Supplementary Information

Multiple gas phase conformations of proline-containing peptides: Is it always cis/trans isomerization?

Christopher B. Lietz,[a] Zhengwei Chen,[a] Chang Yun Son,[a] Xueqin Pang,[b] Qiang Cui,[a] and Lingjun Li[a,b]

[a]Department of Chemistry, University of Wisconsin-Madison, 1101 University Avenue, Madison, WI, 53706

[b]School of Pharmacy, University of Wisconsin-Madison, 777 Highland Avenue, Madison, WI, 53705

1. Ion mobility – mass spectrometry

All ion mobility – mass spectrometry (**IM-MS**) was conducted on the Synapt G2 HMDS (Waters Corporation, Milford, MA, USA) in positive mode with nitrogen as the IM buffer gas. Electrosprayed peptides were synthesized at the Peptide Synthesis Facility of the University of Wisconsin Biotechnology Center (Madison, WI, USA). Analytes were dissolved in 50:50 water: methanol to a concentration of 30 μM and injected into mass spectrometer by direct infusion at the flow rate of 0.4 μL/min. DL-Polyalanine (Sigma-Aldrich, St. Louis, MO, USA) was suspended in 50% ACN, 1% HAc at 0.1 mg mL^{-1} and subsequently used for CCS calibration. Calibration was performed with DriftScope and custom *pepCCScal* software that has been characterized previously.[1] Methods are based off the calibrations described by Bush et al. and used the helium CCS values for calibrants.[2] The IM-MS data of both the analyte and calibrant ions were acquired under five different wave velocity (**WV**, m s^{-1})/wave height (**WH**, V) ratios: 500/30, 500/35, 600/35, 700/40 and 800/40. CCS values listed in figure and the main text are the mean values calculated from these settings. Ion transmission settings for the spray voltage, cone voltage,

extraction cone, and helium cell DC were kept at 1.5 kV – 2.0 kV, 20 V, 4 V, and 15 V, respectively. Trap bias voltage was usually kept at 35 V – 38 V, except when increased to 60V to induce collisional activation.

We would like to address our choice of using the SYNAPT G2 and N_2 as the buffer gas. As explained by Bush et al.,[2] measuring ATDs in N_2 and using the He-CCS values of polyalanine attempts to account for systematic differences between N_2- and He-CCS and allow the calculation of calibrated He-CCS for unknowns. We acknowledge that there is no fundamental theoretical basis for this, and that recent studies have experimentally identified differences in analyte/gas collisions in N_2 and He.[3] However, we believe the empirical success by Bush et al. and in our own previous studies[1] justifies its use in the current investigation.

2. Molecular dynamics and structure clustering

Initial structures without hydrogen atoms were created using Avogadro and consisted of the WT and L7P sequences in an ideal linear configuration. Trans-Pro7 peptide ω- and φ-dihedrals were set to 180º and 149º, respectively. Cis-Pro7 peptide ω- and φ-dihedrals were set to 0º and 158º, respectively. MD was then carried out using GROMACS (version 4.6.5) on the University of Wisconsin-Madison Department of Chemistry Phoenix cluster (NSF Grant CHE-0840494).

Force field parameters are taken from AMBER force field with improved side chain torsion potential (AMBER ff-99SB-ILDN).[4] For non-standard charge states of amino acids in 3+ and 4+ proteins – protonated C-terminal Ala, protonated Gln and backbone-protonated dipeptide Ala24-Leu25 of L25H – parameters are developed following the AMBER force field parameterization procedure. Bonded interaction parameters and nonbonded Lennard Jones parameters are inherited from the existing parameters of equivalent atom types. Hartree-Fock electrostatic potentials (ESP) are calculated with the 6-31G* basis set[5-6] using Gaussian 09 program. Multiple conformation

Restrained ElectroStatic Potential (RESP) fitting[7] is used to fit the partial atomic charges to the obtained ESP data. The derived charges are shown in figure S5.

Hydrogens were added to the initial structures and charging protons were placed on the N-term, Lys5, and Arg6. For simulations of N4H structures, a charging proton was placed on the amide of Asn4. For simulations of L25H structures, a charging proton was placed on the backbone amide of Leu25.

Linear structures were energy-minimized to a tolerance of 10.0 kJ mol^{-1} nm^{-1} (this tolerance was used for all energy minimizations) and then subjected to simulated annealing for 1.55 ns. The cycle began at 300 K and increased by 50 K increments to 1000 K and then back down to 300 K. Increments happened over 9 ps and the system equilibrated for 2 ps after each increment. The final 300 K structure was energy-minimized and used as the "random" starting structure for replica-exchange molecular dynamics (**REMD**).

Each random starting structure was assigned to 34 parallel simulations with the following reference temperatures: 80.00 K, 90.65 K, 101.35 K, 113.45 K, 126.51 K, 140.24 K, 154.64 K, 169.74 K, 185.59 K, 202.26 K, 219.79 K, 238.22 K, 257.62 K, 278.05 K, 299.57 K, 322.25 K, 346.17 K, 371.39 K, 397.97 K, 426.01 K, 455.60 K, 586.80 K, 519.75 K, 554.51 K, 591.23 K, 629.85 K, 670.74 K, 713.9 K, 759.51 K, 807.65 K, 858.48 K, 912.20 K, 968.91 K, and 1028.80 K. The reference temperatures were chosen with the help of free online tools (http://folding.bmc.uu.se/remd/). After allowing the starting structures to equilibrate to their reference temperatures for 200 ps, REMD simulations were carried out for 100 ns. An exchange was attempted every 100 fs. Every 22.5 ps, the current configuration in the 299.57 K simulation (referred to as the "300 K window" in the main text) was output and energy-minimized. When the simulations finished, each WT ensemble consisted of 4444 outputs and each L7P ensemble

consisted of 8888 outputs (from combining the outputs of cis-P7 and trans-P7 initial structures with a given charge configuration).

The theoretical CCS were calculated via the trajectory method (**TM**) using MOBCAL.[8] The program was compiled to run 2.5 x 10^7 trajectories per output at 300 K. Running MOBCAL TM calculations for 26664 structures was accomplished through the use of the Center for High Throughput Computing (**CHTC**) in Madison, WI. We were allotted an average of 2000 CPUs for parallel calculations, and at times were able to obtain over 4000 CPUs.

All structures with theoretical CCS that matched within ± 3% of experimental values were kept for conformation cluster analysis by the MaxCluster algorithm (http://www.sbg.bio.ic.ac.uk/~maxcluster/). Clustering was performed as an "All-vs-All" analysis within each CCS-filtered ensemble. Cis-L7P and trans-L7P structures were analyzed separately. Average-linkage clustering was performed by calculating the rmsd of the N, C_α, C, O, and C_β backbone atoms and adjusting the rmsd threshold until an initial cluster containing at least 1/3 of the CCS-filtered ensemble was obtained. The "representative" structure of a cluster is the centroid that has the lowest average rmsd when compared to every other structure in the cluster.

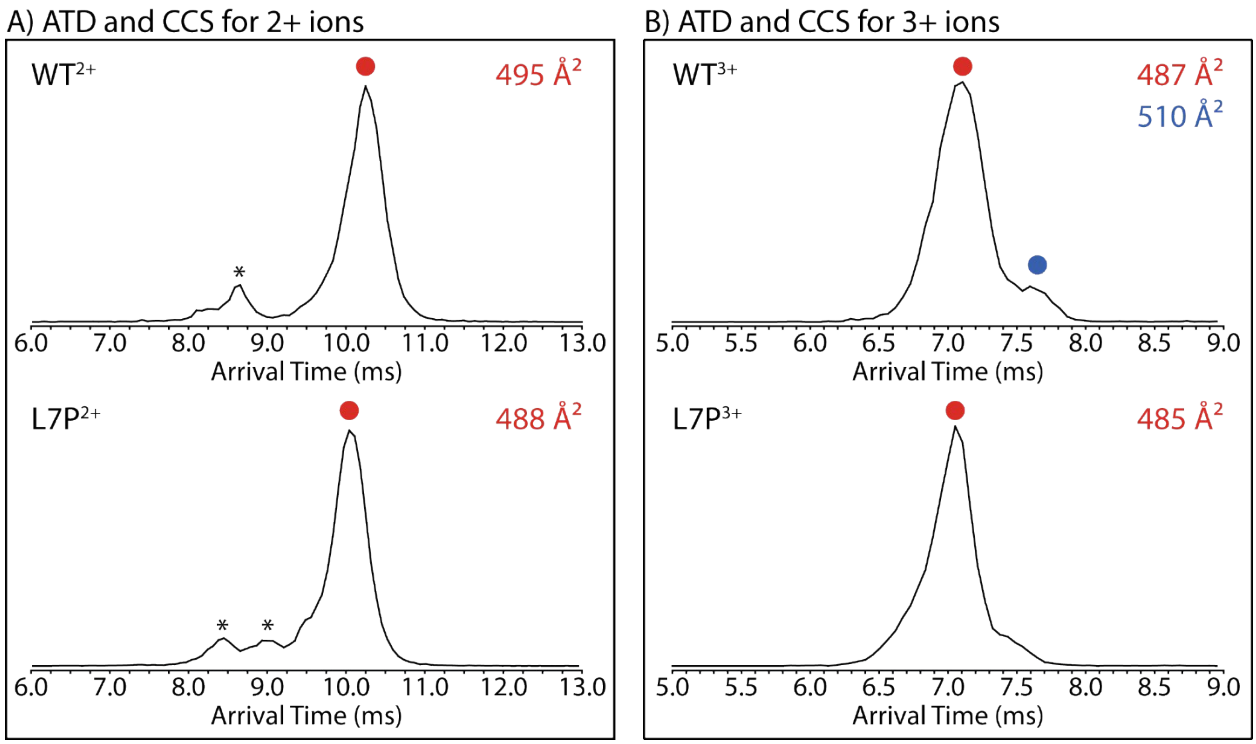


Figure S1. IM-MS ATDs and CCS for 2+ and 3+ ions. **A)** The ATDs and experimental calibrated CCS for WT^{2+} and $L7P^{2+}$ ions. Black asterisks denote peaks corresponding to 4+ dimers. The 2+ ATDs were acquired at wave height and wave velocity settings of 35 V and 500 m s^{-1}, respectively. **B)** The ATDs and experimental calibrated CCS for WT^{3+} and $L7P^{3+}$ ions. The 3+ ATDs were acquired at wave height and wave velocity settings of 30 V and 500 m s^{-1}, respectively. All reported CCS are mean values from acquisitions at wave height and wave velocity settings specified in **Supplemental Information 1**.

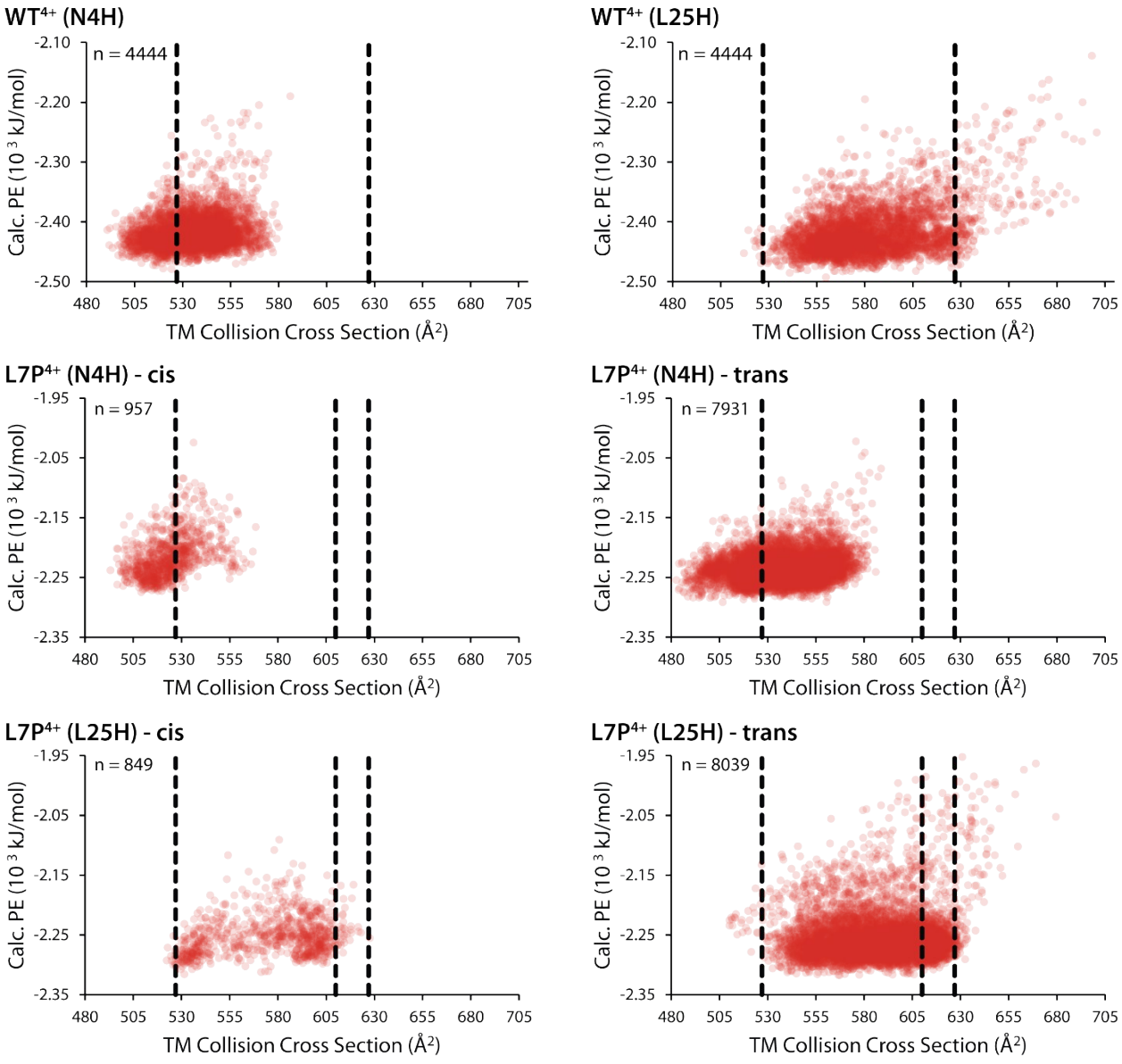


Figure S2. PE versus CCS. Plots of calculated potential energy of energy-minimzed structures versus theoretical CCS for all 300 K REMD outputs. Dotted lines denote the experimental calibrated CCS.

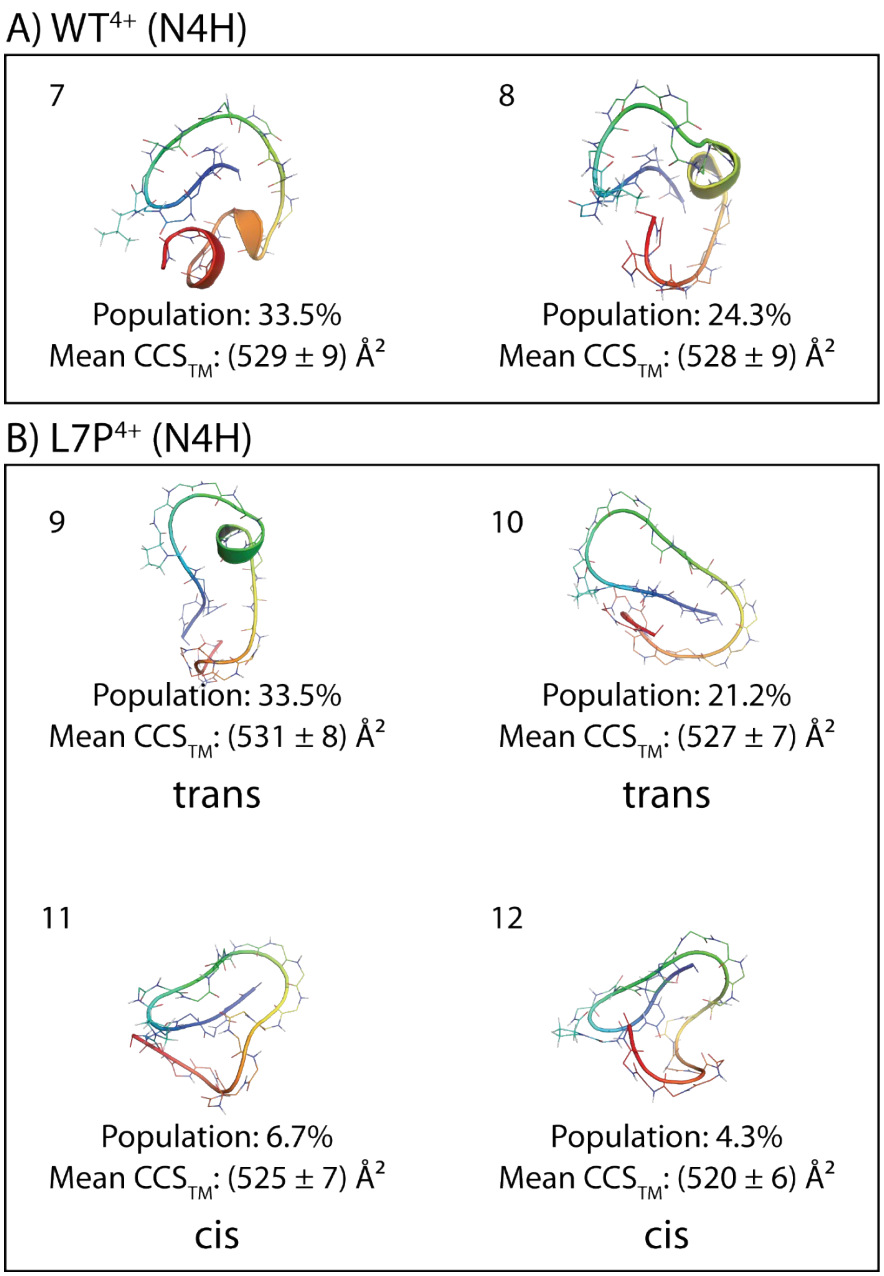
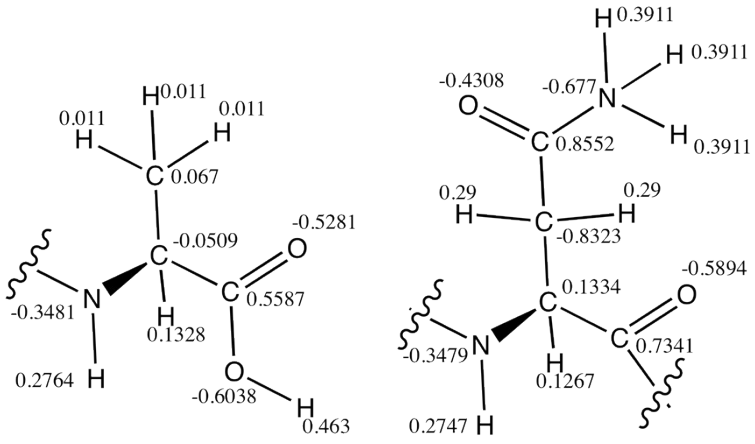
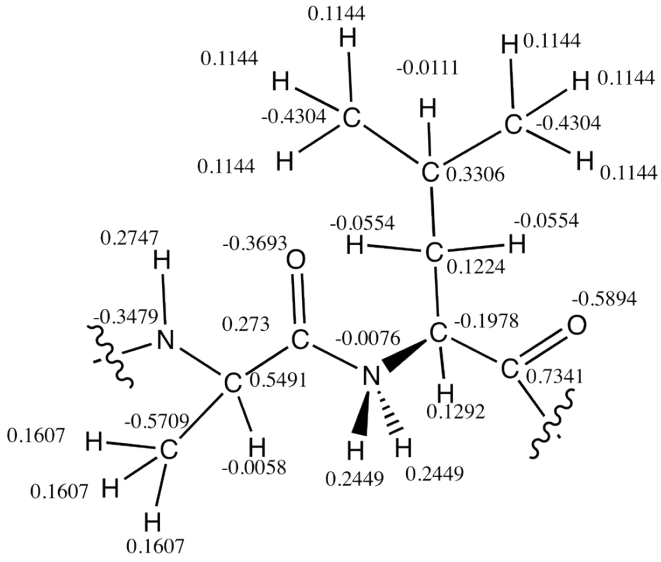


Figure S3. Candidate structures for N4H 4+ ions. Representative centroid structures from the highest population conformation clusters of **A)** WT^{4+} and **B)** $L7P^{4+}$ N4H ions. Atoms from the entire backbone and seventh residue's sidechain are displayed. Structures are color-coded to denote N-terminal (blue) and C-terminal (red) regions. Relative cluster population and mean theoretical CCS of the cluster are listed below the centroid structure. The total populations of the WT^{4+} and $L7P^{4+}$ N4H CCS-filtered ensembles were 2513 and 4565, respectively.



112

113



114

115

116 **Figure S4. Derived charges for custom amino acids.**

Supplemental References

(1) C. B. Lietz, Q. Yu, and L. J. Li, *Journal of the American Society for Mass Spectrometry*, 2014, **25**, 2009-2019.

(2) M. F. Bush, I. D. G. Campuzano, and C. V. Robinson, *Analytical Chemistry*, 2012, **84**, 7124-7130.

(3) C. Larriba-Andaluz, J. Fernandez-Garcia, M. A. Ewing, C. J. Hogan, and D. E. Clemmer, *Physical Chemistry Chemical Physics*, 2015, **17**, 15019-15029.

(4) K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror *et al.*, *Proteins-Structure Function and Bioinformatics*, 2010, **78**, 1950-1958.

(5) Ditchfie.R, W. J. Hehre, and J. A. Pople, *Journal of Chemical Physics*, 1971, **54**, 724-&.

(6) Harihara.Pc, and J. A. Pople, *Chemical Physics Letters*, 1972, **16**, 217-&.

(7) C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *Journal of Physical Chemistry*, 1993, **97**, 10269-10280.

(8) M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, and M. F. Jarrold, *The Journal of Physical Chemistry*, 1996, **100**, 16082-16086.