Journal Name

ARTICLE

The Structures and Properties of Anionic Tryptophan Complexes

Christian Ieritano,[a] Joshua Featherstone,[a] Patrick J. J. Carr,[a] Rick A. Marta,[a] Estelle Loire,[b] Terrance B. McMahon,[a] and W. Scott Hopkins*[a]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

The physicochemical properties of $[Trp_n–H]^-$ and $[Trp_n\cdots Cl]^-$ (n = 1,2) have been investigated in a combined computational and experimental infrared multiple dissociation (IRMPD) study. IRMPD spectra within the 850 - 1900 cm^{-1} region indicate that deprotonation is localized on the carboxylic acid moiety in $[Trp_n–H]^-$ clusters. A combination of hydrogen bonding and higher order charge-quadrupole interactions appear to influence cluster geometries for all investigated system. Calculated global minimum and low energy geometries of $[Trp\cdots Cl]^-$ and $[Trp_2\cdots Cl]^-$ clusters favor coordination of the halide by the indole NH. $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$ exhibit additional π-π interactions between the heterocyclic side chains.

Introduction

The ability of enzymes to adopt the dynamic structures necessary for their biological activity is associated with their tertiary structure,[1-8] which arises from non-covalent interactions between amino acid residues. Since the strength of individual non-covalent interactions range from 4 – 20 kJ mol^{-1},[1,2] introduction of more strongly interacting molecules such as substrates, cofactors, or inhibitors can easily disrupt these transient bonds. It is this conformational flexibility in catalytically active regions of enzymes that ultimately facilitates catalysis.

Perhaps the most versatile amino acid residues which influence tertiary structure are those that contain aromatic side chains (*i.e.*, His, Phe, Tyr, Trp). Specifically, these species can coordinate ions though charge-quadrupole interactions in addition to conventional hydrogen bonding. Recently, the interaction between anions and aromatic systems has attracted interest.[3–9] The anion-π interaction, which has been shown to be nearly isoenergetic with the more familiar cation-π interaction,[10–13] originates from Coulombic attraction between a permanent negative charge and the electron deficient region of the quadrupole moment associated with the heterocycle.[14,15] As such, anions will preferentially occupy sites in the equatorial plane of the heterocyclic side chains. However, introduction of electron withdrawing groups can promote interactions between the anion and the axial face of the aromatic ring.[12,16,17] A combination of these motifs are commonly employed in the development of supramolecular assemblies,[3–6] anion sensing,[7] and have been identified in biological and synthetic anion channels.[8,9] Consequently, anion-π interactions have been incorporated into drugs developed to treat diseases that affect anion transport.[18–20]

Since the first identification of anion-π interactions by X-ray crystallography in 2004,[21] there has been a significant amount of research conducted to investigate their role in biological systems. Protein Data Bank (PDB) searches have identified attractive, edgewise interactions between negatively charged residues (*i.e.*, Asp, Glu) and biological heterocycles that play prominent structural and/or catalytic roles, particularly in serine proteases.[22–30] Anion-π interactions have also been identified as the mechanistic controller behind the voltage-dependant response of prestin, a motor protein that is essential for sensitive hearing in mammalian species.[31]

While these interactions have been identified in condensed phase biological systems, there has been limited work done to investigate isolated, biologically-relevant interactions involving the aromatic amino acids, particularly with anions. An IRMPD study by Burt and coworkers investigated the structural effects of anion-π interactions in Phe analogs, which revealed enhanced binding energies to the anion (*e.g.*, Cl^-, Br^-) owing to the influence of charge-quadrupole interactions.[12] With regard to Trp residues, only studies investigating the binding motifs in cationic alkali[32] and divalent transition and alkaline earth[33] metals in monomeric and dimeric complexes have been completed. These works illustrated the prominence of cation-π interactions between the metal ion and side-chain indole heterocycle. Furthermore, an IRMPD study of protonated Trp monomers, where protonation was localized on the amine, revealed that the protonated amine preferentially coordinated the backbone COOH and was involved in a cation-π interaction with the indole side chain.[34] Similar motifs were observed in an infrared photodissociation (IRPD) study of protonated Trp dimers, where certain low energy conformers exhibited cation-π binding motifs.[35]

[a] *Department of Chemistry, University of Waterloo, Waterloo, ON, N2L 3G1, Canada. E-mail: shopkins@uwaterloo.ca*
[b] *Laboratoire Chimie Physique – CLIO, Bâtiment 201, Porte 2, Campus Universitaire d'Orsay, 91405, France*

†Electronic Supplementary Information (ESI) available: $[Trp_n–H]^-$, $[Trp_n\cdots Cl]^-$ (n = 1,2) *XYZ* coordinates, thermochemical data, scaled harmonic, and anharmonically corrected spectra are available free of charge. See DOI: 10.1039/x0xx00000x

To the best of our knowledge, only a single, condensed phase study has addressed anion-π interactions in Trp-containing species.[36] This work, which involved a search of the PDB, identified numerous protein structures that exhibit close contacts between Trp residues and biologically prominent anions, such as chloride and phosphates.[36] To gain further insight into the interactions exhibited by these species, we have studied the structures and properties of anionic and chloride-bound Trp monomers and dimers in a joint experimental and computational study. Experimental outcomes are supported by detailed basin-hopping molecular dynamics searches of the cluster potential energy surface (PES) and high-level quantum-chemical calculations which provide further insight into the structure of anionic Trp clusters.

Experimental

IRMPD spectra of $[Trp_n–H]^-$ and $[Trp_n\cdots Cl]^-$ (n= 1,2) were recorded at the Centre Laser Infrarouge d'Orsay (CLIO) free electron laser (FEL) facility at the University of Paris XI.[37,38] Details of the experimental apparatus have been previously described.[12,39,40] Briefly, monomeric and dimeric Trp clusters were produced by electrospray ionization (ESI) in negative mode of *ca.* 100 μM solutions of L-tryptophan in methanol/water (50/50 vol%) containing *ca.* 5 μM of NaCl. L-tryptophan and NaCl were obtained from Sigma Aldrich (Oakville, ON, Canada) and used without further purification. The nascent clusters produced were transferred to a Bruker Esquire 3000+ ion trap mass spectrometer, where they were mass selected and subsequently irradiated by the tunable output of the FEL. IRMPD spectra were acquired by irradiating the trapped ions for 400 ms at 5 cm^{-1} intervals in the 850 – 1900 cm^{-1} range. The IR-FEL output was delivered to the trapped ions as a series of 9 ms macropulses emitted with a repetition rate of 25 Hz. Each macropulse consisted of 600 picosecond micropulses. The average IR power for these experiments varied approximately quadratically as a function of photon wavenumber between 800–1200 mW, with the peak power being delivered near the midpoint of the spectrum. Vibrational spectra were generated by recording the fragmentation efficiency of the Trp clusters as a function of FEL wavenumber.

Computational methods

In order to map the PES of the Trp clusters, a custom-written basin-hopping (BH)[40–45] code that is interfaced with Gaussian 09[46] was employed. Clusters were modelled using the AMBER molecular mechanics force field, which utilized partial charges calculated at the B3LYP/6-31+G(d,p) level of theory.[47–49] Each of the $[Trp_n–H]^-$ and $[Trp_n\cdots Cl]^-$ (n = 1,2) were investigated computationally. For each random structural perturbation in the BH search of chloride containing clusters, the charge carrying species was given a random translational step of −0.7 Å ≤ η ≤ 0.7 Å in each of the *X*, *Y*, and *Z* directions. For the $[Trp–H]^-$ and $[Trp\cdots Cl]^-$ species, each of the dihedral angles of the Trp residue were assigned a random rotation of −5° ≤ Φ ≤ 5°. The same dihedral rotation was applied for the $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$ clusters. However, in this case one of the Trp residues was assigned a random translational step of −0.7 Å ≤ η ≤ 0.7 Å in each of the *X*, *Y*, and *Z* directions as well as a random rotation of −5° ≤ θ ≤ 5° about its *x*, *y*, and *z* body-fixed axes.

In total, *ca.* 30 000 structures were sampled by the BH algorithm for each cluster. The BH routine identified more than 500 unique conformations for the $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$ clusters. The 50 lowest energy conformers from each search were pre-optimized at the HF/6-31G level of theory, and unique isomers from this step were then carried forward for treatment using density functional theory (DFT) at the B3LYP/6-311++G(d,p) level of theory employing a correction for empirical dispersion.[48–50] This level of theory has been used successfully in studies of similar species.[13,32,42,51–54] Normal mode analysis was conducted to verify that each isomer corresponded to a minimum on the PES. This also served to calculate the harmonic vibrational spectrum for each structure, which was used in the assignment of the experimental IRMPD spectrum. Calculated spectra are broadened with a 4 cm^{-1} Gaussian linewidth, and scaled by 0.9679.[55] Single-point energy calculations were subsequently conducted on the B3LYP optimized geometries at the MP2/6-311++g(d,p) level of theory. MP2 thermochemistry was determined by combining single point electronic energies with B3LYP thermal corrections determined at 298 K, and is reported as MP2/6-311++G(d,p)//B3LYP/6-311++G(d,p). Unique cluster structures were assigned based on Gibbs corrected energy and geometry. In the case of $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$, partial anharmonic analysis of normal modes was conducted for isomers involving shared proton motifs (see discussion section for further details). Isomers exhibiting unique binding motifs were further optimized at the MP2/6-311++G(d,p)//B3LYP/6-311++G(d,p) level of theory using the Polarizable Continuum Model (PCM) to model isomer energetics using implicit solvation in water.[56] Molecular graphics were generated using the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).[57] Atomic *XYZ* coordinates and pertinent thermochemical information of all calculated isomers are included in the supporting information.

Results and Discussion

$[Trp–H]^-$

Based on chemical intuition, $[Trp–H]^-$ is likely to be formed as the carboxylate anion, as the carboxylic acid moiety is the most acidic site present in the molecule. This is supported by the IRMPD spectrum of $[Trp–H]^-$, which yields fragments consistent with collision-induced dissociation (CID) studies of anionic Trp clusters,[58] where the loss of CO_2 (*m/z* = 44) predominates. Additional product channels associated with the loss of CO_2 + NH_3 (*m/z* 61) and the amino acid backbone (*m/z* 88) are also observed. The IRMPD spectrum of $[Trp–H]^-$ is plotted in Figure 1. Line widths of *ca.* 50 cm^{-1} are larger than the typical CLIO bandwidth of 25 cm^{-1}, and suggests that at least two

conformers are present in the $[Trp–H]^-$ ensemble population.[37] This is supported by our computational study, which identified a total of five conformers of $[Trp–H]^-$ that lie within 10.7 kJ mol^{-1} of one another. The calculated spectra of isomers 1–5 were combined and weighted according to a Maxwell-Boltzmann distribution, where isomer energies are reported relative to isomer 1 (set to 0 kJ mol^{-1}). Further details of this methodology and individual calculated IR spectra corresponding to each isomer are available in the supporting information.

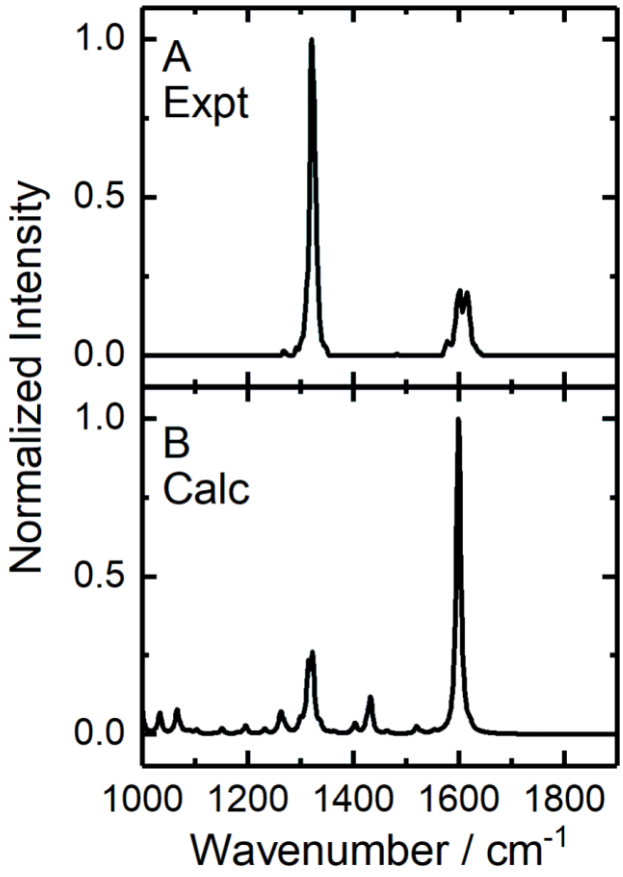

Figure 1. (A) The experimental IRMPD spectrum of $[Trp–H]^-$ and **(B)** Boltzmann-weighted, calculated IR spectra of isomers 1-5. Calculations were conducted at the B3LYP/6-311++G(d,p) level of theory.

Based on the agreement between experimental and predicted spectra, it is possible that all the calculated isomers contribute to the IRMPD spectrum. We can confidently assign the vibrational bands in the IRMPD spectrum to normal modes in the $[Trp–H]^-$ anion. The band at 1600 cm^{-1} corresponds to the simultaneous C=O stretching and NH_2 scissoring of the backbone carboxylate and amine moieties, which provides further evidence that the carboxylic acid is the site of deprotonation. The peak at 1320 cm^{-1} is consistent with CCH/CNH bending modes of the aromatic indole moiety and methylene CH_2 wagging. Note that it is not uncommon in IRMPD spectra to observe intensities dissimilar to those predicted by electronic structure calculations.[52,59,60] Gaussian calculates the absorption cross-section of the molecules, whereas IRMPD monitors fragmentation efficiency (which is determined by absorption cross section and coupling efficiency to dissociative channels). Thus, one cannot directly compare intensities between IRMPD and calculation.

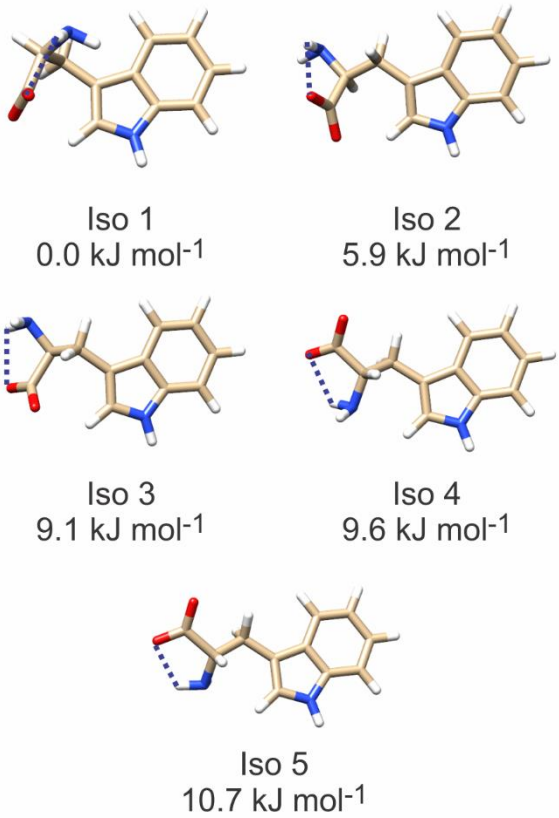

Figure 2. The calculated structures of $[Trp–H]^-$. Calculations were performed at the MP2/6-311++G(d,p)//B3LYP/6-311++G(d,p) level of theory. Electronic energies are Gibbs corrected.

Figure 2 shows the optimized structures of the five lowest energy isomers $[Trp–H]^-$. Structurally, DFT optimized unique isomers identified by the BH routine differ only in the orientation of the amino acid backbone relative to the indole ring. All conformers contain the intramolecular hydrogen bond between the backbone NH_2 and COO^-; stabilization of the global minimum geometry relative to higher energy conformers is apparently accomplished through charge-quadrupole interactions between the carboxylate moiety and the indole side chain. Assessing the interactions present in higher energy structures allows us to gauge their degree of stabilization. These include interactions between the backbone amine proton and the π-cloud of the indole ring, as well as the carboxylate oxygen and the electron deficient equatorial region of the aromatic side chain. For example, isomer 2, which was calculated to be 5.9 kJ mol^{-1} higher in energy relative to the global minimum, does not contain the interaction between the amine NH_2 and π-cloud. In isomer 3, the carboxylate moiety is rotated such that its orientation with respect to the indole ring is skewed, weakening the charge-quadrupole interaction. Interestingly, this migration of the carboxylate anion out of the plane of the

indole ring (isomer 3) raises the energy of the ion by 3.2 kJ mol^{-1}. This highlights the potency of charge-quadrupole interactions and their potential in stabilizing the orientations of Trp moieties in larger species. Removal of the carboxylate from the indole plane in isomers 4 and 5 is (partially) stabilized by the introduction of an interaction between the lone pair on the amine and equatorial plane of the indole heterocycle.

[Trp···Cl]$^-$

Our computational study of [Trp···Cl]$^-$ identified a total of 35 unique isomers (both charge solvated and zwitterionic) within 40.0 kJ mol^{-1} of the global minimum. Calculated geometries often share binding characteristics that can be grouped into structural motifs. The calculated geometries of [Trp···Cl]$^-$ fall within 6 structural motifs. The lowest energy isomer found within each motif is shown in Figure 3, where isomers are reported in order of increasing energy from the global minimum (*i.e.*, isomer 1).

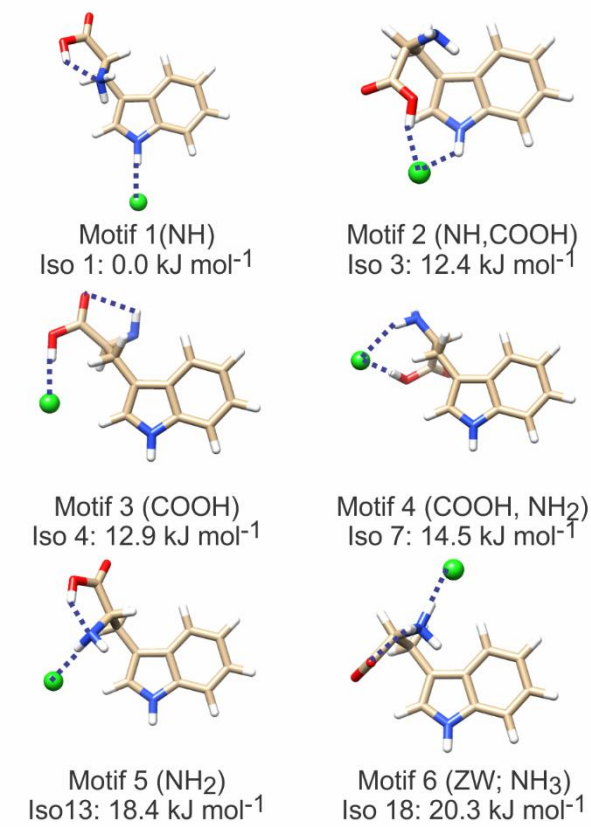

Figure 3. Binding motifs of [Trp···Cl]$^-$ as determined by calculations at the MP2/6-311++G(d,p)//B3LYP/6-311++G(d,p) level of theory. Electronic energies are Gibbs corrected.

The indole side chain, which is inherently an electron rich aromatic system, contains a hydrogen bond donor. As such, the calculated global minimum geometry (motif 1) and higher energy isomers (motif 2) coordinates the halide by a hydrogen bond through the indole NH. An intramolecular hydrogen bond also exists between the NH_2 and COOH moieties of the Trp backbone, with an additional orientation of the NH_2 protons towards the π-cloud of the indole side chain. This secondary interaction was found to be prominent in the [Trp–H]$^-$ anion, and also manifests itself within motifs 2, 5 and 6 of [Trp···Cl]$^-$. Isomers found within motifs 2 and 3 maintain the position of the halide within the equatorial plane of the indole side chain, but modify halide coordination to involve the backbone COOH. Isomers within motif 4 adopt a nearly identical configuration to those found within motif 3 and coordinate the halide by the backbone COOH and NH_2 while slightly skewing the position of the halide such that it is no longer in the same plane as the heterocyclic side chain. Isomers found within motif 5 adopt typical, charge-solvated geometries, coordinating the anion solely through the backbone NH_2. Isomers found within motif 6 adopt a zwitterionic motif, where charge separation is stabilized through coordination of the halide to the cationic backbone amine.

Figure 4 shows the experimental IRMPD spectrum for [Trp···Cl]$^-$. The appearance of two distinct C=O stretching frequencies at 1777 cm^{-1} and 1716 cm^{-1}, in conjunction with peak widths of *ca.* 50 cm^{-1} indicate the presence of multiple isomers/conformers in the [Trp···Cl]$^-$ population.[37] The experimental spectrum is well modeled by isomers from motifs 1–3; there is no evidence of higher energy zwitterionic species. Given the difference of 12.4 kJ mol^{-1} between motifs 1 and 2, and the fact that the calculated position of the C=O band for motif 2 is in good agreement with the band observed at 1716 cm^{-1} (highlighted green in Figure 4), it seems likely that kinetic trapping of higher energy motifs is occurring. This is supported in part by PCM calculations in H_2O (see supporting information), which decreases the energetic separation of the two motifs to 7.1 kJ mol^{-1}, suggesting that higher energy motifs may be populated in the solution phase.

Owing to the energetic spacing between motifs 1, 2, and 3, a Boltzmann-weighted analysis of calculated [Trp···Cl]$^-$ spectra does not capture the band observed at 1716 cm^{-1}. Consequently, we provide the spectra of the representative isomers (1, 3, and 4) for comparison in Figure 4. Given the similar binding characteristics of isomers found within motifs 1-5, each of these isomers exhibits similar calculated IR spectra. Calculations indicate the band at 1777 cm^{-1} corresponds to the C=O stretch of an uncomplexed COOH (motif 1), while the lower frequency mode is the C=O stretch of the COOH moiety coordinated to the chloride anion (motifs 2-4). Further justification for the presence of several low energy, isomeric forms of [Trp···Cl]$^-$ arises from the broad peak centered at 1350 cm^{-1} (assigned as COH bending), which indicates that the OH moiety exists in multiple chemical environments. Coordination of the chloride anion by the COOH moiety results in a red-shifting of the mode compared to the global minimum. Calculations further suggest that this intense band consists of CCH/CNH bending modes associated with the indole heterocycle. Weaker bands at both 1215 cm^{-1} and 1070 cm^{-1} are assigned to methylene/methine CH_2/CCH bending, and a combination of methylene CH_2 as well as further indole CCH wagging/bending modes, respectively.

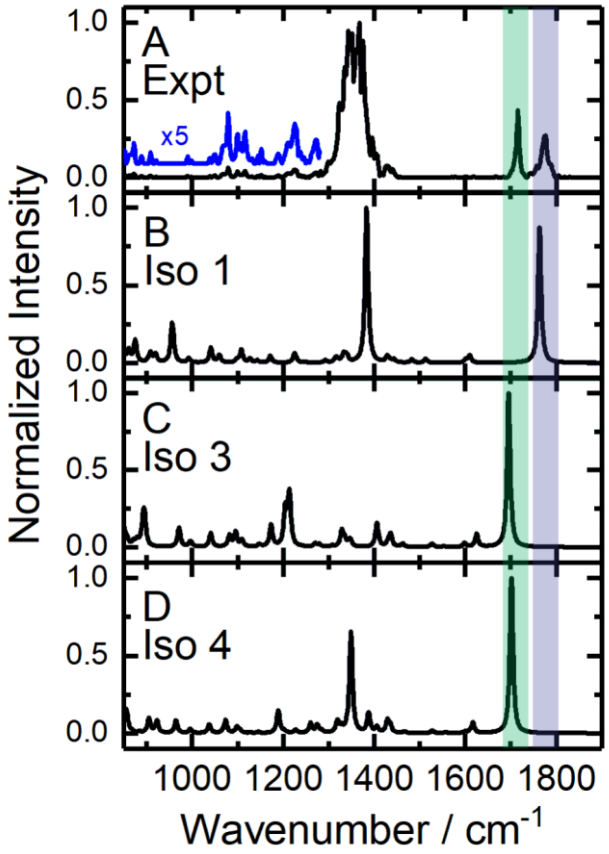

Figure 4. (A) The experimental IRMPD spectrum of $[Trp\cdots Cl]^-$ and the calculated harmonic spectra of the **(B)** global minimum (motif 1), **(C)** isomer 3 (motif 2, 12.4 kJ mol^{-1}), and **(D)** isomer 4 (motif 3, 12.9 kJ mol^{-1}). Calculations were conducted at the B3LYP/6-311++G(d,p) level of theory.

$[Trp_2–H]^-$

Mapping of the PES of $[Trp_2–H]^-$ identified a total of 37 unique conformers. The lowest energy structure for each of the seven lowest energy binding motifs are shown in Figure 5. Isomers are ordered by increasing relative energy from the global minimum (*i.e.*, isomer 1).

In general, the calculated binding motifs of $[Trp_2–H]^-$ are influenced by the Trp side chain; the carboxylate moiety is either coordinated by the indole NH (motifs 1, 3 and 4, and 5), or lies within the equatorial plane of the heterocyclic side chain (motifs 2, 6, and 7). Motif 2 comprises a rather interesting geometry involving a shared proton between the two adjacent carboxylic acids. As a result, the cluster adopts a highly symmetric geometry of near C_2 symmetry (see supporting information for 3D coordinates). All low energy motifs exhibit additional π-π interactions that allows for more intricate cluster geometries. Isomers within motif 2 exhibit a nearly parallel, offset orientation of the indole side chains, while other conformers favour a skewed, edge-to-face stacking interaction in addition to unique binding motifs of the carboxylate moiety.

The experimentally observed IRMPD spectrum of $[Trp_2–H]^-$ is plotted in Figure 6A, and the Boltzmann-weighted spectra for isomers 1-4 are shown in Figure 6B. Due to the prominence of shared proton vibrational modes within low energy $[Trp_2–H]^-$ cluster geometries, it was necessary to calculate anharmonic corrections to the harmonic spectra. To this end, we employed an approach whereby large amplitude shared proton modes were treated with a full anharmonic correction, and all other modes were modeled with a harmonic potential. Anharmonic corrections were employed on isomers 1 and 2. Since isomers 3 and 4 belong to the same structural motif as isomer 1 (motif 1), conducting anharmonic calculations would be redundant as the scaling factor of 0.9679 was able to reproduce the anharmonically corrected spectrum corresponding to isomer 1 (see supporting information for further details).

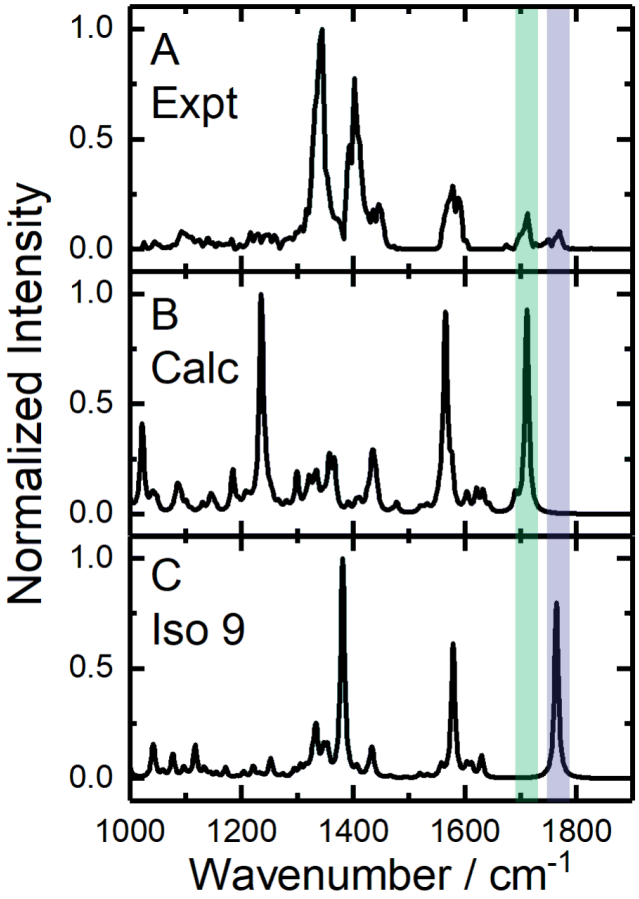

Figure 6. (A) The experimental IRMPD spectrum of $[Trp_2–H]^-$ and **(B)** the computed IR spectrum of $[Trp_2–H]^-$. The computed spectrum includes a Boltzmann-weighted convolution of IR spectra for isomers 1-4, and anharmonic frequency corrections for large amplitude shared proton modes in isomers 1 and 2. **(C)** The calculated harmonic IR spectrum for isomer 30. Calculations were conducted at the B3LYP/6-311++G(d,p) level of theory.

The anharmonically corrected Boltzmann-weighted convolution of the spectra for isomers 1-4 provides a reasonable reproduction of the experimental IRMPD spectrum. Based on the calculations, we can assign the major vibrational bands. The bands at 1770 cm^{-1} and 1715 cm^{-1} are associated with C=O stretches in carboxylic acid moieties. The band at 1580 cm^{-1} is assigned to simultaneous carboxylate C=O stretching and NH_2 scissoring. The band at 1405 cm^{-1} arises from indole CNH/CCH bending modes. Additional bands at 1340 cm^{-1} correspond to

the combined CH bend of the methine moiety and CH_2 bend of the beta methylene carbon, each with a substantial displacement of proton shared between the carboxylate groups. The band at 1022 cm^{-1} is assigned to the methine CH wag, methylene CH_2 twist, and displacement of the shared proton.

As was observed with the monomeric species, a peak width in the IRMPD spectrum of *ca.* 50 – 75 cm^{-1} indicates that multiple isomers/conformers are present in the sample population.[37] This is further supported by the two bands corresponding to the carbonyl stretch of a COOH moiety at 1770 cm^{-1}, and 1715 cm^{-1}, which suggests that at least two isomers contribute to the overall gas-phase population. The calculated global minimum conformer, which coordinates the deprotonated carboxylate with the COOH of the adjacent Trp, exhibits a single carboxylate stretching band at 1577 cm^{-1} due to symmetry about the shared proton. The C=O stretching mode of the carboxylate moiety in isomers 2 and 3 are calculated to occur at 1565 cm^{-1} and 1591 cm^{-1}, respectively. However, isomers 2 and 3 also contain intramolecular H-bonding interactions between the COO^- moiety and the indole NH and backbone amine. This results in C=O stretching vibrations associated with the H-bonded COOH at 1711 cm^{-1} and 1709 cm^{-1} (highlighted in green in Figure 6). The COH bending mode of isomer 1 is calculated to lie at 1394 cm^{-1}, while the COH bending modes of isomers two and 3 (both of which involve a bidentate coordination of the carboxylate) are predicted to occur at 1234 cm^{-1} and 1340 cm^{-1}, respectively.

In total, there are 13 isomers calculated to lie within 20 kJ mol^{-1} of the global minimum, the spectra of which all agree with experiment. However, the carbonyl stretching band observed at 1770 cm^{-1} is only reproduced by isomer 9 (motif 5; highlighted in blue in Figure 6). Observation of such high energy species (22.0 kJ mol^{-1}) suggests the kinetic trapping is occurring in the ESI process.[61,62] Implicit solvation in water decreases this energetic gap to 9.9 kJ mol^{-1}, which supports the trapping of higher energy species (see supporting information).

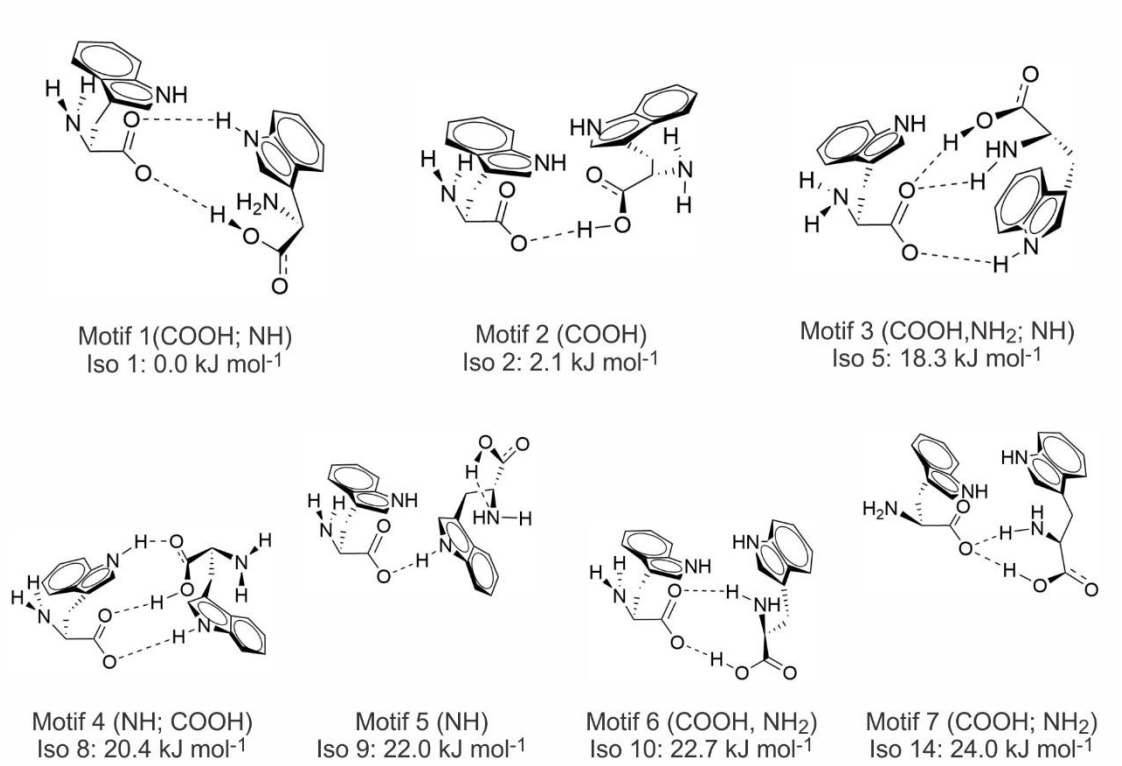

Figure 5. Low energy binding motifs of $[Trp_2–H]^-$. Geometries correspond to the lowest energy isomer found within a particular binding motif, as calculated at the MP2/6-311++G(d,p)//B3LYP/6-311++G(d,p) level of theory. Electronic energies are Gibbs corrected.

$[Trp_2\cdots Cl]^-$

Several structural motifs were identified in the computational study of $[Trp_2\cdots Cl]^-$. The lowest energy isomer in each of the four lowest energy binding motifs are shown in Figure 7. In total, 40 unique isomers were found within 88.1 kJ mol^{-1} of the global minimum. Isomers are ordered by increasing relative energy relative to the global minimum (*i.e.*, isomer 1). All low-energy $[Trp_2\cdots Cl]^-$ binding motifs exhibit coordination of the halide through at least one indole NH moiety, indicating that H-bonding with the heterocycle provides stabilization for the complex. Moreover, nearly all binding motifs exhibit skewed edge-to-face or staggered π-stacking of the Trp side chains. Within isomers that adopt identical halide coordination schemes (*e.g.*, motifs 1 and 3, and 2 and 4), structural differences stem from the variety of intra/intermolecular hydrogen bonding possible between the Trp residues.

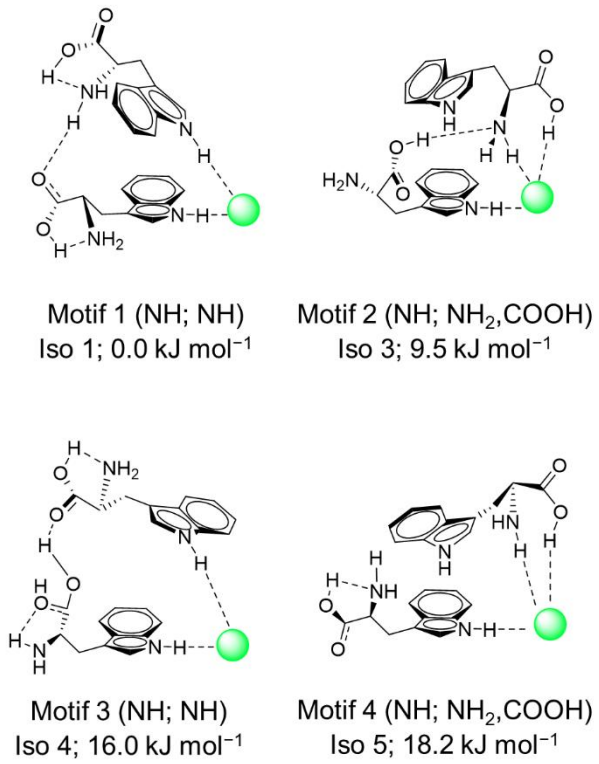

Figure 7. Low energy binding motifs of $[Trp_2\cdots Cl]^-$. Geometries correspond to the lowest energy isomer found within a particular binding motif, as calculated at the MP2/6-311++G(d,p)//B3LYP/6-311++G(d,p) level of theory. Electronic energies are Gibbs corrected.

The experimentally observed IRMPD spectrum of $[Trp_2\cdots Cl]^-$ is plotted in Figure 8A. The Boltzmann-weighted convolution of the calculated spectra of the lowest energy isomers corresponding to motifs 1 and 2 are plotted in Figure 8B. The prominence of vibrational modes involving shared protons in low energy isomers compelled us to employ partial anharmonic corrections to calculated vibrational spectra using the methodology described above. Details concerning the anharmonic correction can be found in the supporting information.

The appearance of two C=O stretching frequencies and the fact that peak widths of *ca.* 50 cm^{-1} are observed indicates the presence of multiple isomers in the gas-phase population.[37] The band at 1777 cm^{-1} corresponds to the carbonyl stretch of a coordinated COOH moiety. Coordination could be either with the halide or with the adjacent amine group via an intramolecular hydrogen bond. The band at 1745 cm^{-1}, which is also associated with a C=O stretching motion, provides further evidence of the multiple coordination environments for the carboxylic acid functional groups. This is reinforced by the broad absorption corresponding to the COH bending mode, which spans the region from 1315 – 1470 cm^{-1}. The absorption is consistent with our calculations, and shows that the low energy motifs favour coordination of the COOH moiety in one of three environments: (i) to a backbone amine of one of the Trp residues, (ii) to the other COOH moiety, and/or (iii) to the halide. The remaining features of the experimental spectrum are assigned to NH_2 scissoring (*ca.* 1580 cm^{-1}), the combined methylene CH_2 wag/twist and methine CCH twist (*ca.* 1190 cm^{-1} and 1230 cm^{-1}), HCN bending (*ca.* 1110 cm^{-1}), and NH_2 wagging (*ca.* 1000 cm^{-1}), where each vibration involves displacement of a shared proton.

Given the similarities in cluster binding characteristics, it was surprising to see such a large spread in energetics between motifs, especially when higher energy isomers (*i.e.,* isomers 4 and 5) returned reasonable reproductions of the bands present in the experimental spectrum. For example, motifs 1 and 3 differ only by an O-H···N vs. an O-H···O intermolecular hydrogen bond, despite having respective energy differences of 16.0 kJ mol^{-1}. The experimental spectrum can be accounted for by isomers 1, 2, and 3. However, given the observations for $[Trp_2–H]^-$ and $[Trp\cdots Cl]^-$, it is possible that higher energy kinetically trapped species contribute to the probed ensemble. As indicated by PCM calculations, it is unlikely that higher energy species account for a substantial portion of the ensemble population, as the energetic gaps between motifs 1 to 4 are roughly maintained (see supporting information).

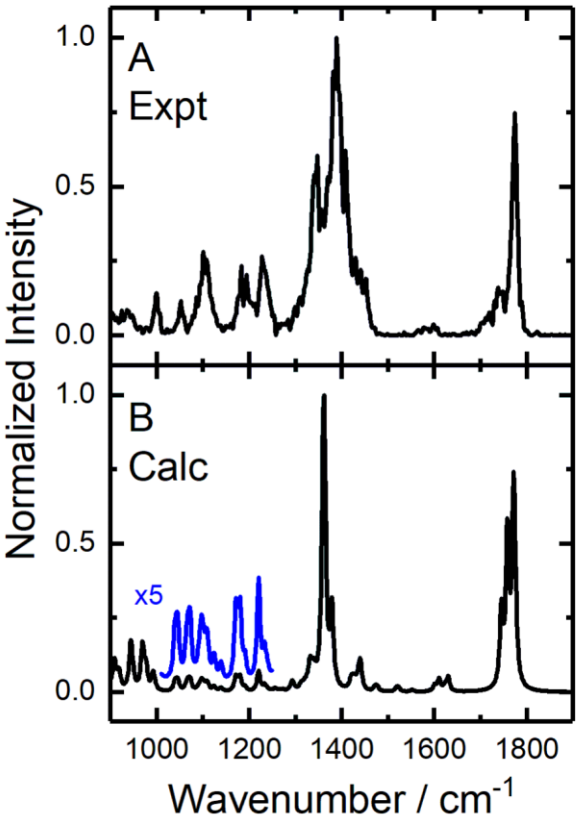

Figure 8. (A) The experimental IRMPD spectrum of $[Trp_2\cdots Cl]^-$ and **(B)** the Boltzmann-weighted convolution of the calculated IR spectra of isomers 1-3. Anharmonic frequency corrections were performed for large amplitude, shared proton modes. Calculations are performed at the B3LYP/6-311++g(d,p) level of theory.

Identifying higher order interactions in $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$

To investigate the prominence of π-π interactions in $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$, a geometric analysis was conducted. A description of the aromatic system is provided in Figure 9, which uses the benzene dimer as an example. The orientation of one ring relative to the other is described using three components: (i) the mass-weighted centroid separation (R), (ii) the dihedral angle between the ring normal vectors (α), and (iii) the distance between the normal vector of one aromatic ring to the centroid of the other (aka. offset).[63] For dihedral angles greater than 90°, α was normalized by 180 – α, and reported as the absolute value.

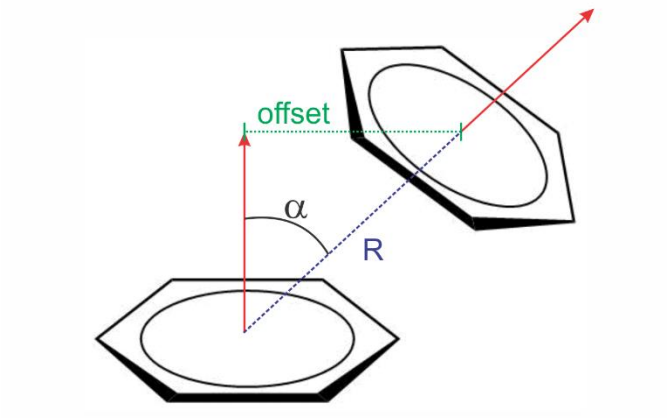

Figure 9. Geometric parameters defining the π-π interaction. For simplicity, benzene is used as a model. Red arrows represent surface normal vectors.

In their 1990 article, Hunter and Sanders developed geometric rules for identifying π-π interactions.[64] In short, these rules state that repulsive interactions occur in a face-to-face π-stacked geometry, while attractive interactions result from T-shaped and/or offset π-stacked orientations. Much of the empirical evidence for these rules came from inspection of the crystal structures of simple aromatics, but π-π interactions have also been identified in larger systems. For example, Hunter, Singh, and Thornton investigated the geometric orientations of Phe-Phe contacts in proteins and derived a computational model to identify regions of electrostatic attraction and repulsion as a function of α *vs*. offset.[65] These regions are plotted in Figure 10; the white regions show attractive interactions and the red regions indicate repulsive interactions. The top panel of Figure 10 shows the geometries corresponding to the different regions of the π-π interaction PES.

To visualize the π-π interactions in the $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$ dimers, the geometric orientations defined by Hunter et al. have been calculated for each isomer found in our computational study (see supporting information). The (α, offset) coordinates for each of the $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$ isomers are plotted in the middle and bottom panel of Figure 10, respectively. Note that datapoints are colour coordinated, with low energy structures ($E_{relative}$ < 10 kJ mol^{-1}) plotted in black.

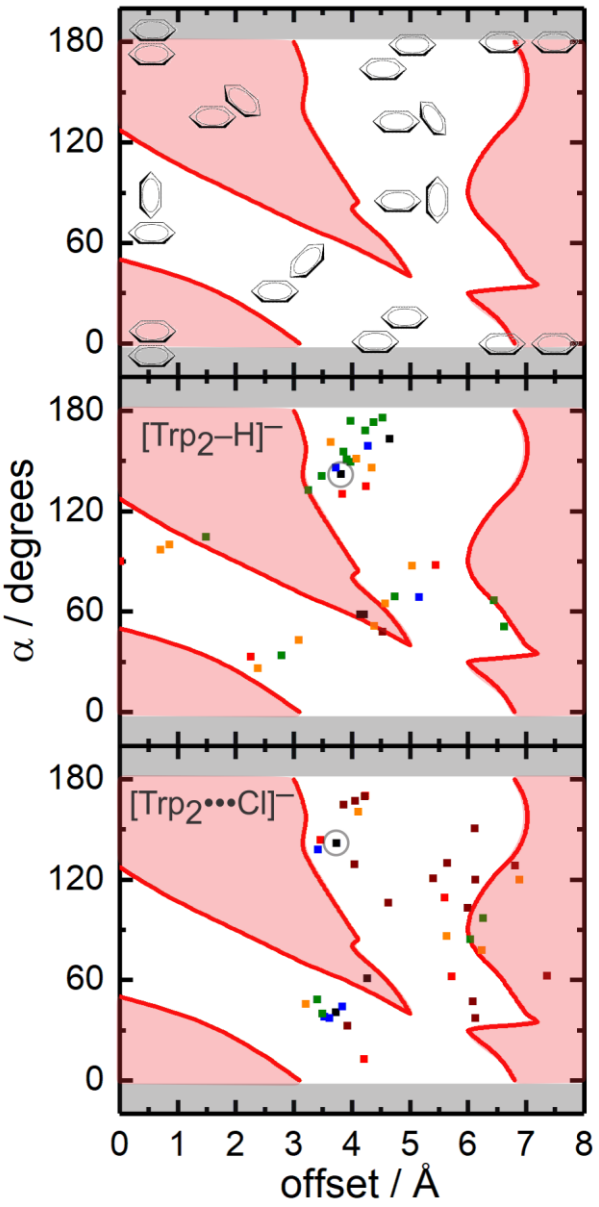

Figure 10. (top) Orientations of Phe-Phe interactions in proteins with respect to the offset and dihedral angle of the two side chains. Figure adapted from reference 63. Offset and dihedral angles of $[Trp_2–H]^-$ **(middle)** and $[Trp_2\cdots Cl]^-$ **(bottom)**. $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$ global minima are circled. (α, offset) data points are colour coded in terms of relative energy: (black) 0 – 10 kJ mol^{-1}, (blue) 10 – 20 kJ mol^{-1}, (green) 20 – 30 kJ mol^{-1}, (orange) 30 – 40 kJ mol^{-1}, (red) 40 – 50 kJ mol^{-1}, (maroon) > 50 kJ mol^{-1}.

Most of the calculated isomers exhibit geometric parameters that lie within the attractive (*viz*., white) regions of the π-π interaction PES. Since each (α, offset) pair corresponds to an orientation of the indole rings relative to one another, the scatter plots shown in Figure 10 allow one to determine trends in π-stacking via inspection. Figure 10 shows that the calculated $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$ isomers predominantly adopt offset, stacked π-π interactions, while T-shaped orientations of the indole rings occur to a lesser extent. The distribution of heterocycle orientations in $[Trp_2\cdots Cl]^-$ is much more

sporadic than those of $[Trp_2–H]^-$; this is likely a result of coordination of the halide via the indole NH moiety. Anion coordination to the indole heterocycle provides a greater energetic stabilization than π-π interactions for the isolated, gas-phase system. Note that the distribution of geometric parameters resembles a similar study that involved identifying π-π interactions in condensed-phase heteroaromatics, including indole-indole interactions.[66] Interestingly, the offset stacked conformation is widely observed in proteins,[67,68] which may suggest that this quantitative analysis on isolated, gas-phase geometries can be utilized in structural or biological studies.

Conclusions

$[Trp_n–H]^-$ and $[Trp_n\cdots Cl]^-$ (n = 1,2) have been studied for the first time using a combined computational and experimental approach. The potential energy surfaces of the clusters have been mapped using a basin hopping algorithm. Combining IRMPD with quantum chemical calculations performed at the DFT level of theory facilitated the assignment of the observed vibrational bands. Deprotonation is found to be localized on the carboxylate of $[Trp_n–H]^-$ clusters. Through a comparison of the unique structural motifs across all clusters investigated, we find that low energy conformers favorably coordinate anionic residues via the indole NH or situate the charge site within the equatorial plane of the heterocycle (owing to anion-π interactions).

In the dimeric species, a series of closely related isomers were identified in which the halide exhibited either a bidentate coordination between the indole NH of both Trp residues, or a tridentate coordination with the amino NH_2, carboxylic OH, and indole NH of adjacent residues. Calculated geometries also suggest that π-π stacking interactions provide additional stabilization to the clusters. Owing to the prevalence of intra- and intermolecular hydrogen bonding, anharmonic corrections were necessary to accurately model the vibrational transitions associated with large amplitude shared proton modes. Geometric analyses reveal that π-π stacking interactions are prominent in the calculated geometries of $[Trp_2–H]^-$ and $[Trp_2\cdots Cl]^-$. These higher order interactions favour parallel, offset geometries, which appear to be more prominent in $[Trp_2–H]^-$ than $[Trp_2\cdots Cl]^-$. This is likely a result of the preference for coordination of the chloride anion to the indole NH moiety. We expect that the other aromatic amino acids (Tyr, His, Phe) in either their deprotonated or halide-bound counterparts will adopt similar binding motifs to maximize hydrogen bonding, anion-π, and π-stacking interactions. Whether this is the case is still an open question. Further exploration of these systems will provide atomistic insight into analogous dynamic ensembles that could be prevalent in biological systems.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

We gratefully acknowledge the high-performance computing support from the SHARCNET consortium of Compute Canada. We are also indebted to the Centre Laser Infrarouge d'Orsay (CLIO) team and support staff for their assistance and hospitality. The authors would like to acknowledge the financial support provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Notes and references

1 D. A. Dougherty, *Science*, 1996, **271**, 163–168.
2 A. R. Fersht, J. P. Shi, J. Knill-Jones, D. M. Lowe, A. J. Wilkinson, D. M. Blow, P. Brick, P. Carter, M. M. Y. Waye and G. Winter, *Nature*, 1985, **314**, 235–238.
3 J. Kobylarczyk, D. Pinkowicz, M. Srebro-Hooper, J. Hooper and R. Podgajny, *Dalt. Trans.*, 2017, **46**, 3482–3491.
4 H. T. Chifotides and K. R. Dunbar, *Acc. Chem. Res.*, 2013, **46**, 894–906.
5 G. Aragay, A. Frontera, V. Lloveras, J.´ Vidal-Gancedo and P. Ballester, *J. Am. Chem. Soc.*, 2013, **135**, 2620–2627.
6 H. T. Chifotides, I. D. Giles and K. R. Dunbar, *J. Am. Chem. Soc.*, 2013, **135**, 3039–3055.
7 S. Guha and S. Saha, *J. Am. Chem. Soc.*, 2010, **132**, 17674–17677.
8 V. K. Dickson, L. Pedi and S. B. Long, *Nature*, 2014, **516**, 213–218.
9 R. B. Stockbridge, L. Kolmakova-Partensky, T. Shane, A. Koide, S. Koide, C. Miller and S. Newstead, *Nature*, 2015, **525**, 548–551.
10 D. Kim, P. Tarakeshwar and K. S. Kim, *J. Phys. Chem. A*, 2004, **108**, 1250–1258.
11 R. Wu and T. B. McMahon, *J. Am. Chem. Soc.*, 2008, **130**, 12554–12555.
12 M. Burt, K. Wilson, R. Marta, M. Hasan, W. S. Hopkins and T. McMahon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 24223–24234.
13 W. Fu, P. J. J. Carr, M. J. Lecours, M. Burt, R. A. Marta, V. Steinmetz, E. Fillion, T. B. McMahon and W. S. Hopkins, *Phys. Chem. Chem. Phys.*, 2017, **19**, 729–734.
14 A. Frontera, P. Gamez, M. Mascal, T. J. Mooibroek and J. Reedijk, *Angew. Chem. Int. Ed.*, 2011, **50**, 9564–9583.
15 B. L. Schottel, H. T. Chifotides and K. R. Dunbar, *Chem. Soc. Rev.*, 2008, **37**, 68–83.
16 M. Giese, M. Albrecht, A. Valkonen and K. Rissanen, *European J. Org. Chem.*, 2013, **16**, 3247–3253.
17 X. Lucas, A. Frontera, D. Quifionero and P. M. Deyà, *J. Phys. Chem. A*, 2010, **114**, 1926–1930.
18 I. V. Krieger, J. S. Freundlich, V. B. Gawandi, J. P. Roberts, V. B. Gawandi, Q. Sun, J. L. Owen, M. T. Fraile, S. I. Huss, J. L. Lavandera, T. R. Ioerger and J. C. Sacchettini, *Chem. Biol.*, 2012, **19**, 1556–1567.
19 S. Li, Y. Xu, Q. Shen, X. Liu, J. Lu, Y. Chen, T. Lu, C. Luo, X. Luo, M. Zheng and H. Jiang, *Curr. Pharm. Des.*, 2013, **19**, 6522–6533.
20 P. Gamez, T. J. Mooibroek, S. J. Teat and J. Reedijk, *Acc.*

Chem. Res., 2007, **40**, 435–444.
21 S. Demeshko, S. Dechert and F. Meyer, *J. Am. Chem. Soc.*, 2004, **126**, 4508–4509.
22 V. M. Philip, J. Harris, R. Adams, D. Nguyen, J. Spiers, J. Baudry, E. E. Howell and R. J. Hinde, *Biochemistry*, 2011, **50**, 2939–2950.
23 M. V. Zlatović, S. Z. Borozan, M. R. Nikolić and S. Đ. Stojanović, *RSC Adv.*, 2015, **5**, 38361–38372.
24 K. Kapoor, M. R. Duff, A. Upadhyay, J. C. Bucci, A. M. Saxton, R. J. Hinde, E. E. Howell and J. Baudry, *Biochemistry*, 2016, **55**, 6056–6069.
25 L. M. Breberina, M. K. Milčić, M. R. Nikolić and S. Đ. Stojanović, *J. Biol. Inorg. Chem.*, 2015, **20**, 475–485.
26 J. P. Schwans, F. Sunden, J. K. Lassila, A. Gonzalez, Y. Tsai and D. Herschlag, *Proc. Natl. Acad. Sci.*, 2013, **110**, 11308–11313.
27 A. N. Bueno, R. K. Shrestha, J. A. Ronau, A. Babar, M. J. Sheedlo, J. E. Fuchs, L. N. Paul and C. Das, *Biochemistry*, 2015, **54**, 6038–6051.
28 S. Z. Borozan, M. V. Zlatović and S. Đ. Stojanović, *J. Biol. Inorg. Chem.*, 2016, **21**, 357–368.
29 Y. Cotelle, V. Lebrun, N. Sakai, T. R. Ward and S. Matile, *ACS Cent. Sci.*, 2016, **2**, 388–393.
30 Y. N. Imai, Y. Inoue, I. Nakanishi and K. Kitaura, *Protein Sci.*, 2008, **17**, 1129–1137.
31 S. Lovas, D. Z. Z. He, H. Liu, J. Tang, J. L. Pecka, M. P. D. Hatfield and K. W. Beisel, *J. Biol. Chem.*, 2015, **290**, 24326–24339.
32 N. C. Polfer, J. Oomens and R. C. Dunbar, *Phys. Chem. Chem. Phys.*, 2006, **8**, 2744–2751.
33 R. C. Dunbar, J. D. Steill, N. C. Polfer and J. Oomens, *J. Phys. Chem. A*, 2009, **113**, 845–51.
34 J. Mino, Warren K., K. Gulyuz, D. Wang, C. N. Stedwell and N. C. Polfer, *J. Phys. Chem. Lett.*, 2011, **2**, 299–304.
35 R. Feng, H. Yin and X. Kong, *Rapid Commun. Mass Spectrom.*, 2016, **30**, 24–28.
36 A. Robertazzi, F. Krull, E.-W. Knapp and P. Gamez, *CrystEngComm*, 2011, **13**, 3293–3300.
37 L. Mac Aleese, A. Simon, T. B. McMahon, J. M. Ortega, D. Scuderi, J. Lemaire and P. Maître, *Int. J. Mass Spectrom.*, 2006, **249**–**250**, 14–20.
38 J. M. Ortega, F. Glotin and R. Prazeres, *Infrared Phys. Technol.*, 2006, **49**, 133–138.
39 W. S. Hopkins, R. A. Marta and T. B. McMahon, *J. Phys. Chem. A*, 2013, **117**, 10714–10718.
40 W. S. Hopkins, R. A. Marta, V. Steinmetz and T. B. McMahon, *Phys. Chem. Chem. Phys.*, 2015, **17**, 28548–28555.
41 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.
42 C. Ieritano, P. J. J. Carr, M. Hasan, M. Burt, R. A. Marta, V. Steinmetz, E. Fillion, T. B. Mcmahon and W. S. Hopkins, *Phys. Chem. Chem. Phys.*, 2016, **18**, 4704–4710.
43 C. Liu, J. C. Y. Le Blanc, J. Shields, J. S. Janiszewski, C. Ieritano, G. F. Ye, G. F. Hawes, W. S. Hopkins and J. L. Campbell, *Analyst*, 2015, **14**, 6897–6903.
44 M. Burt, K. Wilson, R. Marta, M. Hasan, W. Scott Hopkins and T. McMahon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 24223–24234.
45 C. Liu, J. C. Y. Le Blanc, B. B. Schneider, J. Shields, J. J. Federico, H. Zhang, J. G. Stroh, G. W. Kauffman, D. W. Kung, C. Ieritano, E. Shepherdson, M. Verbuyst, L. Melo, M. Hasan, D. Naser, J. S. Janiszewski, W. S. Hopkins and J. L. Campbell, *ACS Cent. Sci.*, 2017, **3**, 101–109.
46 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci and G. A. Petersson, *et al.*, Gaussian, Inc., Wallingford, CT, 2009.
47 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
48 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
49 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
50 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
51 T. Shi, C. K. Siu, K. W. M. Siu and A. C. Hopkinson, *Angew. Chemie Int. Ed.*, 2008, **47**, 8288–8291.
52 W. Scott Hopkins, R. A. Marta, V. Steinmetz and T. B. McMahon, *Phys. Chem. Chem. Phys.*, 2015, **17**, 28548–28555.
53 W. Fu and W. S. Hopkins, *J. Phys. Chem. A*, 2018, **122**, 167–171.
54 U. H. Verkerk, J. Zhao, I. S. Saminathan, J. K. C. Lau, J. Oomens, A. C. Hopkinson and K. W. M. Siu, *Inorg. Chem.*, 2012, **51**, 4707–4710.
55 M. P. Andersson and P. Uvdal, *J. Phys. Chem. A*, 2005, **109**, 2937–2941.
56 S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.
57 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
58 L. Feketeová, G. N. Khairallah, R. A. J. O'Hair and S. B. Nielsen, *Rapid Commun. Mass Spectrom.*, 2015, **29**, 1395–1402.
59 R. J. Nieckarz, J. Oomens, G. Berden, P. Sagulenko and R. Zenobi, *Phys. Chem. Chem. Phys.*, 2013, **15**, 5049–5056.
60 P. Parneix, M. Basire and F. Calvo, *J. Phys. Chem. A*, 2013, **117**, 3954–3959.
61 A. L. Patrick, A. P. Cismesia, L. F. Tesler and N. C. Polfer, *Int. J. Mass Spectrom.*, 2017, **418**, 148–155.
62 J. M. Voss, K. C. Fischer and E. Garand, *J. Mol. Spectrosc.*, 2018, **347**, 28–34.
63 G. B. Mcgaughey, M. Gagne and A. K. Rappe, *J. Biol. Chem.*, 1998, **273**, 15458–15463.
64 C. A. Hunter and J. K. M. Sanders, *J. Am. Chem. Soc.*, 1990, **112**, 5525–5534.
65 C. A. Hunter, J. Singh and J. M. Thornton, *J. Mol. Biol.*, 1991, **218**, 837–846.
66 S. M. Malathy Sony and M. N. Ponnuswamy, *Cryst. Growth Des.*, 2006, **6**, 736–742.
67 J. B. O. Mitchell, C. L. Nandi, I. K. McDonald, J. M. Thornton and S. L. Price, *J. Mol. Biol.*, 1994, 239, 315–331.
68 S. Marsili, R. Chelli, V. Schettino and P. Procacci, *Phys. Chem.*

Chem. Phys., 2008, **10**, 2673–2685.