rev.*, 2014, **114**, 6589–6631.
- 20 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1**–2, 19–25.
- 21 R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles, *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.

- 22 W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 23 T. Yanai, D. P. Tew and N. C. Handy, A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP), *Chem. Phys. Lett.*, 2004, **393**, 51–57.
- 24 M. J. Frisch, G. W. Trucks, 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. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. 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, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, Gaussian, Inc., Wallingford, CT, USA, 2009.
- 25 T. Lu and F. Chen, Multiwfn: a multifunctional wavefunction analyzer, *J. Comput. Chem.*, 2012, **33**, 580–592.
- 26 R. Dennington, T. Keith and N. J. Millam, *GaussView, Version 5*, Semichem Inc., Shawnee Mission, KS, 2009.
- 27 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, The Protein Data Bank, *Nucleic Acids Res.*, 2000, **28**, 235–242.
- 28 C. A. Guido, P. Cortona, B. Mennucci and C. Adamo, On the Metric of Charge Transfer Molecular Excitations: A Simple Chemical Descriptor, *J. Chem. Theory Comput.*, 2013, **9**, 3118–3126.
- 29 T. Le Bahers, C. Adamo and I. Ciofini, A Qualitative Index of Spatial Extent in Charge-Transfer Excitations, *J. Chem. Theory Comput.*, 2011, **7**, 2498–2506.
- 30 D. M. Rogers, N. A. Besley, P. O'Shea and J. D. Hirst, Modeling the absorption spectrum of tryptophan in proteins, *J. Phys. Chem. B*, 2005, **109**, 23061–23069.
- 31 P. Stepanek and P. Bour, Multi-scale modeling of electronic spectra of three aromatic amino acids: importance of conformational averaging and explicit solute-solvent interactions, *Phys. Chem. Chem. Phys.*, 2014, **16**, 20639–20649.
- 32 D. Jacquemin, E. A. Perpète, G. E. Scuseria, I. Ciofini and C. Adamo, TD-DFT performance for the visible absorption spectra of organic dyes: Conventional versus long-range hybrids, *J. Chem. Theory Comput.*, 2008, **4**, 123–135.
- 33 M. J. G. Peach, P. Benfield, T. Helgaker and D. J. Tozer, Excitation energies in density functional theory: An evaluation and a diagnostic test, *J. Chem. Phys.*, 2008, **128**, 044118.
- 34 M. Campetella, F. Maschietto, M. J. Frisch, G. Scalmani, I. Ciofini and C. Adamo, Charge transfer excitations in TDDFT: A ghost-hunter index, *J. Comput. Chem.*, 2017, **38**, 2151–2156.
- 35 M. N. Paddon-Row, Investigating long-range electron-transfer processes with rigid, covalently linked donor-(norbornylogous bridge)-acceptor systems, *Acc. Chem. Res.*, 1994, **27**, 18–25.

- 36 O. S. Wenger, How Donor–Bridge–Acceptor Energetics Influence Electron Tunneling Dynamics and Their Distance Dependences, *Acc. Chem. Res.*, 2011, **44**, 25–35.
- 37 A. Dreuw, J. L. Weisman and M. Head-Gordon, Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange, *J. Chem. Phys.*, 2003, **119**, 2943–2946.