

The Structure of the Protonated Serine Octamer

Valeriu Scutelnic,^{†,‡} Marta A. S. Perez,^{†,§} Mateusz Marianski,^{⊥,#} Stephan Warnke,^{‡,⊥} Aurelien Gregor,[§] Ursula Rothlisberger,[§] Michael T. Bowers,^{||} Carsten Baldauf,[⊥] Gert von Helden,[⊥] Thomas R. Rizzo,[‡] and Jongcheol Seo^{*,⊥}

[‡]Laboratory of Molecular Physical Chemistry, Ecole Polytechnique Fédérale de Lausanne, Station 6, CH-1015 Lausanne, Switzerland

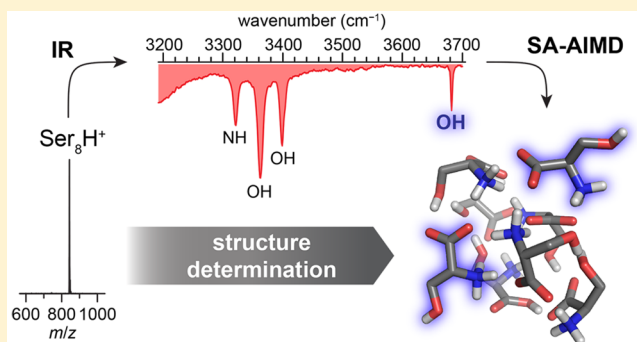
[§]Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

[⊥]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195 Berlin, Germany

^{||}Department of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, California 93106, United States

S Supporting Information

ABSTRACT: The amino acid serine has long been known to form a protonated “magic-number” cluster containing eight monomer units that shows an unusually high abundance in mass spectra and has a remarkable homochiral preference. Despite many experimental and theoretical studies, there is no consensus on a Ser_8H^+ structure that is in agreement with all experimental observations. Here, we present the structure of Ser_8H^+ determined by a combination of infrared spectroscopy and ab initio molecular dynamics simulations. The three-dimensional structure that we determine is $\sim 25 \text{ kcal mol}^{-1}$ more stable than the previous most stable published structure and explains both the homochiral preference and the experimentally observed facile replacement of two serine units.



INTRODUCTION

Enhanced intensities for certain sizes of specific atomic, molecular or ionic clusters in mass spectra are indicative of species with unusually high stabilities and, in many cases, special geometries. Well-known examples are fullerenes,¹ icosahedral rare gas clusters,² or metallocarbohedranes.³ While amino acids have the tendency to self-assemble when produced in the gas phase by electrospray ionization (ESI), in most cases, the intensity distributions in the mass spectrum vary smoothly with cluster size, suggesting a smooth and continuous evolution of properties.^{4,5} However, clusters of the amino acid serine behave differently. Some time ago, it was observed that the ESI mass spectra of solutions containing serine yield either cationic or anionic serine octamer clusters of exceptional abundance relative to their neighboring clusters. The first of these observations was on the protonated serine octamer (Ser_8H^+),^{6–9} followed by several alkali ion complexed octamers such as Ser_8Li^+ and Ser_8Na^+ .¹⁰ Later, exceptionally abundant anionic serine octamer clusters were found in the form of the doubly charged dihalide complexes $\text{Ser}_8\text{Cl}_2^{2-}$ and $\text{Ser}_8\text{Br}_2^{2-}$.¹¹ An additional striking property of serine octamer clusters is their homochiral preference. When D- and L-serine are mixed, the resulting octamer ions do not contain a statistical distribution of the two but prefer eight serines of the same chirality, and this has led to the suggestion that such clusters may have played a role in homochirogenesis.^{12–19}

These observations triggered numerous experimental and theoretical studies aimed at revealing the special structures of serine octamer ions, most of which focused on the protonated species, Ser_8H^+ . For example, ion mobility spectrometry (IMS) has been applied to Ser_8H^+ , and its relatively small collision cross section (CCS) (191 and 187 \AA^2)^{7,9} suggested a compact structure. Another notable experimental observation is that up to two serine units in the homochiral Ser_8H^+ can be easily substituted with either other amino acids, or serine of opposite chirality.^{12,16,20–22} Several stable, compact structures with eight zwitterionic serines have been proposed to explain these experimental observations.^{7,9,23} For example, Counterman and Clemmer proposed a tightly packed, distorted block-like octamer, consisting of seven zwitterionic serines with the eighth having a protonated amine and a neutral carboxylic acid.⁷ Julian et al. suggested a zwitterionic structure as well, but one in which eight serines form a flattened cube consisting of two tetramer units with eight points of contact between them.⁹ Schalley and Weis further explored the conformational space and proposed another zwitterionic flattened cube type structure that contains chains of $\text{O-H}\cdots\text{O}$ hydrogen bonds between the hydroxyl side chain and the carboxylate of an adjacent serine.²³ Although all of these proposed structures for Ser_8H^+ show good

Received: February 21, 2018

Published: April 11, 2018

agreement with the observed CCS value, none of them provide a clear reason for either the homochiral preference or the facile substitution of two serine units. For the anionic $\text{Ser}_8\text{Cl}_2^{2-}$ cluster, a symmetric homochiral structure has been determined; however, this structure is likely very different from any cationic structure.²⁴

More recently, spectroscopic techniques have been applied in an effort to make a more definitive structural assignment of the protonated serine octamer. Infrared (IR) spectroscopy is a particularly powerful tool for structural identification, because vibrational frequencies provide critical insights into the nature of hydrogen-bonding interactions. Nevertheless, previous room-temperature IR spectroscopic studies on Ser_8H^+ in the N–H and O–H stretch region ($3000\text{--}4000\text{ cm}^{-1}$)^{25–28} have not led to a clear structural assignment.

In this work, we use a combination of cryogenic IR messenger-tagging spectroscopy, room-temperature mid-IR multiple-photon dissociation (IRMPD) spectroscopy, and first-principles theory to determine the structure of Ser_8H^+ . Sharp, well-resolved vibrational bands, together with isotopic labeling studies, provide strong validation points for conformational searches on the ab initio potential-energy surface (PES) to yield a global minimum structure of Ser_8H^+ . In particular, we use simulated annealing ab initio molecular dynamics (SA-AIMD) for the conformational search, as density-functional theory (DFT) makes no assumptions about the protonation sites and is also capable of describing shared protons.

MATERIALS AND METHODS

Experimental Approach. All amino acids except isotopically labeled L-serine were obtained from Sigma-Aldrich (Taufkirchen, Germany). The isotopically labeled serines, L-serine-¹⁸O₂, and L-serine-¹³C₃, ¹⁵N₁ were purchased from Cambridge Isotope Laboratories, Inc. (MA, USA). Serine solutions (5 mM) are prepared either in pure water or in 50:50:0.1 (v/v) water/methanol/acetic acid solvent. Mixed solutions of L-serine with D-serine, L-cysteine, or L-alanine were obtained by mixing a 5 mM solution of L-serine with a solution of the same concentration of the partner amino acid.

A detailed description of the tandem quadrupole mass spectrometer instrument for cryogenic ion spectroscopy is given elsewhere.²⁹ Briefly, we generate protonated serine cluster ions by nanoelectrospray ionization, focus them in an ion funnel, and mass-select Ser_8H^+ (m/z 841) in a quadrupole mass filter. The m/z -selected octamer ions are guided to a cryogenic octopole ion trap, which is held at a temperature of 3.5 K, where they are cooled in collisions with helium buffer gas and subsequently tagged with He. We then irradiate the tagged ions with an infrared pulse from a tunable optical parametric oscillator (OPO). When the OPO is tuned in resonance with a vibrational transition, IR radiation is absorbed and then redistributed, causing the weakly bound He atom to evaporate. Monitoring the depletion of the tagged ions as a function of the IR wavenumber allows us to acquire an infrared absorption spectrum.

IRMPD spectroscopy experiments are done using a home-built, hybrid drift-tube, quadrupole time-of-flight (Q-TOF) mass spectrometer, which is described in detail elsewhere.^{30,31} Serine cluster ions generated by nanoelectrospray ionization are pulsed into the 80 cm long drift tube filled with helium buffer gas (~ 4 mbar). Ions traverse the drift tube under the influence of a weak constant electric field and are separated according to their collisional cross section (CCS). The ion mobility-separated ions are guided into the high vacuum region where m/z -selection occurs via a quadrupole mass filter. Isolated Ser_8H^+ ions are then irradiated with intense, tunable infrared radiation provided by the Fritz Haber Institute free electron laser (FHI FEL).³² After the resonant absorption of multiple photons, fragmentation can occur, which is monitored by a time-of-flight mass spectrometer. An

IRMPD spectrum is obtained by plotting the fragmentation yield as a function of IR wavenumber.

Computational Approach. To identify the global minimum of the Ser_8H^+ complex, several search strategies have been adopted that are described in the Supporting Information. These approaches include: (1) screening with force-field-based replica-exchange molecular dynamics (FF-REMD) followed by optimizations at the DFT level, (2) multiple optimizations from symmetrized and randomly generated starting geometries, (3) ab initio replica-exchange molecular dynamics (AI-REMD),^{33,34} and (4) simulated-annealing ab initio molecular dynamics (SA-AIMD).³⁵ Among these four different approaches, ultimately the lowest-energy structure that fulfills all experimental constraints has been found by SA-AIMD.

The SA-AIMD simulations were performed with the Terachem program^{36–38} using the B3LYP³⁹ functional and 6-31G or 6-31G(d,p) basis sets. Briefly, a set of initial complexes, which include symmetric and nonsymmetric clusters, were assembled and optimized at the DFT level. These conformers were then repeatedly heated and slowly cooled to 6 K to accelerate phase-space sampling. To maximize sampling of the conformational space, different maximal heating temperatures (between 700 and 900 K) and different cooling rates were applied. The total simulation time was on average 50 ps/run. In addition to the final conformer, 50 000 geometries were extracted throughout the annealing runs. The optimization of these structures yielded 150 unique conformational families. Finally, 200 low-energy structures obtained from the different searches were combined with 9 structures previously proposed in the literature to compare their energetics. All those structures were reoptimized using the PBE+vdW^{TS} functional^{40,41} and tight basis set settings, as implemented in FHI-aims,⁴² and clustered with a tight 0.1 Å RMSD cutoff. This procedure yielded 185 unique structures which were then again reoptimized at the B3LYP/6-31G(d,p) level of theory, as implemented in Gaussian09.⁴³

For calculating the vibrational spectra of serine octamer, we tested several DFT functionals (B3LYP, B3LYP-D3,⁴⁴ M05-2X,⁴⁵ M11,⁴⁶ and PBE0-D3⁴⁷). Substantial variations were observed in predicted vibrational spectra with different DFT functionals. Among tested methods, we identified B3LYP/6-31G(d,p) as the best method for energies and IR spectra in the NH stretch region ($>2800\text{ cm}^{-1}$), and M05-2X/6-31G(d,p) for IR spectra in the $1000\text{--}2200\text{ cm}^{-1}$ region. Scaling factors of 0.951 for the IR spectrum with B3LYP/6-31G(d,p) in the NH stretch region and 0.943 for the IR spectrum with M05-2X/6-31G(d,p) were used.⁴⁸ The orientationally averaged collision cross sections (CCS) are calculated using the trajectory method (TM) as implemented in Mobcal^{49,50} software with an increased number of impact parameter (total 1500 points in the Monte Carlo integration of the impact factor and orientation). This modification decreases the standard deviation of the calculated CCS to 0.5%.

RESULTS

Mass spectra of electrosprayed serine solutions, measured using either the quadrupole mass-spectrometer setup or the ion mobility mass spectrometer setup, are shown in Figures S1a and S2a, respectively. An exceptionally abundant Ser_8H^+ (m/z 841) peak can be observed, consistent with previous studies.^{6,7,9} Moreover, the mass spectrum obtained after spraying a 50:50 mixture of D-serine and L-serine-¹³C₃, ¹⁵N₁ (Figure S1b) demonstrates the strong homochiral preference of Ser_8H^+ .

The arrival time distribution at m/z 841 measured in the ion mobility mass spectrometer (Figure S2c) shows three peaks which can be assigned to $\text{Ser}_{24}\text{H}_3^{3+}$, $\text{Ser}_{16}\text{H}_2^{2+}$, and Ser_8H^+ .^{5,7,9} The major portion is Ser_8H^+ . The experimentally determined CCS value for Ser_8H^+ in the present study is $191 \pm 2\text{ \AA}^2$, in good agreement with previous measurements.^{7,9} In addition, the ion mobility mass spectrometry results show that besides the small peak from the protonated dimer, protonated serine octamer is the only singly protonated cluster observed (see

Figure S2b), which further indicates the uniqueness of the protonated serine octamer cluster.

The IR spectrum of Ser_8H^+ , obtained under cryogenic conditions in the 2900–3700 cm^{-1} region, is shown in Figure 1.

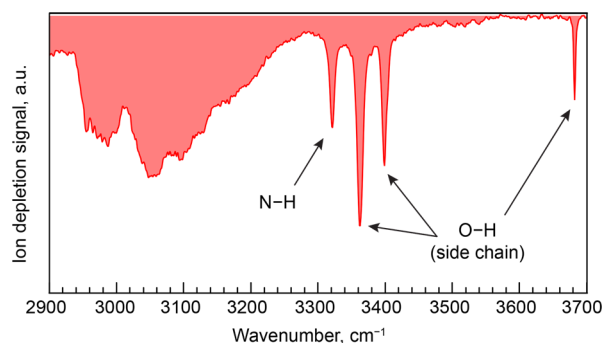


Figure 1. Cold ion infrared spectrum of the He-tagged Ser_8H^+ .

The spectrum displays broad features in the region 2900–3200 cm^{-1} and four sharp peaks appearing at 3317, 3358, 3395, and 3677 cm^{-1} . We can assign the sharp lines by isotope labeling of the serine amine group with ^{15}N or by labeling the carboxylic acid groups with ^{18}O . ^{15}N -labeling shifts the peak at 3317 cm^{-1} (Figure S3), allowing for its assignment as an N–H stretch band. The sharp bands at 3358, 3395, and 3677 cm^{-1} do not shift upon either ^{15}N or carboxylic acid ^{18}O -labeling, indicating that they correspond to O–H stretch modes of the serine hydroxyl side-chain.

An extensive exploration of the Ser_8H^+ conformational space using various search strategies has been performed as described above. The relative energies including zero point energy corrections (ZPE) at the B3LYP/6-31G(d,p) level and the CCS of the merged pool of candidate structures are presented in Figure 2. Intriguingly, all structures derived from the AIMD simulations consistently converge to nonsymmetric Ser_8H^+ structures. The lowest-energy structure (A), which was found using SA-AIMD, is more stable than the previously reported most stable structure (L)²³ by more than 25 kcal mol^{-1} . The lowest energy structure obtained from the exhaustive force-field screening (FF) followed by DFT optimizations lies ~ 8 kcal mol^{-1} above the structure (A). The low-energy region of the plot is populated solely by zwitterionic, nonsymmetric complexes, and we observe that imposing symmetry or constraining to canonical (nonzwitterionic) forms of serine yield higher energy conformers.

The lowest-energy structure A (Figure 3) is stabilized by multiple short, strong hydrogen bonds in which carboxylate groups strongly interact with neighboring $-\text{CH}_2\text{OH}$ or $-\text{NH}_3^+$ groups. It has three protons which are strongly shared between adjacent functional groups. Two of them are in $-\text{NH}_3^+$ groups, each of which is shared with an adjacent carboxylate group ($\text{COO}^- \cdots \text{H}^+ \cdots \text{NH}_2$). The extra proton on the cluster is also shared between two carboxylate groups ($\text{COO}^- \cdots \text{H}^+ \cdots \text{OOC}$) with O–H distances of 1.1 and 1.4 Å. These shared protons are highlighted as spheres in Figure 3.

The structure A contains six serines with their side chains tightly incorporated into the hydrogen bonding network and two serines that have their side chains either free (serine α) or loosely hydrogen-bonded to the core (serine β). The second lowest energy structure in Figure 2, as well as numerous other conformers removed during clustering (see Figure S4) are similar to the structure A, and its serine β has a side-chain

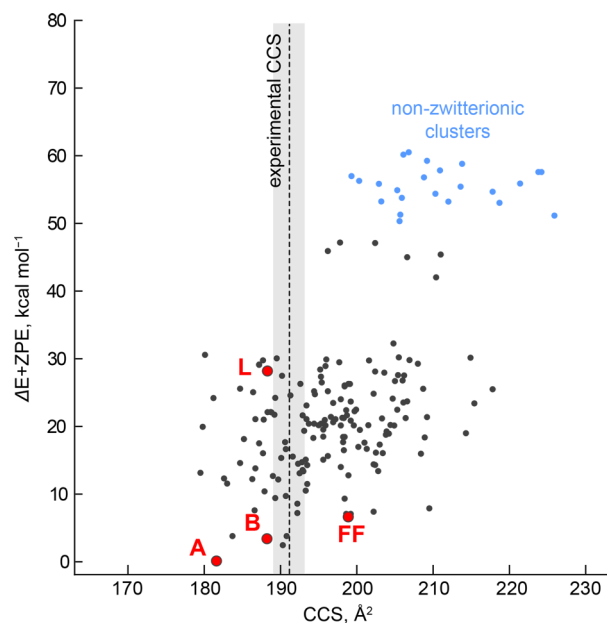


Figure 2. Relative energies and collision cross sections (CCS) of candidate structures for the protonated serine octamer. The x-axis represents the CCS value calculated with the trajectory method. The y-axis represents the relative energies, including zero-point corrections, at the B3LYP/6-31G(d,p) level. The most stable structure found by force field based sampling (FF), the zwitterionic flattened cube structure based on the previous literature (L),²³ the lowest-energy structure overall (A), and the structural candidate (B) are labeled in red. The dotted line represents our measured CCS, and the shaded region is the estimated range of error. Twelve conformers that represent small variations of the conformers A and B are shown in Figure S4.

forming a loose hydrogen bond with a different carboxylate. These lowest-energy structures of type A can easily adopt structure B (Figure 3), in which the hydrogen bond of the hydroxy group in serine β is broken. While the structures A and B differ in energy by 3.3 kcal mol^{-1} at 0 K, their free energy difference at 300 K decreases to 1.6 kcal mol^{-1} using harmonic vibrational corrections. In the AIMD simulations at 300 K, which inherently include anharmonic effects, structure B is more abundant than structure A: a 10 ps trajectory (B3LYP/6-31G(d,p)) shows that the probability of the occurrence of a $-\text{CH}_2\text{OH} \cdots \text{O}$ hydrogen bond in serine β is 24% and further decreases to less than 10% at 600 K. (see Figure S5). Structure B has a CCS value of $189 \pm 1 \text{ \AA}^2$, which is in good agreement with the experimental values either in the present work ($191 \pm 2 \text{ \AA}^2$) or in the previous literature (187 and 191 \AA^2).^{7,9}

Figure 4 compares the measured IR spectrum (red trace) with the calculated spectrum of structure B (red sticks). In the latter, we find two stretch modes of the free side chain OH groups (α and β in Figure 3) at 3653 and 3646 cm^{-1} , which fit well to the experimental band position of 3677 cm^{-1} . Stretch modes of hydrogen-bonded OH side-chains appear at 3430 and 3353 cm^{-1} , in good agreement with the measured bands at 3395 and 3358 cm^{-1} , which were assigned to side chain OH stretch modes by isotope labeling (Figure 1 and Figure S3). Two vibrational bands from the NH_2 asymmetric stretch of $-\text{NH}_3^+$ groups that have one H weakly hydrogen-bonded appear at 3280 cm^{-1} (corresponding to the measured band at 3317 cm^{-1}). In the region between 2900 and 3200 cm^{-1} , a large number of ammonium stretch modes gives rise to unresolved

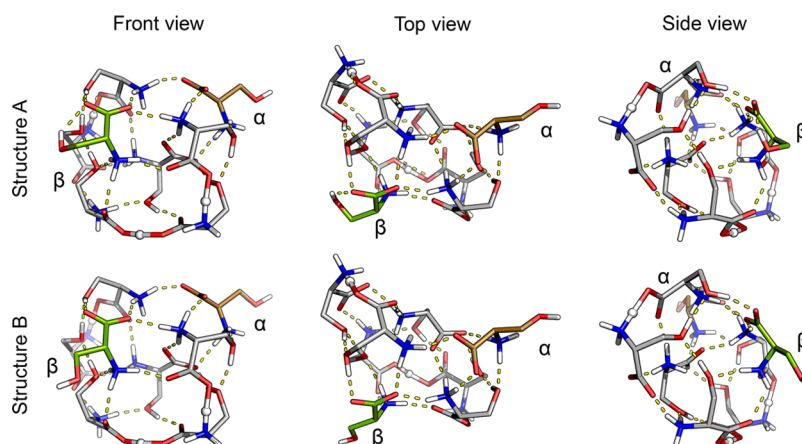


Figure 3. Different views of two structures of the protonated serine octamer (structures A and B in Figure 2). Dashed lines represent the hydrogen bonds. Nonpolar hydrogens are omitted for clarity. The white spheres represent a shared proton between two carboxylate groups and another two shared protons at $\text{COO}^- \cdots \text{H}^+ \cdots \text{NH}_2$. See Movie S1, which shows the structure rotating around 360° .

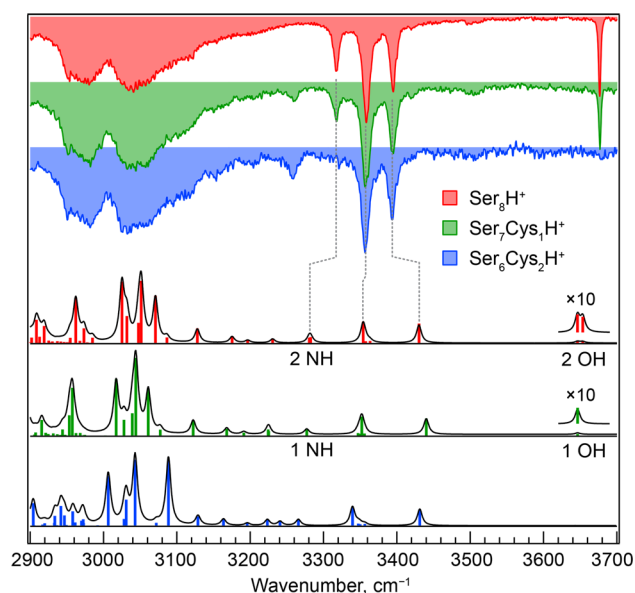


Figure 4. Experimental and theoretical IR spectra of the protonated serine octamer with cysteine substitutions. Red, green, and blue spectra represent the protonated serine octamer with no, one, and two cysteine substitutions, respectively. The IR spectra were scaled to the peak at 3358 cm^{-1} .

overlapping bands in both the experiment and the calculated spectra. The good agreement between the calculations and the well-resolved experimental bands gives strong support for an assignment to conformer B.

To confirm this structural assignment, we measured spectra of mixed serine/cysteine octamers. As shown in Figure 4, IR spectra of $\text{Ser}_7\text{Cys}_1\text{H}^+$ and $\text{Ser}_6\text{Cys}_2\text{H}^+$ resemble that of Ser_8H^+ , with a few subtle, but important, differences. Upon the first substitution (green trace), the bands at 3320 and 3677 cm^{-1} decrease in intensity and a new band appears at 3258 cm^{-1} , whereas other parts of the spectrum remain unchanged. The second substitution (blue trace) leads to similar changes, but now the two bands at 3317 and 3677 cm^{-1} disappear completely, and the band at 3258 cm^{-1} grows by a factor of 2 in intensity. The disappearance of the free OH band at 3677 cm^{-1} upon these two substitutions suggests that the two serines

with a free OH side-chain are substituted by cysteines, because the SH stretch band should appear below 2700 cm^{-1} .⁵¹

These experimental observations are confirmed by theory. Calculations of $\text{Ser}_7\text{Cys}_1\text{H}^+$ and $\text{Ser}_6\text{Cys}_2\text{H}^+$ show that the substitutions of up to two serines by cysteines induce only minimal perturbations to the octamer structure, and the calculated and experimental spectra for $\text{Ser}_7\text{Cys}_1\text{H}^+$ and $\text{Ser}_6\text{Cys}_2\text{H}^+$ are in good agreement (see green and blue stick spectra in Figure 4). The two cysteines replace the two serine units in which the OH side chain is not hydrogen-bonded (serine α and β in Figure 3). This replacement preserves the strong hydrogen-bond character of the cluster, maintaining its maximum stability. Cysteine replacement in any other position results in a less stable structure and yields an IR spectrum that does not match with the experiment. $\text{Ser}_7\text{Cys}_1\text{H}^+$ lacks one free OH from Ser_8H^+ , which appears at 3653 cm^{-1} , because it is replaced by SH that shifts this line down to 2562 cm^{-1} . Likewise, the intensity of the band at 3280 cm^{-1} is decreased for $\text{Ser}_7\text{Cys}_1\text{H}^+$, because the replacement of serine red-shifts the NH stretch mode by $\sim 50\text{ cm}^{-1}$. Thus, the calculated spectrum of $\text{Ser}_7\text{Cys}_1\text{H}^+$ exhibits one peak at $\sim 3280\text{ cm}^{-1}$ and two peaks at $\sim 3230\text{ cm}^{-1}$, justifying the decrease in the measured intensity of the NH band at 3317 cm^{-1} and the increase in intensity at 3258 cm^{-1} . In $\text{Ser}_6\text{Cys}_2\text{H}^+$, both serine α and β are replaced by cysteine, causing the second OH band at 3646 cm^{-1} to disappear. Moreover, the second NH band that is observed at 3280 cm^{-1} for Ser_8H^+ and for $\text{Ser}_7\text{Cys}_1\text{H}^+$ is red-shifted by 11 cm^{-1} in $\text{Ser}_6\text{Cys}_2\text{H}^+$, in agreement with the total disappearance of the band at 3317 cm^{-1} in the measured spectrum.

We also measured spectra of serine octamers with mixed chirality in the $2900\text{--}3700\text{ cm}^{-1}$ region (see Figure S6). In this case, theory predicts that for $(\text{L-Ser})_7(\text{D-Ser})_1\text{H}^+$, the L-serine α is replaced by D-serine. For $(\text{L-Ser})_6(\text{D-Ser})_2\text{H}^+$, both L-serines α and β are then substituted by D-serine. Their theoretical vibrational spectra match well with the measured IR spectra.

To further confirm the structural assignment, IR spectra of serine octamer ions in the wavenumber range of $1000\text{--}2200\text{ cm}^{-1}$ were measured and compared with theoretical predictions (Figure 5). The room-temperature IRMPD and He-tagged cryogenic IR spectra have similar features (Figure 5b, c). For amino acids, the C=O stretch bands of a neutral carboxylic acid should appear in the range $1700\text{--}1800\text{ cm}^{-1}$.^{52,53}

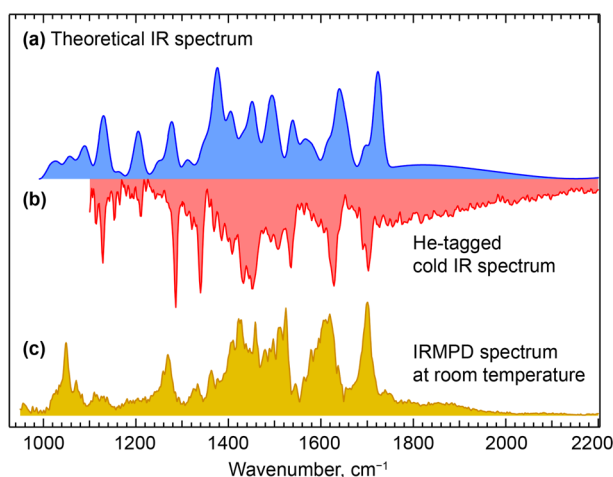


Figure 5. (a) Theoretical IR spectrum of structure B of the Ser_8H^+ , (b) experimental cold IR spectrum of the He-tagged Ser_8H^+ , and (c) experimental IRMPD spectrum of the Ser_8H^+ at room temperature.

Carboxylate groups, on the other hand, give symmetric and antisymmetric stretch bands in the 1300–1450 and 1600–1650 cm^{-1} region, respectively.^{54,55}

In the experimental spectra, the IR bands at $\sim 1620 \text{ cm}^{-1}$ can be assigned to carboxylate groups and the NH bending mode of $-\text{NH}_3^+$. The band at $\sim 1420 \text{ cm}^{-1}$ confirms the presence of carboxylates. The observed IR band at $\sim 1695 \text{ cm}^{-1}$ indicates the presence of a strongly hydrogen-bonded carbonyl (C=O) group of a carboxylic acid. Additional experiments further support these assignments, which show that the intense IR bands at ~ 1695 and $1410\text{--}1430 \text{ cm}^{-1}$ do not shift to lower wavenumber upon H/D exchange (see Figure S7). Some broad features at $1560\text{--}1620 \text{ cm}^{-1}$ also remain after H/D exchange. These results suggest the presence of both carboxylic acids and carboxylates.

The calculated IR spectrum for structure B in this wavenumber region is shown in Figure 5a. Each line in the theoretical vibrational spectrum is represented by a Gaussian with full width at half-maximum (fwhm) of 20 cm^{-1} except for bands $>1750 \text{ cm}^{-1}$ where vibrational modes involving motion of the shared protons appear. Proton motions between two carboxylates and/or at $\text{CO}\cdots\text{H}\cdots\text{N}$ give rise to bands at 2346, 1896, and 1780 cm^{-1} . Typically, vibrational modes of such shared protons are highly anharmonic and give rise to broad spectral features.^{56,57} For this reason, we set the fwhm of the Gaussian of the vibrations in the region $>1750 \text{ cm}^{-1}$ to 200 cm^{-1} . The experimentally observed broad IR features at $>1750 \text{ cm}^{-1}$ fits well to theory, which further confirms the presence of the proton-sharing carboxylates. Complete H/D exchange shifts all these broad features in both the cryogenic IR spectrum as well as in the IRMPD spectrum (see Figure S7c–f), which gives additional support for the assignment as shared proton modes.

Theory predicts IR bands at $\sim 1695 \text{ cm}^{-1}$ corresponding to antisymmetric stretch modes of three carboxylates with shared protons ($\text{COO}^-\cdots\text{H}^+\cdots\text{OOC}$ and $\text{COO}^-\cdots\text{H}^+\cdots\text{NH}_2$). The strongly shared protons increase acidic character of carboxylates, which makes antisymmetric stretches of carboxylate similar to the carbonyl stretches of strongly hydrogen-bonded carboxylic acids. Therefore, they appear at $\sim 1695 \text{ cm}^{-1}$, which is higher than typically found for the antisymmetric stretch of carboxylates. Antisymmetric and symmetric stretch modes of all

other COO^- groups are calculated to appear at ~ 1620 and $\sim 1440 \text{ cm}^{-1}$, which are once again in agreement with the experimental observations. All these results in the $1000\text{--}2200 \text{ cm}^{-1}$ region corroborate our proposed structure for Ser_8H^+ .

DISCUSSION

There are several constraints derived from experiments that a Ser_8H^+ structure has to fulfill: (a) it has to be consistent with the measured IR spectra and collisional cross-section; (b) it has to be of exceptional stability, implying a (near) perfect hydrogen bonding pattern; (c) it has to have a strong preference for homochirality; (d) it must permit the facile substitution of up to two amino acids, either serine of different chirality or other amino acids.

As demonstrated above, the proposed structure B fulfills all the above-mentioned constraints. The calculated IR spectra are in excellent agreement with the measured spectra of Figures 4 and 5, as verified by isotopic substitution. The predicted CCS value ($189 \pm 1 \text{ \AA}^2$) also agrees well with the experimental value. All eight serines are zwitterionic and form a tight and nearly complete hydrogen-bonding network, which is severely disrupted upon the removal of a single serine. Furthermore, this octameric assembly does not have docking points for the addition of another amino acid as only the two hydroxyl groups in serines α/β and a few N–Hs are exposed to the outside. Therefore, the formation of Ser_9H^+ by adding one serine to Ser_8H^+ is not favored. This is confirmed by additional theoretical calculations to determine the lowest-energy structure of the Ser_9H^+ . The predicted structure of Ser_9H^+ suggests that the compact octamer has a complete shell, and an extra serine monomer initiates a new shell that surrounds Ser_8H^+ (see Figure S8). Moreover, the calculated absolute binding energy per monomer in Ser_9H^+ ($22.8 \text{ kcal mol}^{-1}$) is $2.7 \text{ kcal mol}^{-1}$ smaller than that in Ser_8H^+ ($25.5 \text{ kcal mol}^{-1}$). This clearly rationalizes the magic number of eight.

The preference for homochirality is the consequence of the three-point interactions between the serine units, as previously suggested.^{9,23} In such three-point interactions, three functional groups of a serine (a carboxylate, an ammonium, and a side chain hydroxyl) form intermolecular hydrogen bonds with adjacent serines. Because of the chirality of serine, such three-point interactions will have a certain directionality, and the entire network of three-point interactions will depend on the chiral nature of its components. However, while the protonated serine octamer is strongly homochiral, it is not absolutely homochiral as shown in (see Figure S1), and the structure we propose nicely explains this. In conformer B, two serine units (α and β) have free side chains, and thus are only involved in intrinsically nonchiral, two-point interactions, while other six are forming three-point interactions. They are thus more loosely bound in the cluster than the other six and more easily able to be substituted by other amino acids (including serine of opposite chirality) without inducing a major change of the octamer structure. Theory predicts that replacing the serine α or β in homochiral $(\text{L-Ser})_8\text{H}^+$ with a D-serine does not significantly destabilize the cluster ($<4 \text{ kcal mol}^{-1}$). Furthermore, the spectra of $\text{Ser}_7\text{Ala}_1\text{H}^+$ and $\text{Ser}_7\text{Cys}_1\text{H}^+$ in the $1000\text{--}2200 \text{ cm}^{-1}$ region are practically identical with that of Ser_8H^+ (see Figure S7a–c), which confirms that the internal hydrogen-bonding network of the octamer is not affected by the substitution in the α and β positions. However, substitution in other position breaks the internal hydrogen bonding network, greatly destabilizing the entire cluster. Theory predicts that (L-

$(\text{Ser})_7(\text{D-Ser})_1\text{H}^+$ is 8–14 kcal mol⁻¹ less stable than the homochiral cluster when a D-serine is in a location other than α or β position. In such cases, the side chain OH of the substituted D-serine cannot be involved in the hydrogen bonding network and is exposed outside (see Figure S9).

The serine octamer has been observed to exhibit special behavior not only in its protonated form but also as adducts with two halides (chlorides and/or bromides) to form anionic species.²⁴ Cooks and co-workers suggested that both cationic and anionic octamers (Ser_8H^+ and $\text{Ser}_8\text{X}_2^{2-}$, X = Cl or Br) may be structural analogues, and further proposed the possible presence of a common neutral octamer structure which maintains high stability as well as homochiral preference.^{6,11} However, since the charge plays a significant role in determining the structure for Ser_8H^+ , there is not likely to be a strong correlation between the structures of the protonated and neutral species. Moreover, the structure of Ser_8H^+ is significantly different from that of $\text{Ser}_8\text{Cl}_2^{2-}$, which is highly symmetric and in which eight zwitterionic serines form a cage, encapsulating two halogen ions.²⁴

For both Ser_8H^+ and $\text{Ser}_8\text{X}_2^{2-}$, the charge plays a crucial role in determining the structure. In the case of the anions, the halogens stabilize the structure, and removing them comes at a large energy penalty. For Ser_8H^+ , the additional proton is shared by two carboxylates, serving as “electrostatic glue” that mitigates their mutual repulsion. Removing this proton would thus lead to a complete structural reorganization. It thus seems clear that the anionic, cationic, and a presumably neutral cluster have structurally little in common. Therefore, the same magic number of eight for both cationic and anionic serine octamer may be just a coincidence. Nevertheless, this seems somewhat surprising, as such magic numbers are only observed for serine, and only for eight monomer units.

CONCLUSIONS

In summary, we report the structure of the Ser_8H^+ as determined by the combination of cryogenic ion spectroscopy of He-tagged Ser_8H^+ , room-temperature mid-infrared multiple photon dissociation spectroscopy, and simulated annealing ab initio molecular dynamics. The proposed structure is outstandingly stable, maximizing the number of strong hydrogen bonds. The presence of two free side-chain OH groups (serine α and β) explains the facile substitution of only two serines. We observe that if any serine other than α and β is substituted, either by another amino acid or by D-serine, it strongly disrupts the hydrogen bonding network of the entire cluster.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b02118.

Additional computational details of conformational search techniques and the supplementary data which include (i) mass spectra as well as ion mobility spectrometry (IMS) results of serine solutions, (ii) additional infrared spectra of the protonated serine octamer with various isotopic substitutions or amino acid substitutions, (iii) 3D structure of 12 low-energy structures, (iv) theoretical calculation for Ser_9H^+ , and (v) theoretical calculation for $(\text{L-Ser})_7(\text{D-Ser})_1\text{H}^+$. Separate files containing (i) xyz-coordinates of the structures A and B, and (ii) . (PDF)

xyz-coordinates of the structure A (XYZ)
xyz-coordinates of the structure B (XYZ)
Movie S1, rotation of structure B around 360° (MPG)

AUTHOR INFORMATION

Corresponding Author

*E-mail: jseo@fhi-berlin.mpg.de

ORCID

Ursula Rothlisberger: 0000-0002-1704-8591

Carsten Baldauf: 0000-0003-2637-6009

Gert von Helden: 0000-0001-7611-8740

Thomas R. Rizzo: 0000-0003-2796-905X

Jongcheol Seo: 0000-0001-5844-4585

Present Address

#M.M. is currently at Hunter College, The City University of New York

Author Contributions

†V.S. and M.P. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

G.v.H. and J.S. acknowledge the expert assistance of the FHI free electron laser facility staff, in particular S. Gewinner and W. Schöllkopf. M.M. and C.B. are grateful to Matthias Scheffler (FHI Berlin) for support of the computational work. T.R.R. and U.R. thank the EPFL, and T.R.R. the Swiss National Science Foundation for the support of this work under grant No. 200020_165908. M.T.B. thanks the Alexander von Humboldt Foundation and the National Science Foundation (USA) for support under grant No. CHE-1565941. M.M., C.B., G.v.H., and T.R.R. thank the joint Max-Planck-EPFL Center for Molecular Nanoscience and Technology for financial support.

REFERENCES

- (1) Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. *Nature* **1985**, *318*, 162–163.
- (2) Miehle, W.; Kandler, O.; Leisner, T.; Echt, O. *J. Chem. Phys.* **1989**, *91*, 5940–5952.
- (3) Guo, B. C.; Wei, S.; Purnell, J.; Buzza, S.; Castleman, A. W. *Science* **1992**, *256*, 515–516.
- (4) Nemes, P.; Schlosser, G.; Vekey, K. *J. Mass Spectrom.* **2005**, *40*, 43–49.
- (5) Do, T. D.; de Almeida, N. E. C.; LaPointe, N. E.; Chamas, A.; Feinstein, S. C.; Bowers, M. T. *Anal. Chem.* **2016**, *88*, 868–876.
- (6) Cooks, R. G.; Zhang, D.; Koch, K. J.; Gozzo, F. C.; Eberlin, M. N. *Anal. Chem.* **2001**, *73*, 3646–3655.
- (7) Counterman, A. E.; Clemmer, D. E. *J. Phys. Chem. B* **2001**, *105*, 8092–8096.
- (8) Koch, K. J.; Gozzo, F. C.; Zhang, D. X.; Eberlin, M. N.; Cooks, R. G. *Chem. Commun.* **2001**, 1854–1855.
- (9) Julian, R. R.; Hodyss, R.; Kinnear, B.; Jarrold, M. F.; Beauchamp, J. L. *J. Phys. Chem. B* **2002**, *106*, 1219–1228.
- (10) Nanita, S. C.; Sokol, E.; Cooks, R. G. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 856–868.
- (11) Nanita, S. C.; Cooks, R. G. *J. Phys. Chem. B* **2005**, *109*, 4748–4753.
- (12) Takats, Z.; Nanita, S. C.; Cooks, R. G. *Angew. Chem., Int. Ed.* **2003**, *42*, 3521–3523.
- (13) Julian, R. R.; Myung, S.; Clemmer, D. E. *J. Am. Chem. Soc.* **2004**, *126*, 4110–4111.
- (14) Nanita, S. C.; Takats, Z.; Cooks, R. G.; Myung, S.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1360–1365.
- (15) Takats, Z.; Cooks, R. G. *Chem. Commun.* **2004**, 444–445.

- (16) Nanita, S. C.; Cooks, R. G. *Angew. Chem., Int. Ed.* **2006**, *45*, 554–569.
- (17) Yang, P. X.; Xu, R. F.; Nanita, S. C.; Cooks, R. G. *J. Am. Chem. Soc.* **2006**, *128*, 17074–17086.
- (18) Perry, R. H.; Wu, C. P.; Nefliu, M.; Cooks, R. G. *Chem. Commun.* **2007**, 1071–1073.
- (19) Weissbuch, I.; Lahav, M. *Chem. Rev.* **2011**, *111*, 3236–3267.
- (20) Koch, K. J.; Gozzo, F. C.; Nanita, S. C.; Takats, Z.; Eberlin, M. N.; Cooks, R. G. *Angew. Chem., Int. Ed.* **2002**, *41*, 1721–1724.
- (21) Takats, Z.; Nanita, S. C.; Schlosser, G.; Vekey, K.; Cooks, R. G. *Anal. Chem.* **2003**, *75*, 6147–6154.
- (22) Gronert, S.; O'Hair, R. A. J.; Fagin, A. E. *Chem. Commun.* **2004**, 1944–1945.
- (23) Schalley, C. A.; Weis, P. *Int. J. Mass Spectrom.* **2002**, *221*, 9–19.
- (24) Seo, J.; Warnke, S.; Pagel, K.; Bowers, M. T.; von Helden, G. *Nat. Chem.* **2017**, *9*, 1263–1268.
- (25) Kong, X.; Tsai, I. A.; Sabu, S.; Han, C.-C.; Lee, Y. T.; Chang, H.-C.; Tu, S.-Y.; Kung, A. H.; Wu, C.-C. *Angew. Chem., Int. Ed.* **2006**, *45*, 4130–4134.
- (26) Kong, X.; Lin, C.; Infusini, G.; Oh, H.-B.; Jiang, H.; Breuker, K.; Wu, C.-C.; Charkin, O. P.; Chang, H.-C.; McLafferty, F. W. *ChemPhysChem* **2009**, *10*, 2603–2606.
- (27) Sunahori, F. X.; Yang, G.; Kitova, E. N.; Klassen, J. S.; Xu, Y. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1873–1886.
- (28) Liao, G. H.; Yang, Y. J.; Kong, X. L. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1554–1558.
- (29) Svendsen, A.; Lorenz, U. J.; Boyarkin, O. V.; Rizzo, T. R. *Rev. Sci. Instrum.* **2010**, *81*, 073107.
- (30) Warnke, S.; Baldauf, C.; Bowers, M. T.; Pagel, K.; von Helden, G. *J. Am. Chem. Soc.* **2014**, *136*, 10308–10314.
- (31) Warnke, S.; Seo, J.; Boschmans, J.; Sobott, F.; Scrivens, J. H.; Bleiholder, C.; Bowers, M. T.; Gewinner, S.; Schöllkopf, W.; Pagel, K.; von Helden, G. *J. Am. Chem. Soc.* **2015**, *137*, 4236–4242.
- (32) Schöllkopf, W.; Gewinner, S.; Junkes, H.; Paarmann, A.; von Helden, G.; Bluem, H. P.; Todd, A. M. M. *Proc. SPIE* **2015**, *9512*, 95121L.
- (33) Baldauf, C.; Rossi, M. *J. Phys.: Condens. Matter* **2015**, *27*, 493002.
- (34) Schubert, F.; Rossi, M.; Baldauf, C.; Pagel, K.; Warnke, S.; von Helden, G.; Filsinger, F.; Kupser, P.; Meijer, G.; Salwiczek, M.; Koks, B.; Scheffler, M.; Blum, V. *Phys. Chem. Chem. Phys.* **2015**, *17*, 7373–7385.
- (35) Aseev, O.; Perez, M. A. S.; Rothlisberger, U.; Rizzo, T. R. *J. Phys. Chem. Lett.* **2015**, *6*, 2524–2529.
- (36) Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.* **2009**, *5*, 1004–1015.
- (37) Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.* **2009**, *5*, 2619–2628.
- (38) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Chem. Theory Comput.* **2013**, *9*, 213–221.
- (39) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (40) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (41) Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 73005.
- (42) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. *Comput. Phys. Commun.* **2009**, *180*, 2175–2196.
- (43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision D.01*; Gaussian Inc.: Wallingford, CT, 2009.
- (44) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
- (45) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.
- (46) Peverati, R.; Truhlar, D. G. *J. Phys. Chem. Lett.* **2011**, *2*, 2810–2817.
- (47) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982.
- (48) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683–11700.
- (49) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996**, *100*, 16082–16086.
- (50) Shvartsburg, A. A.; Jarrold, M. F. *Chem. Phys. Lett.* **1996**, *261*, 86–91.
- (51) Miller, B. J.; Howard, D. L.; Lane, J. R.; Kjaergaard, H. G.; Dunn, M. E.; Vaida, V. *J. Phys. Chem. A* **2009**, *113*, 7576–7583.
- (52) Antony, J.; von Helden, G.; Meijer, G.; Schmidt, B. *J. Chem. Phys.* **2005**, *123*, 014305.
- (53) Linder, R.; Seefeld, K.; Vavra, A.; Kleinermanns, K. *Chem. Phys. Lett.* **2008**, *453*, 1–6.
- (54) Kapota, C.; Lemaire, J.; Maitre, P.; Ohanessian, G. *J. Am. Chem. Soc.* **2004**, *126*, 1836–1842.
- (55) Oomens, J.; Steill, J. D.; Redlich, B. *J. Am. Chem. Soc.* **2009**, *131*, 4310–4319.
- (56) Roscioli, J. R.; McCunn, L. R.; Johnson, M. A. *Science* **2007**, *316*, 249–254.
- (57) Van Hoozen, B. L., Jr.; Petersen, P. B. *J. Chem. Phys.* **2015**, *142*, 104308.