Supporting information

The Structure of the Protonated Serine Octamer

Valeriu Scutelnic,[1,‡] Marta A. S. Perez,[2,‡] Mateusz Marianski,[3,†] Stephan Warnke,[1,3] Aurelien Gregor,[2] Ursula Rothlisberger,[2] Michael T. Bowers,[4] Carsten Baldauf,[3] Gert von Helden,[3] Thomas R. Rizzo,[1] Jongcheol Seo[3,*]

[1] Laboratory of Molecular Physical Chemistry, Ecole Polytechnique Fédérale de Lausanne, Station 6, CH-1015 Lausanne, Switzerland.
[2] Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.
[3] Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195 Berlin, Germany.
[4] Department of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, California 93106, USA.

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

† Currently at Hunter College, The City University of New York

‡ These authors contributed equally to this work.

Table of Contents

1. Supplementary Computational Details

In addition to the simulated annealing *ab initio* MD (SA-AIMD), we used three additional search techniques: (1) screening with force field-based replica-exchange MD (REMD) followed by optimizations at the DFT level, (2) multiple optimizations from symmetrized and randomly generated starting geometries, and (3) *ab initio* REMD (AI-REMD). The details are the following:

Force field REMD. The force field REMD screening was performed in Gromacs 5.1.1.[1] Since common empirical potentials prohibit any proton transfer during simulations, we constructed all 20 tautomers of protonated serine octamer that feature zwitterionic, canonical or charged (positive or negative) species together with an excess proton. The parameters for the amino acid were derived from the OPLS/AA force field [2] augmented with an extra set of parameters for zwitterionic amino acids. Each tautomer has been simulated in 16 entangled trajectories which spanned the temperature range from 300 to 900 K and featured an exchange probability of 0.65 [3]. Each trajectory has been simulated for at least a few hundred nanoseconds, accumulating for a total time of approximately 50 μs for 20 tautomers. Next, the 300 K trajectories were clustered with strict RMSD criteria (0.1 Å), and the central structures of the 10 most populated clusters were optimized using the dispersion-corrected PBE+vdW^{TS} functional [4-5] and *tight* basis set setting in FHI-aims[6].

Multiple optimizations strategy. Alternatively, approximately 4000 symmetric and random clusters of Ser_8H^+ (only in canonical or zwitterionic forms) were constructed as initial structures (see SI for details). The calculations were carried out in FHI-aims using the PBE+vdW^{TS} functional and *light* basis set setting and loose convergence criteria. Next, the 50 lowest energy conformers from each symmetric and random set of structures have been reoptimized with *tight* basis set.

***Ab initio* REMD.** The AI-REMD[7] is initiated from structures obtained from FF-based low-mode molecular dynamics as implemented in MOE[8-9] and were performed in FHI-aims using PBE+vdW^{TS} functional and *light* basis set settings. Four independent runs, each featuring either 8 or 12 trajectories, mounted to 812 ps of simulation time. Next, 8129 conformers (a snapshot every 100 fs) are extracted and reoptimized using loose convergence criteria. From the resulting structures, 100 unique ones were selected and reoptimized with tight basis set settings.

2. Mass Spectrometry and Ion Mobility Spectrometry Results

The protonated serine octamer (Ser_8H^+) is exceptionally abundant in the ESI mass spectra compared with other serine clusters (Figure S1a and S2a) and exhibits a strong homochiral preference (Figure S1b). Ion mobility spectrometry-mass spectrometry (IMS-MS) confirms that Ser_8H^+ is the only singly-protonated serine cluster (Figure S2b). The collision cross section value of the Ser_8H^+ measured by ion mobility spectrometry is 191 $Å^2$ (Figure S2c).

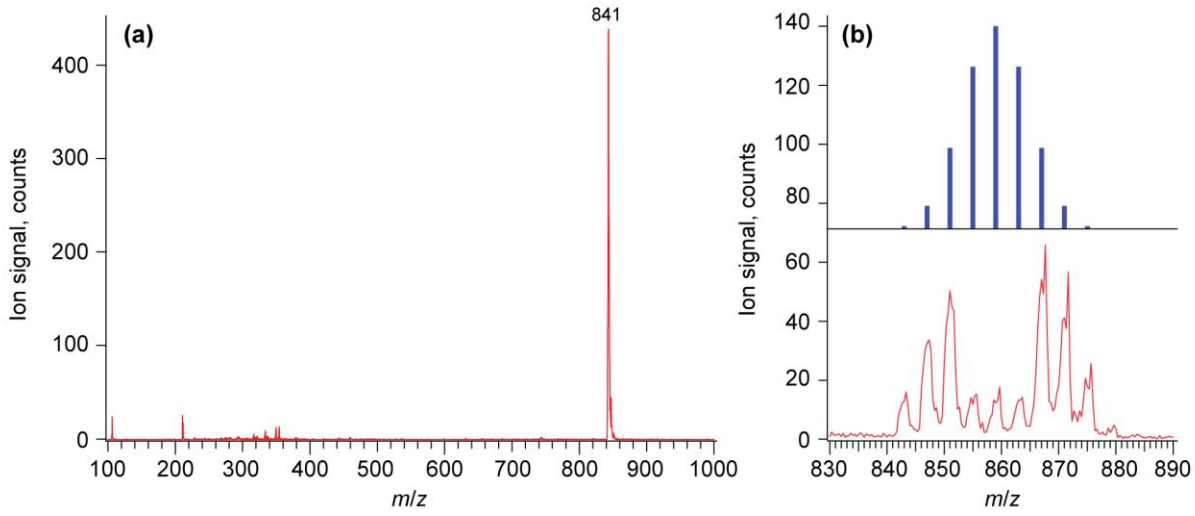

Figure S1. nESI-mass spectra of (a) the acidified enantiome rically pure L-serine solution (5 mM) in water/methanol and (b) the mixture of D-serine and L-serine-$^{13}C_3$,$^{15}N_1$ measured in the quadrupole mass spectrometer for cryogenic ion spectroscopy. The mass spectrum of the racemic serine octamer is compared with a statistical distribution (upper panel). The sample was nanosprayed into air under ambient conditions as described in Ref. 10. The mass spectra are obtained by passing the ions through the first quadrupole in RF-only mode, and through the second quadrupole mass filter in resolving mode, detecting the ion current with a channel electron multiplier.

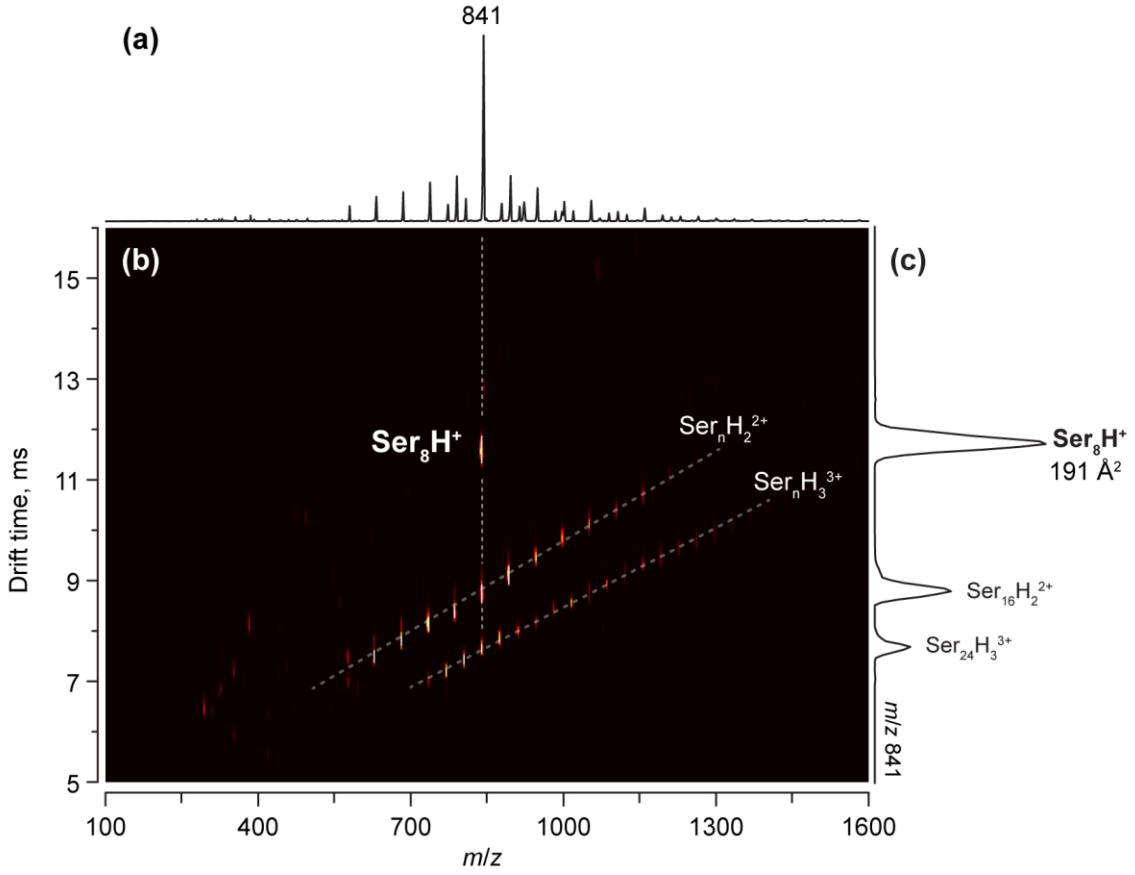

Figure S2. (a) nESI-mass spectra of the aqueous L-serine solution (5 mM), (b) drift time versus *m*/*z* plot of the protonated serine clusters, and (c) arrival time distribution (ATD) of *m*/*z* 841 ions which contain the protonated serine octamer (Ser_8H^+). All these results are obtained in the drift tube hybrid quadrupole time-of-flight mass spectrometer for infrared multiple photon dissociation (IRMPD) spectroscopy. Instrumental details are given in Ref. 11. Generated cluster ions were separated in the drift tube by their geometrical sizes. This 2-D plot was obtained by passing cluster ions through the first and second quadrupoles in RF-only mode and varying the timing of a high-voltage pulse (0.1 ms step size in the drift time range of 5–16 ms) which pulsed a fraction of ion mobility-separated ions into an orthogonal TOF analyzer. Drift time-selected mass spectra are summed to build the total mass spectrum, and the ATD was built by extracting the drift time-dependent TOF signal of specific *m*/*z* ions. In this figure, the determined collision cross section (CCS) value of Ser_8H^+ is also given.

3. IR spectra of serine octamers with various isotopic substitution

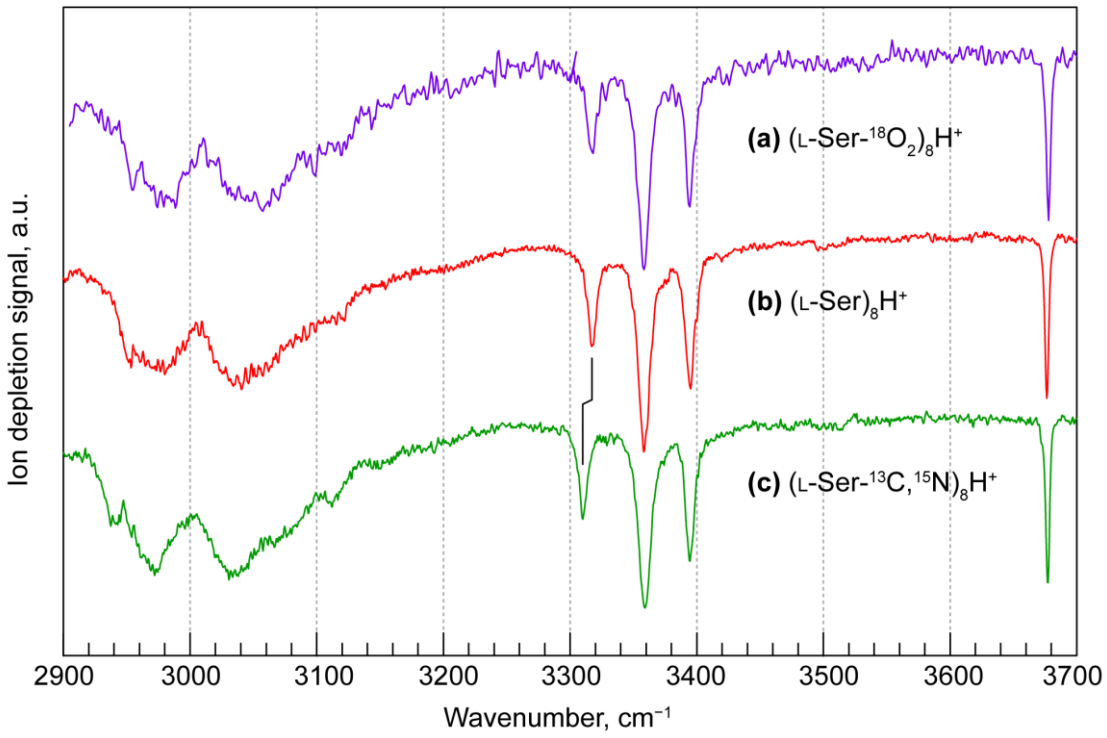

Figure S3. Cold ion infrared spectra of the serine octamers with various isotopic substitutions, (a) (L-Ser-$^{18}O_2)_8H^+$, (b) $(L\text{-Ser})_8H^+$ and (c) (L-Ser-$^{13}C_3$,$^{15}N_1)_8H^+$ in the 2900–3700 cm^{-1} region.

One can see in Figure S3c that the band at 3317 cm^{-1} shifts by 8 cm^{-1} upon labeling the serine nitrogen with ^{15}N while all other bands appear at the same wavenumber, indicating that only this one band belongs to NH oscillators. The lack of shifts in the spectrum of Figure S3a upon ^{18}O substitution of the carboxylic acid oxygens indicates that the remaining sharp lines belong to side chain OH groups.

4. Dynamic behavior of the cluster

In addition to the structures A and B (shown in Figure 2 and 3), several more conformers were obtained with relative 0 K energies between these two conformers (A and B). These conformers, presented in Figure S4, belong to one conformational family and mainly differ in the rotamer orientation of serine β leading to a slightly different H-bond pattern.

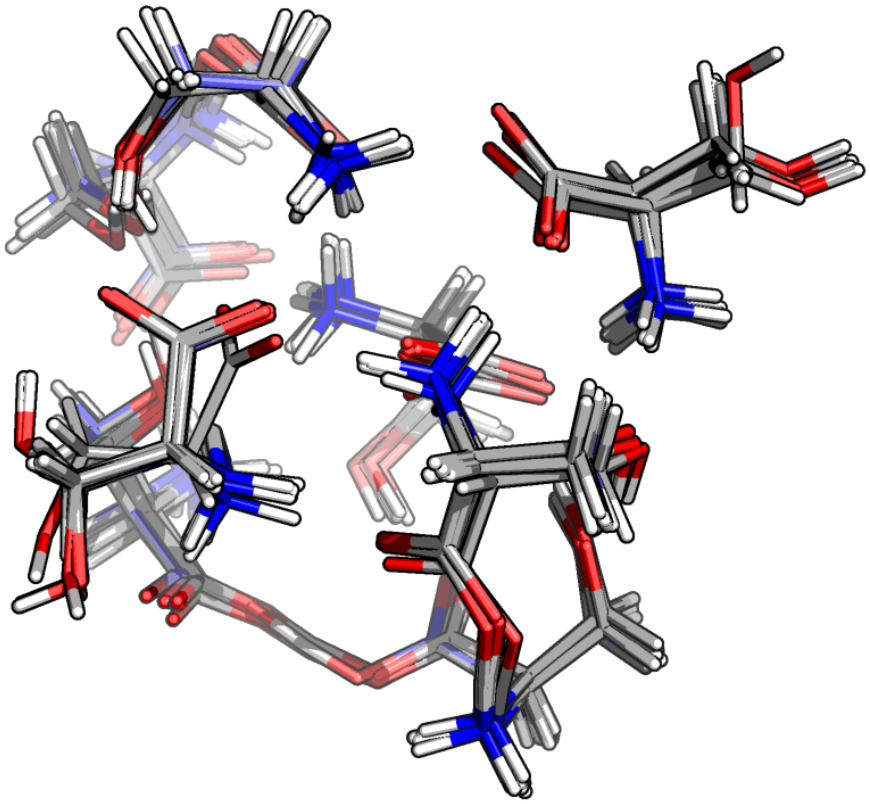

Figure S4 Twelve additional low-energy conformers that slightly differ at B3LYP/6-31G(d,p) level. These conformers are similar to the structure A, and each of their serines β forms a loose hydrogen bond with a carboxylate.

Ab initio MD (AIMD) allows gaining some insight into the finite temperature behavior of the cluster. Regarding the evolution of the H-bond characteristics of the cluster, we observe that the $–CH_2OH$ hydrogen-bond pattern of serine α stays the same throughout the simulations at room temperature. In contrast, serine β displays local structural fluctuations in the $–CH_2OH$ group,

occasionally forming short-lived –CH_2OH···O bonds with neighboring atoms. AIMD simulations of the cluster at 300 K with B3LYP/6-31G(d,p) show that the occurrence of CH_2OH ···O bonds in serine β is 24%. Although at 0 K the conformers with the serine β side chain hydrogen bonded are lower in energy, this is not the case at 300 K. Hydrogen bond distances for serine β during the *ab initio* MD at 300 K are presented in Figure S5.

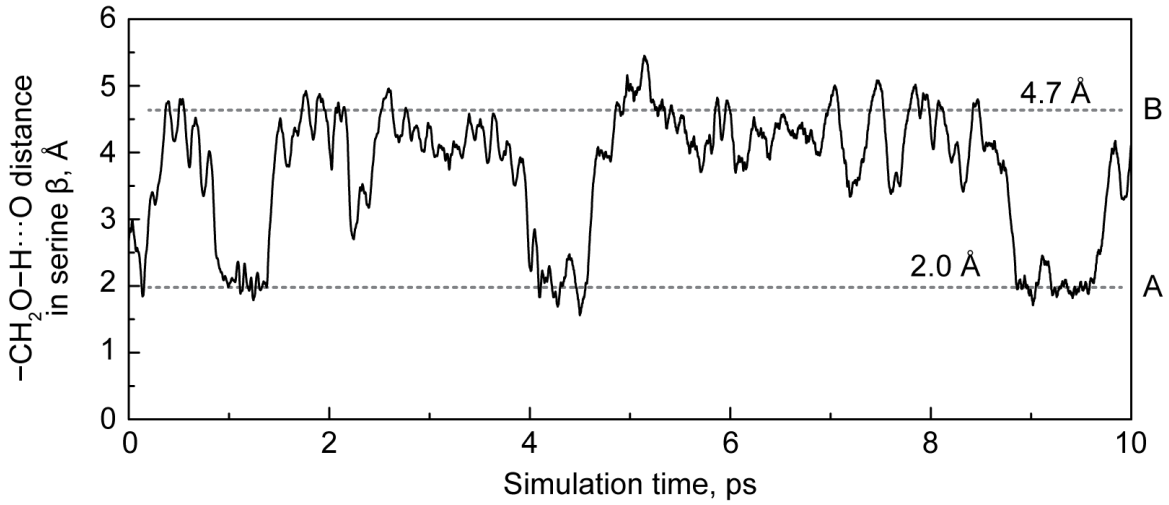

Figure S5. The –CH_2OH···O distance in serine β during the *ab initio* MD simulation with B3LYP/6-31G(d,p) level of theory at 300 K. The –CH_2OH···O distances in the structures A and B (2.0 and 4.7 Å, respectively) are shown as dashed lines.

5. Heterochiral substitution of Ser_8H^+

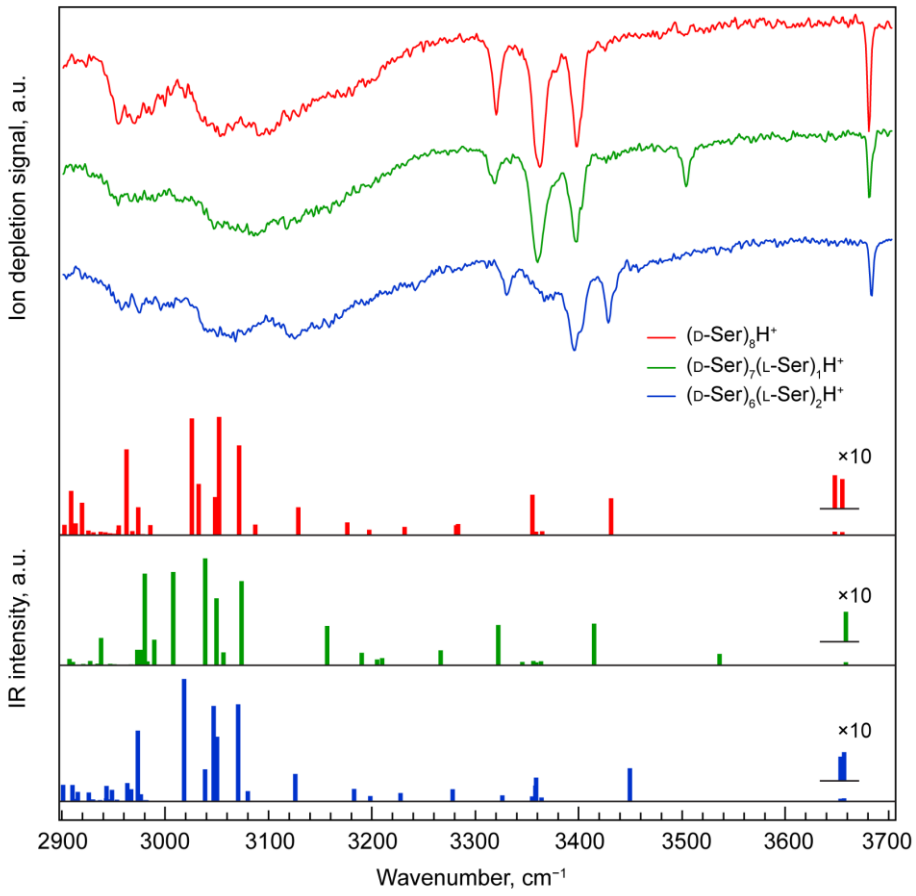

Figure S6. Chiral replacement in the serine octamer. D-serine was mixed with isotopically labeled L-serine (L-serine-$^{18}O_2$) for isolating the $(L\text{-}Ser)_7(D\text{-}Ser)_1H^+$ and $(L\text{-}Ser)_6(D\text{-}Ser)_2H^+$ species by *m*/*z*-selection.

$(L\text{-}Ser)_7(D\text{-}Ser)_1H^+$ consists in the replacement of L-serine α by D-serine. The side chain of D-serine is then weakly H-bonded, which justifies the appearance of the peak at 3534 cm^{-1}. The N–H stretch mode of the D-serine is less H-bonded and appears blue shifted to 3344 cm^{-1}, justifying the shoulder of the band experimentally observed at 3358 cm^{-1} and the decrease in intensity of the band at 3317 cm^{-1}.

The molecular structure of $(L\text{-}Ser)_6(D\text{-}Ser)_2H^+$ consists in the replacement of subunits α and β by two D-serines. Energetically $(L\text{-}Ser)_7(D\text{-}Ser)_1H^+$ is ~1 kcal mol^{-1} less stable than structure A at 300K and $(L\text{-}Ser)_6(D\text{-}Ser)_2H^+$ is already ~4 kcal mol^{-1} less stable. Some variations in the H-bond pattern occur.

6. Substitution of a serine with an alanine or a cysteine, and full deuteration of serine octamer

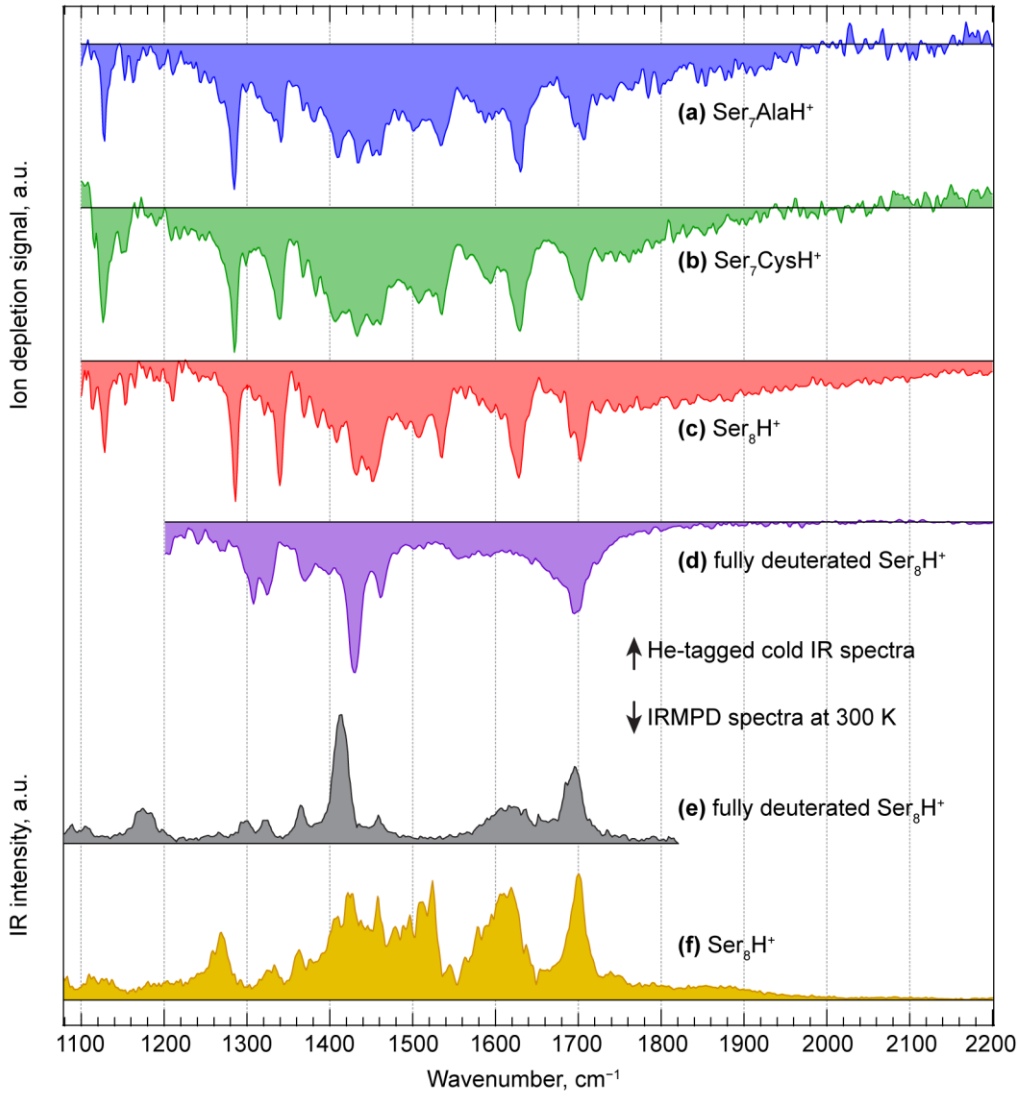

Figure S7. The cold ion spectra of He-tagged octamer clusters at 1000–2200 cm^{-1} region with various substitutions; (a) a serine substituted by an alanine (Ser_7AlaH^+), (b) a serine substituted by a cysteine (Ser_7CysH^+), (c) without substitution (Ser_8H^+), and (d) all 33 labile protons substituted by deuteriums (Ser_8H^+ − 33H + 33D); and the IRMPD spectra of octamer clusters with (e) full deuteration and (f) no substitution.

7. Comparison between Ser_8H^+ and Ser_9H^+

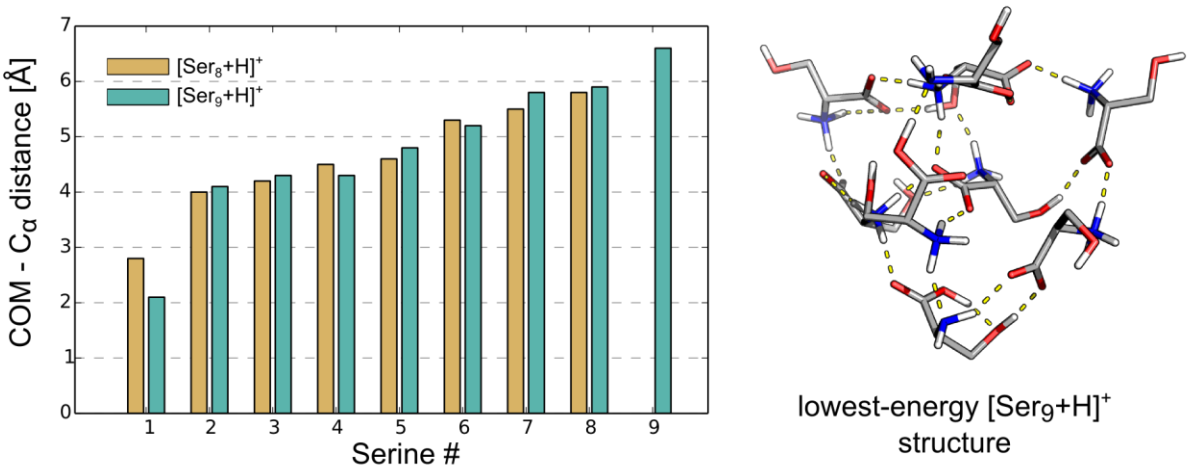

Figure S8. Distances from C_α of each serine to the center of mass of the cluster presented in brown for Ser_8H^+ and in green for Ser_9H^+. The lowest-energy structure of Ser_9H^+ obtained with SA-AIMD is shown on the right.

8. Structures and Energies of the Species (L-Ser)$_7$(D-Ser)$_1$H$^+$

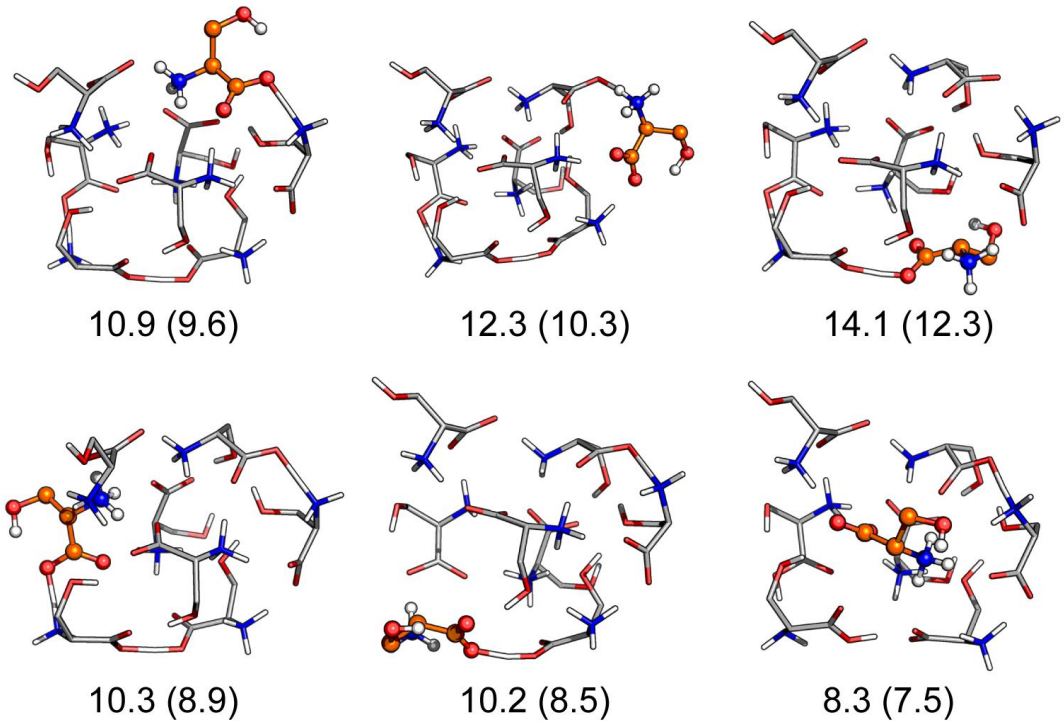

Figure S9. Structures of (L-Ser)$_7$(D-Ser)$_1$H$^+$ with a D-serine in locations other than positions α and β. The Δ*E* + ZPE and Gibbs free energy (in parentheses) at 300 K relative to the homochiral cluster A are also given (unit: kcal mol^{-1}). The D-serine is represented with ball-and-stick model and colored by orange. The sidechain OH group of substituted D-serine is exposed outside and thus is not involved in the three-point hydrogen bonding network.

REFERENCES

1. Abraham, M. J.; Murtola, T.; Schulz, R.; Pll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
2. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
3. Lei, H.; Duan, Y., Improved sampling methods for molecular simulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 187–191.
4. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
5. Tkatchenko, A.; Scheffler, M., Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102* (7), 73005.
6. Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M., Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **2009**, *180*, 2175–2196.
7. Schubert, F.; Rossi, M.; Baldauf, C.; Pagel, K.; Warnke, S.; von Helden, G.; Filsinger, F.; Kupser, P.; Meijer, G.; Salwiczek, M.; Koksch, B.; Scheffler, M.; Blum, V., Exploring the conformational preferences of 20-residue peptides in isolation: Ac-Ala $_{19}$ -Lys + H^+ vs. Ac-Lys-Ala $_{19}$ + H^+ and the current reach of DFT. *Phys. Chem. Chem. Phys.* **2015**, *17*, 7373–7385.
8. Chemical Computing Group, U. L. C., Molecular Operating Environment, 2013.08. **2017**.
9. Labute, P., LowModeMD—Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops. *J. Chem. Inf. Model.* **2010**, *50*, 792–800.
10. Svendsen, A.; Lorenz, U. J.; Boyarkin, O. V.; Rizzo, T. R., A new tandem mass spectrometer for photofragment spectroscopy of cold, gas-phase molecular ions. *Rev. Sci. Instrum.* **2010,** *81*, 073107.
11. 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., Protomers of Benzocaine: Solvent and Permittivity Dependence. *J. Am. Chem. Soc.* **2015**, ***137***, 4236–4242.