Exploring Structures and Dynamics of Protamine Molecules through Molecular Dynamics Simulations

Hossain Shadman[1], Caleb Edward Gallops[1], Jesse D. Ziebarth[1], Jason E. DeRouchey[2], Yongmei Wang[1]*

1: Department of Chemistry, The University of Memphis, Memphis, Tennessee 38154

2:Department of Chemistry, The University of Kentucky, Lexington, Kentucky 40506

Supplementary Information

Table S1: Technical details of MD simulations

	Number of Simulations	Water Model	Force Field	Box Size Prior to Starting Explicit Solvent Simulation	Time of Implicit Solvent Simulation	Time of Explicit Solvent Simulation
Salmon	13	TIP3P	*ff14SB*	73Å x 73Å x 73Å	Varied from ~40ns to ~320ns	200 ns
Bull P1	13	TIP3P	*ff14SB*	199Å x 199Å x 199Å	~200ns	200 ns
Human P1	13	TIP3P	*ff14SB*	210Å x 210Å x 210Å	~200ns	200 ns

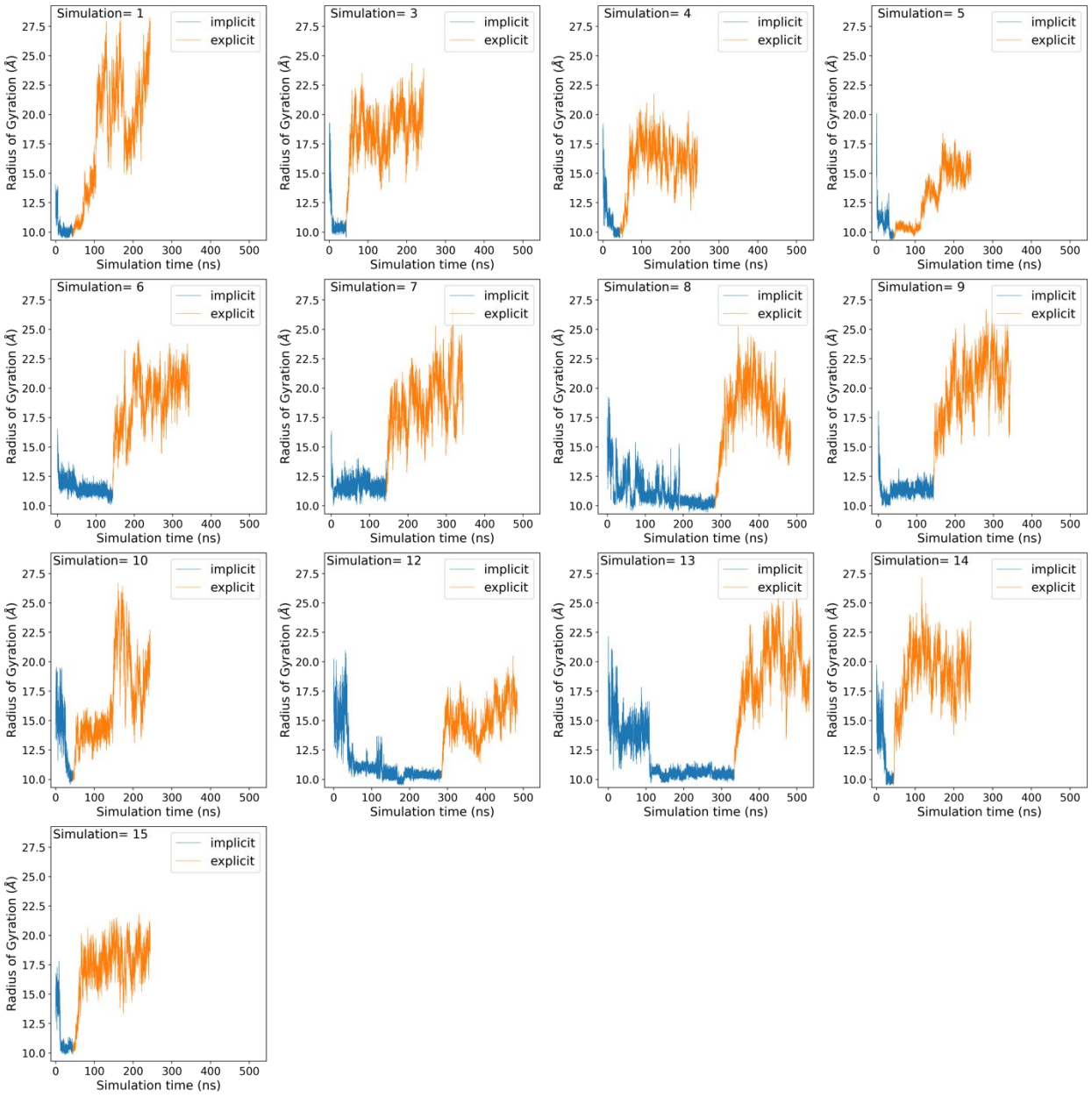

Figure S1: A simulation by simulation view of radius of gyration values for salmon protamine, starting from implicit solvent simulation (blue) to explicit solvent simulation (orange).

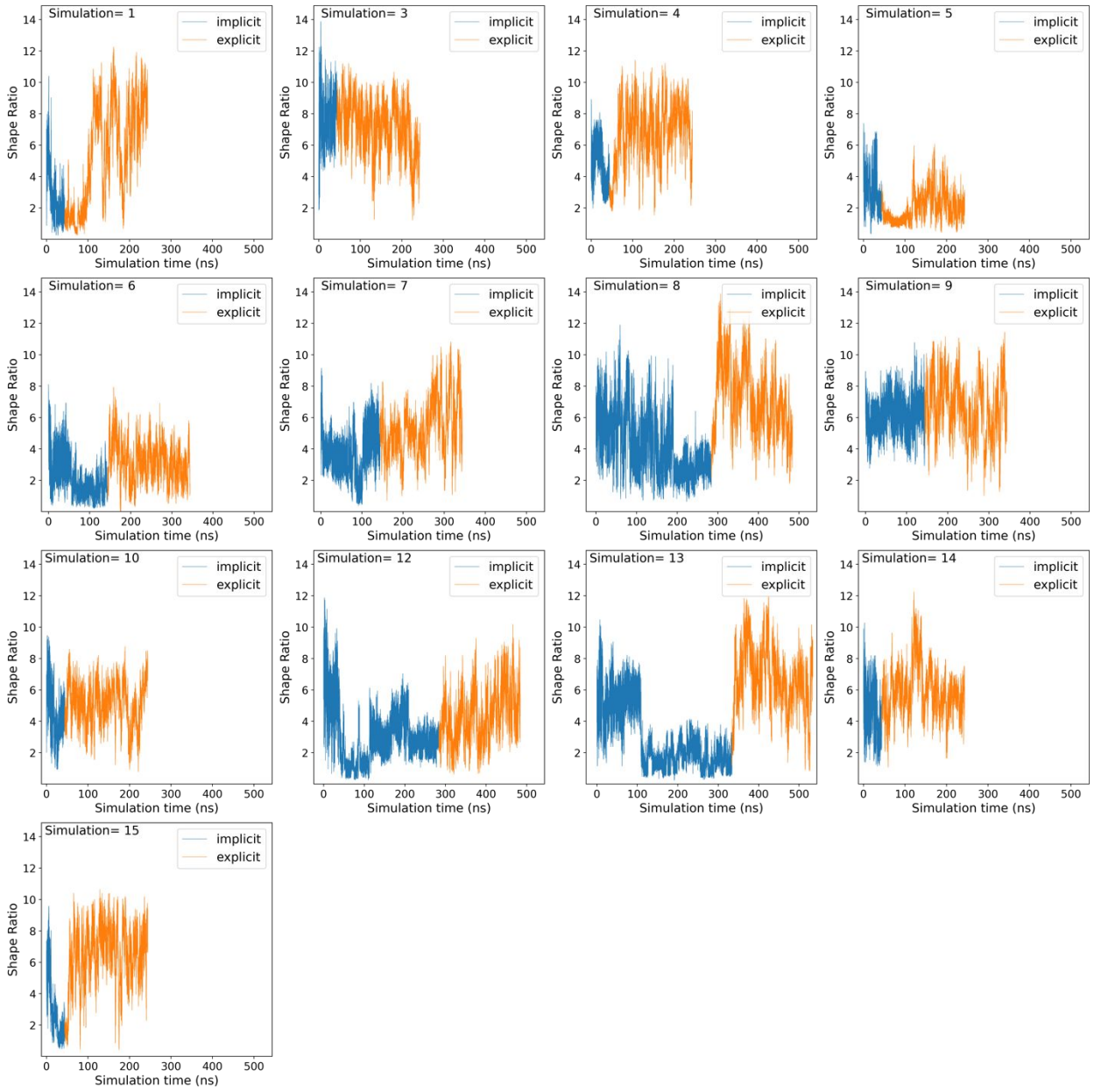

Figure S2: A simulation by simulation view of shape ratio values for salmon protamine, starting from implicit solvent simulation (blue) to explicit solvent simulation (orange)

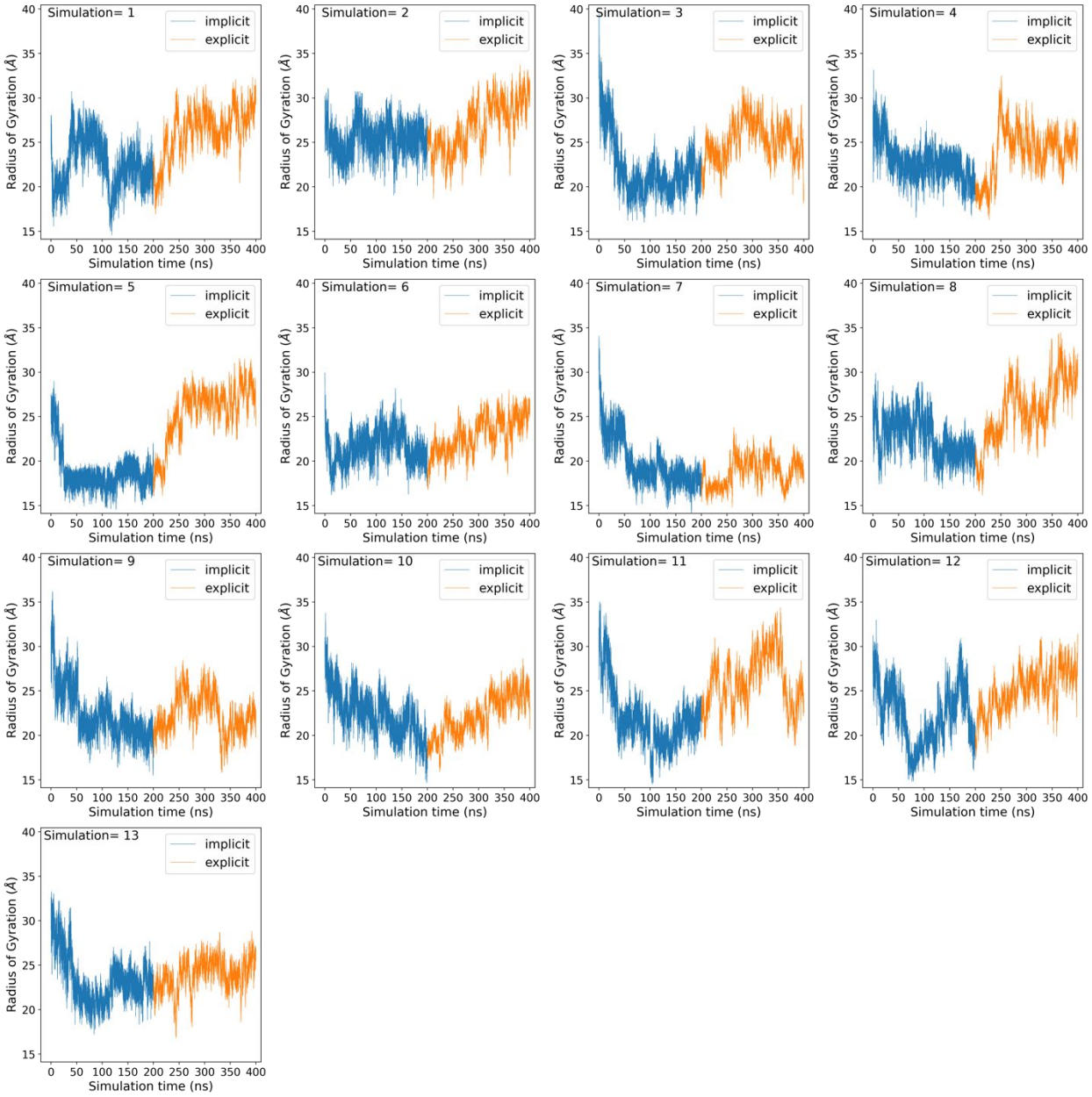

Figure S3: A simulation by simulation view of radius of gyration values for bull P1 protamine, starting from implicit solvent simulation (blue) to explicit solvent simulation (orange)

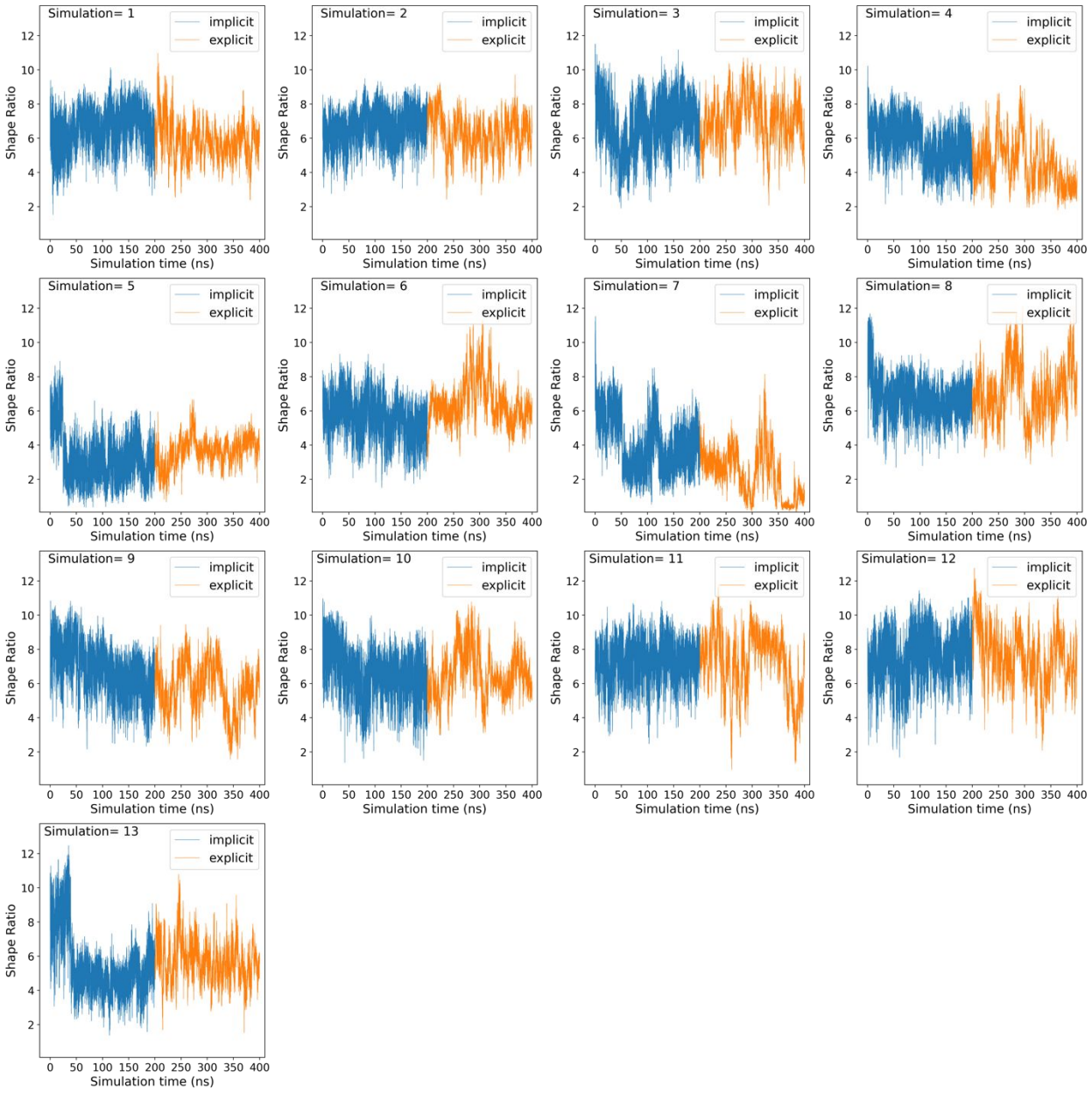

Figure S4: A simulation by simulation view of shape ratio values for bull P1 protamine, starting from implicit solvent simulation (blue) to explicit solvent simulation (orange)

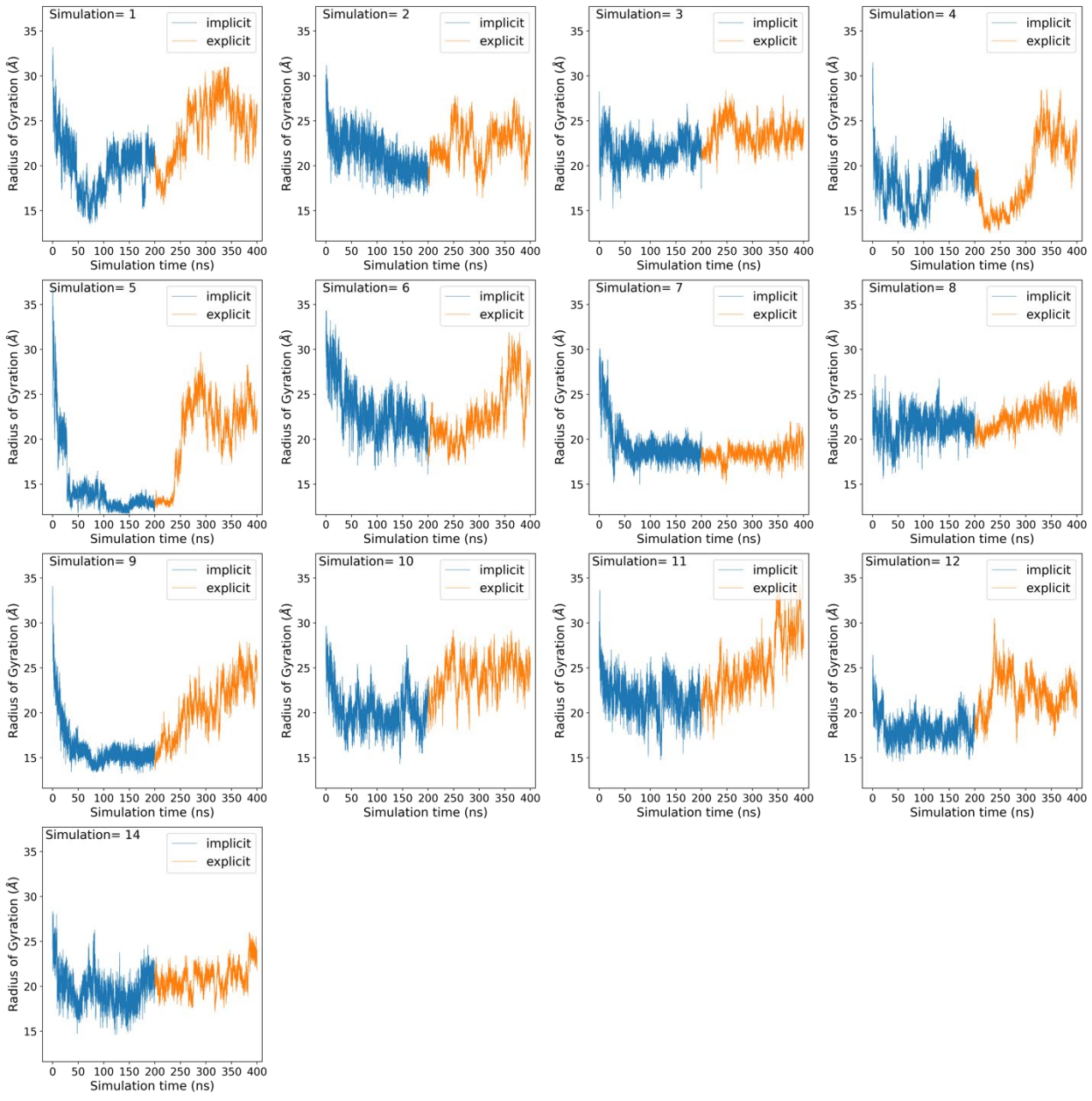

Figure S5: A simulation by simulation view of radius of gyration values for human P1 protamine, starting from implicit solvent simulation (blue) to explicit solvent simulation (orange)

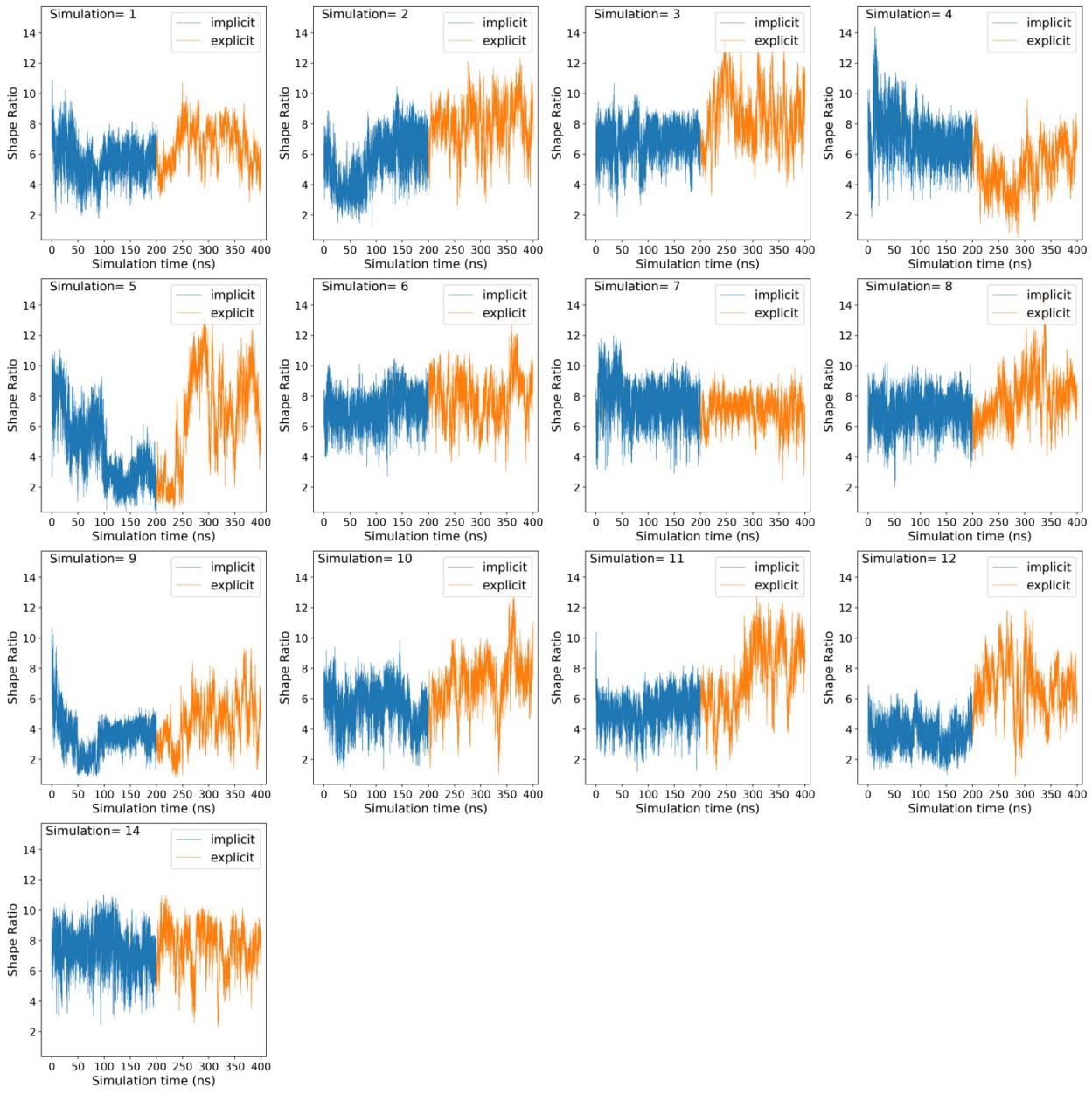

Figure S6: A simulation by simulation view of shape ratio values for human P1 protamine, starting from implicit solvent simulation (blue) to explicit solvent simulation (orange)

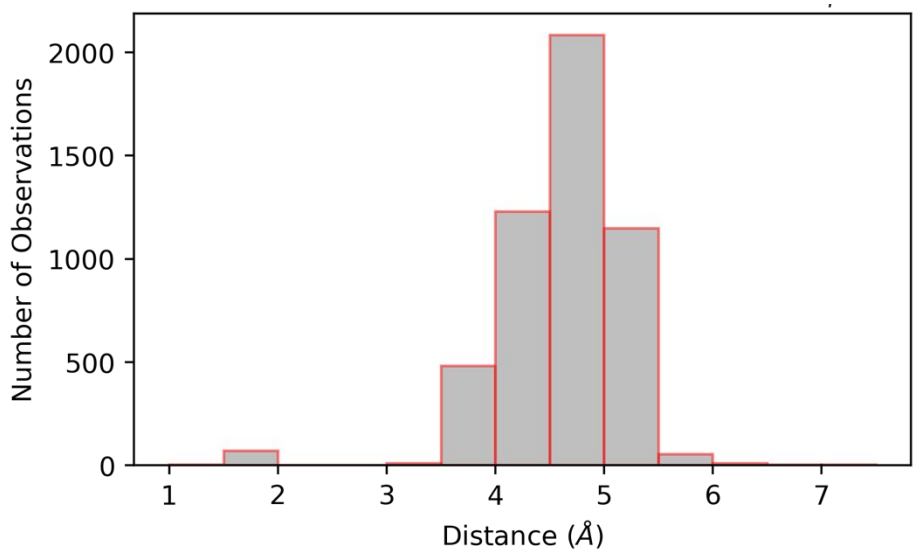

Figure S7: Histogram of inter-cysteine (C_α-C_β) distances in disulfide bonds, based on PISCES[1] dataset compiled and published by Gao *et al.*[2]

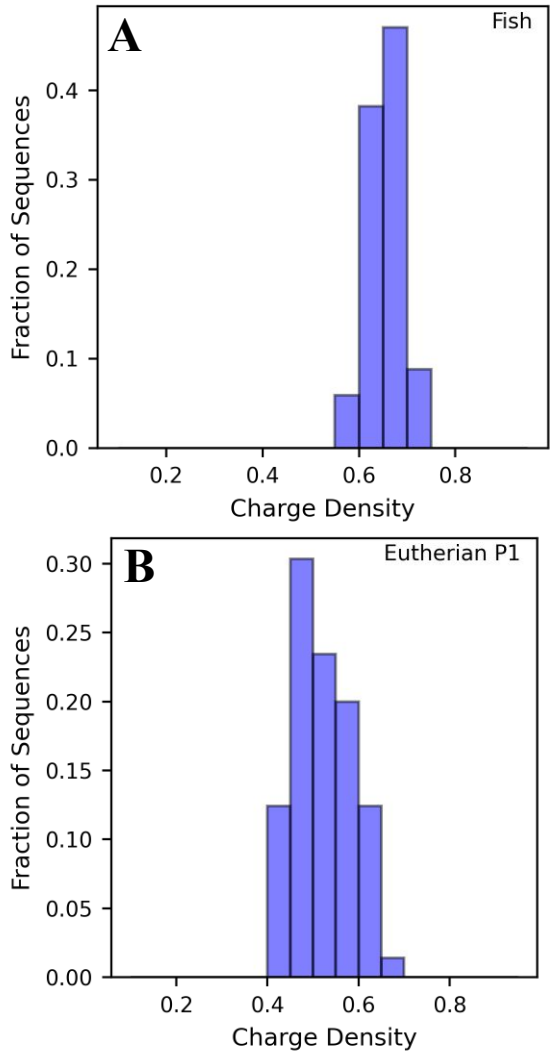

Figure S8: Charge densities of (A) fish and (B) eutherian P1 protamine sequences. Sequence dataset obtained from Powell *et al.*[3]

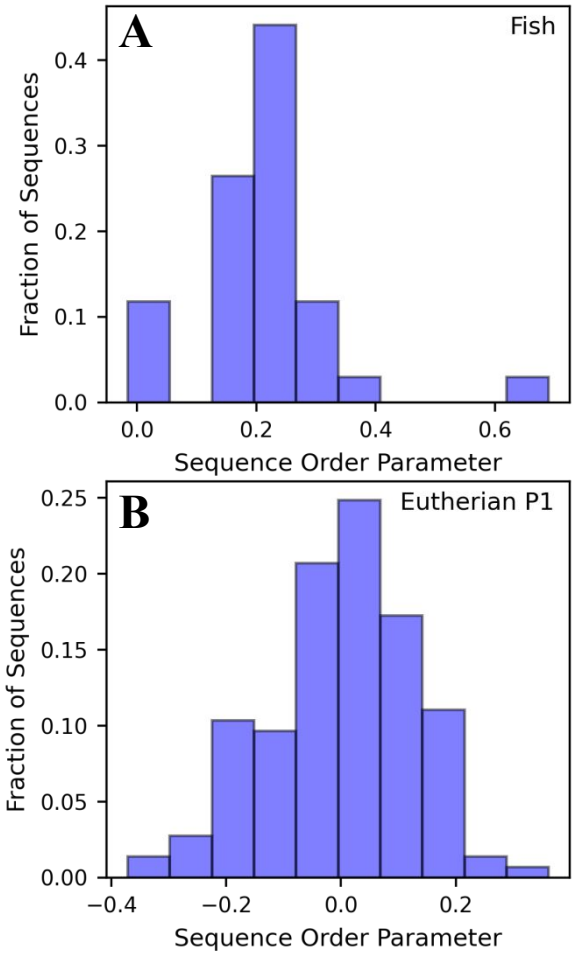

Figure S9: Distributions of sequence order parameter (λ) values, for (A) fish protamine and (B) eutherian P1 protamine sequences obtained from Powell *et al.*[3] λ values were calculated for each sequence by considering residues as either an arginine or a non-arginine. Fish protamines have an average λ value of 0.21 with standard deviation 0.12. Eutherian P1 protamines have an average λ value of 0.0015 with standard deviation 0.13

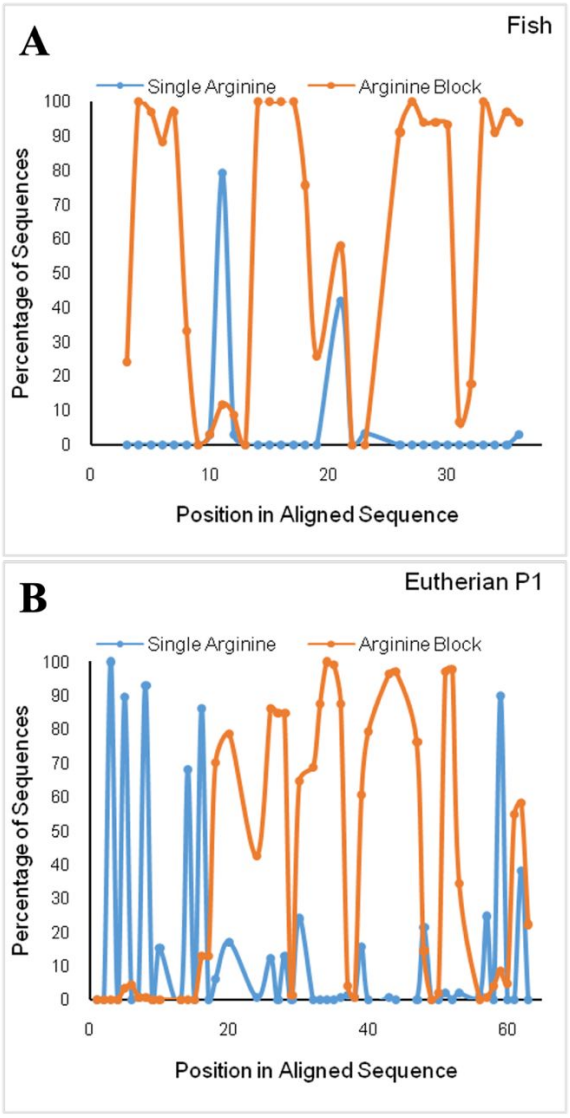

Figure S10: Locations of single arginines and grouped arginines present in (A) fish protamines and (B) eutherian P1 protamines. The positions are based on aligned sequences. In the aligned

sequences, an '-X-arg-X-' is identified as an instance of a single arginine, where X is a non-arginine residue. All other cases of arginine residues (i.e. stretches of 2 or more consecutive arginines) are identified as 'grouped' arginines (i.e. an arginine 'block'). Positions in aligned sequences found to be blank in more than 40% of sequences were disregarded. Aligned sequence dataset obtained from Powell *et al.*[3]

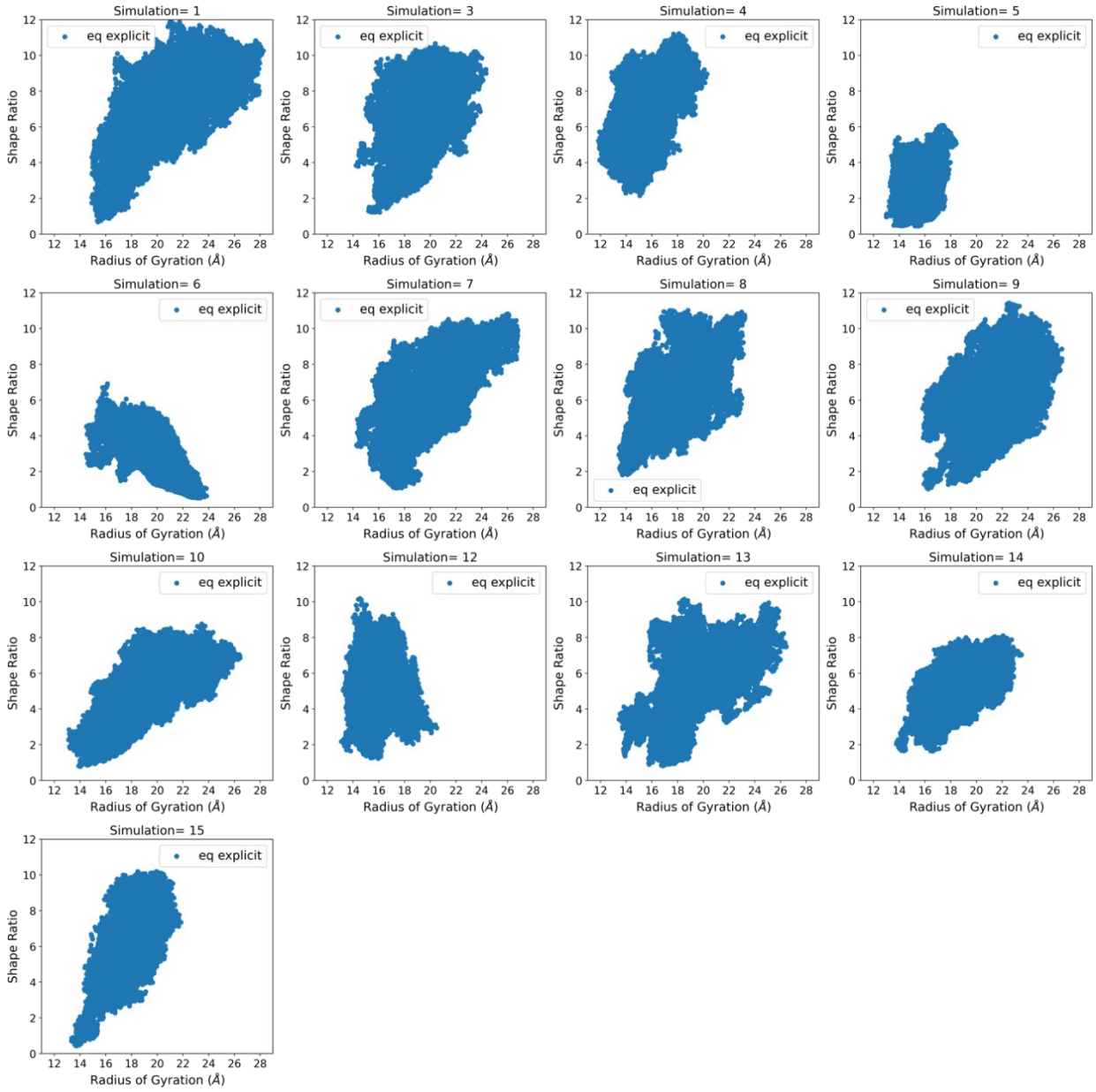

Figure S11: A simulation by simulation presentation of how shape ratio varies with radius of gyration for salmon protamine

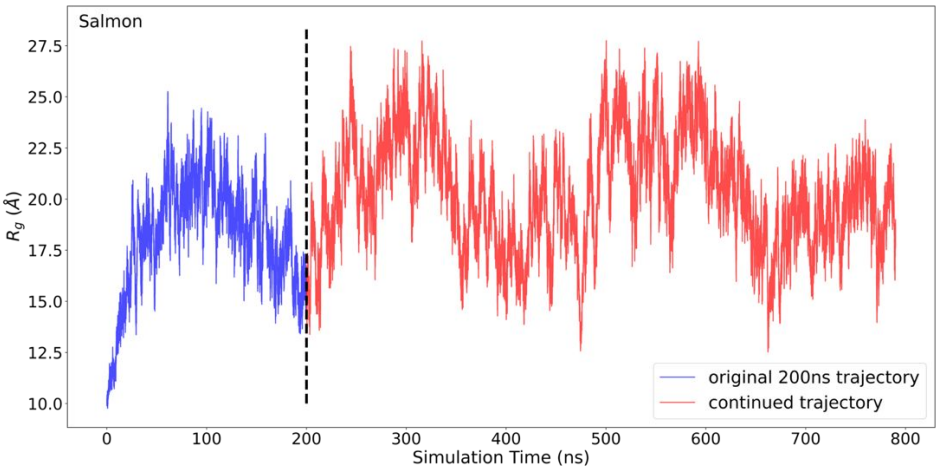

Figure S12: Radius of gyration (R_g) values for one of the 200 ns salmon explicit solvent trajectories (simulation #8) continued for an additional 590 ns, with the additional time shown in red

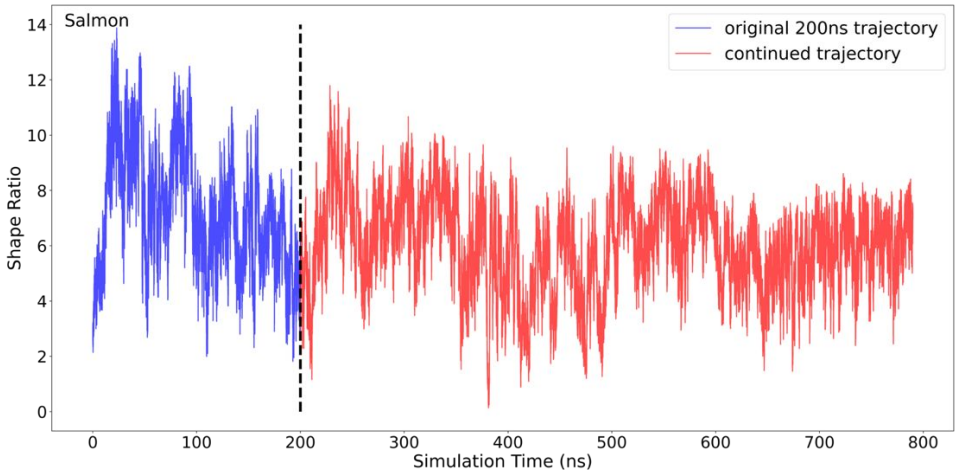

Figure S13: Shape ratio values for one of the 200 ns salmon explicit solvent trajectories (simulation #8) continued for an additional 590 ns, with the additional time shown in red

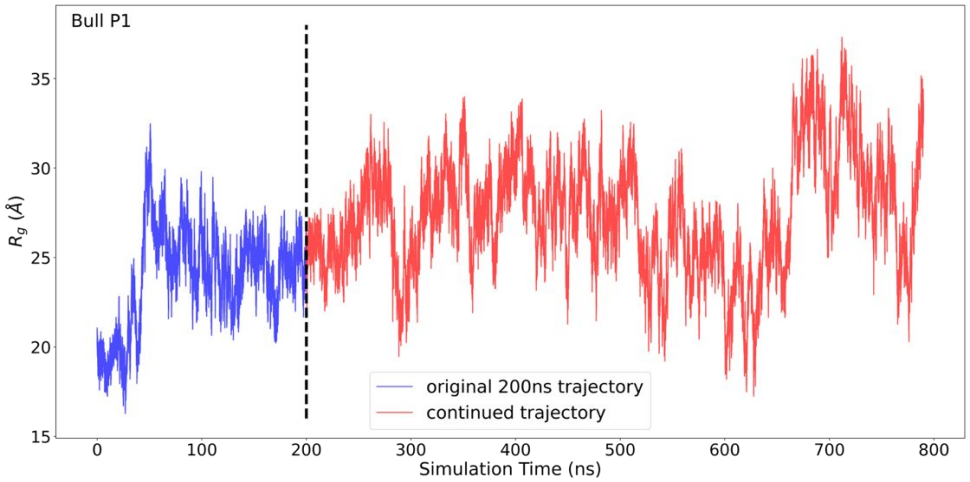

Figure S14: Radius of gyration (R_g) values for one of the 200 ns bull explicit solvent trajectories (simulation #4) continued for an additional 590 ns, with the additional time shown in red

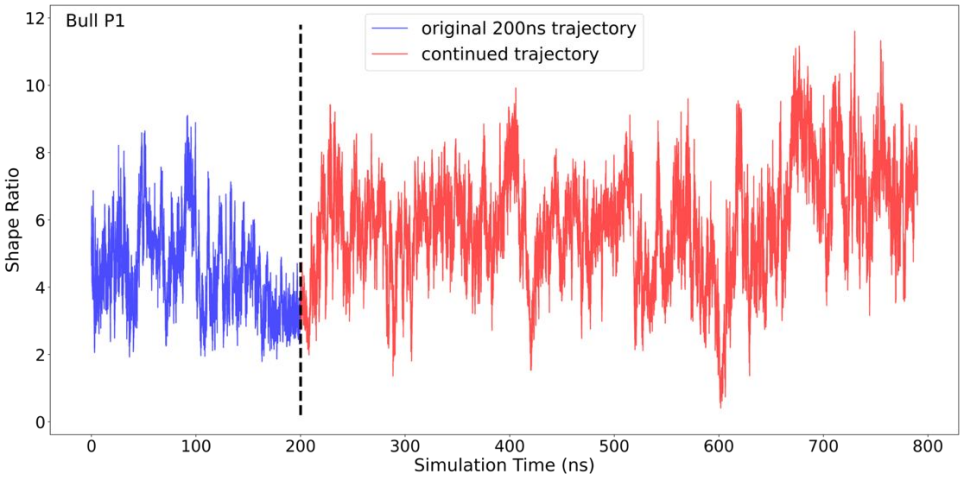

Figure S15: Shape ratio values for one of the 200 ns bull explicit solvent trajectories (simulation #4) continued for an additional 590 ns, with the additional time shown in red

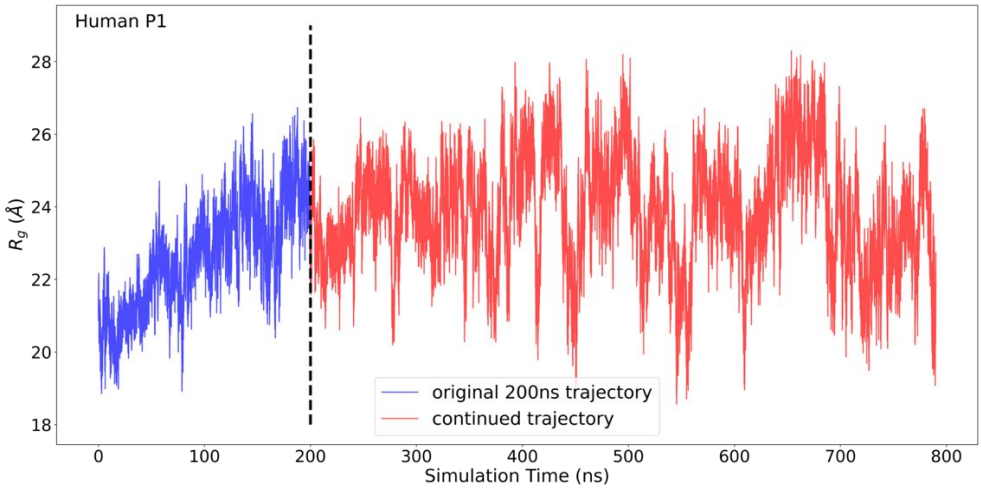

Figure S16: Radius of gyration (R_g) values for one of the 200 ns human P1 explicit solvent trajectories (simulation #8) continued for an additional 590 ns, with the additional time shown in red

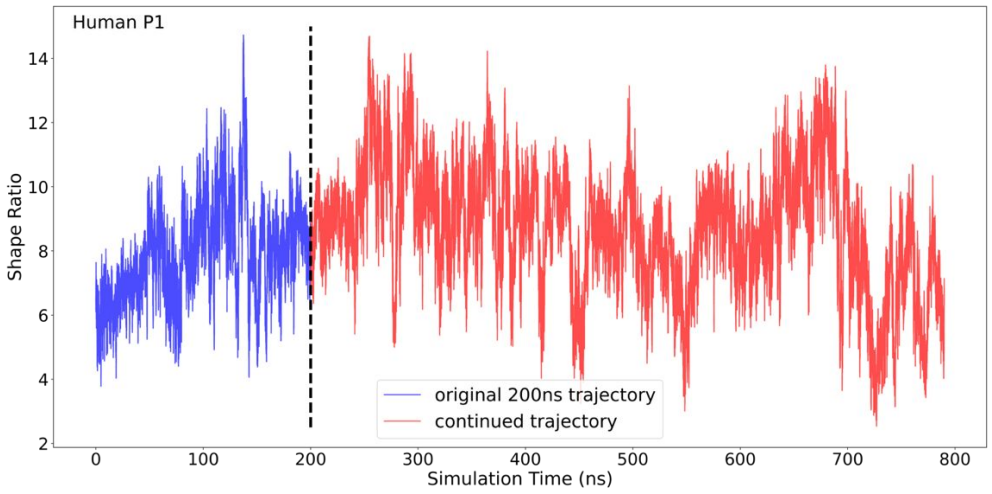

Figure S17: Shape ratio values for one of the 200 ns human P1 explicit solvent trajectories (simulation #8) continued for an additional 590 ns, with the additional time shown in red

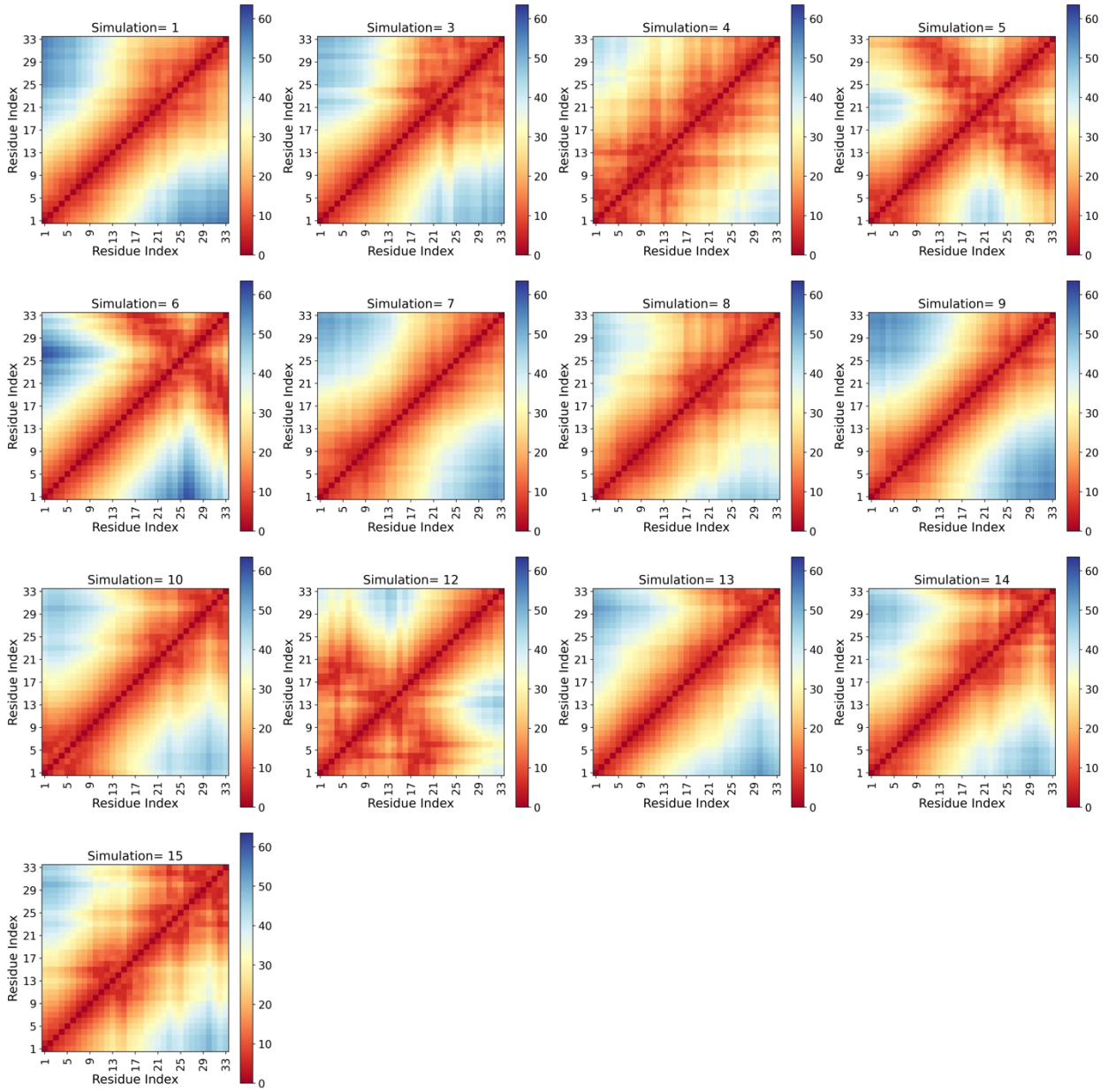

Figure S18: Inter-residue distance heatmap for every salmon trajectory. Figure 1 has representative conformation images for some of these simulations. For example, simulation #9 is represented by

image #1 in figure 1. Simulation #5 is represented by image #7 in figure 1. Distance values are in Angstrom

Figure S19: A simulation by simulation presentation of how shape ratio varies with radius of gyration for bull P1 protamine

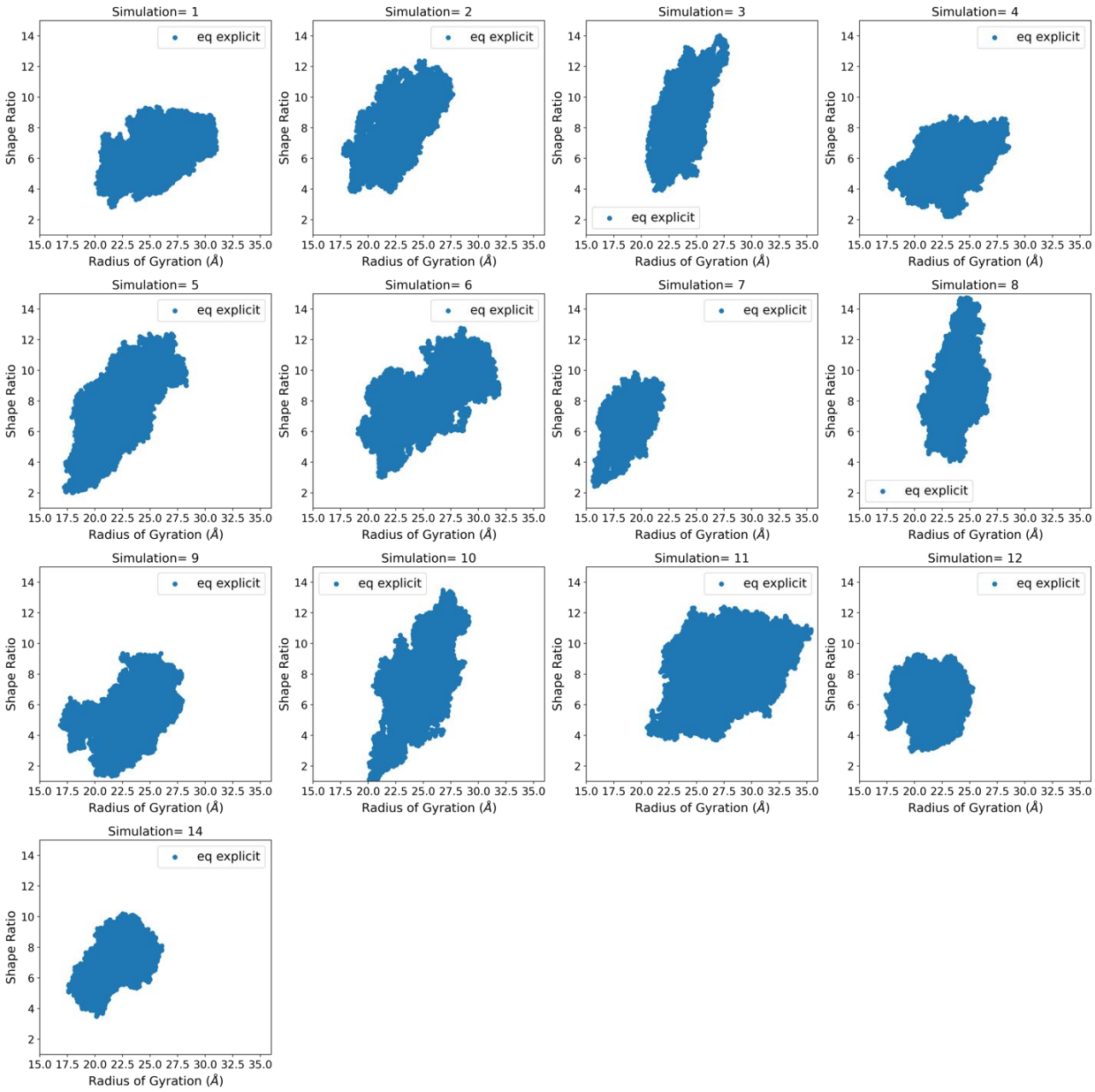

Figure S20: A simulation by simulation presentation of how shape ratio varies with radius of gyration for human P1 protamine

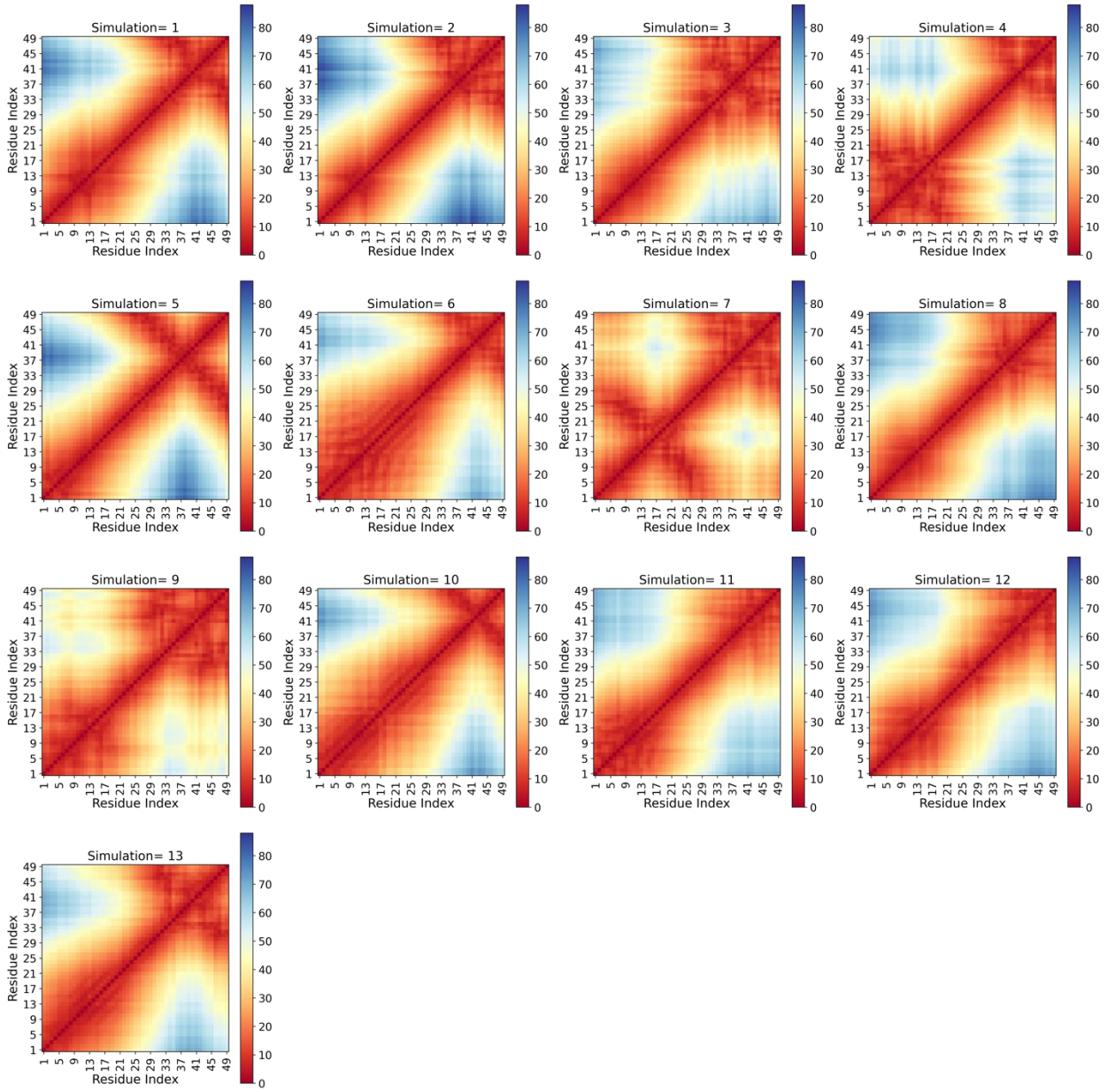

Figure S21: Inter-residue distance heatmap for every bull P1 trajectory. Figure 3 has representative conformation images for some of these simulations. For example, simulation #8 is represented by image #7 in figure 3. Simulation #5 is represented by image #2 in figure 3. Distance values are in Angstrom

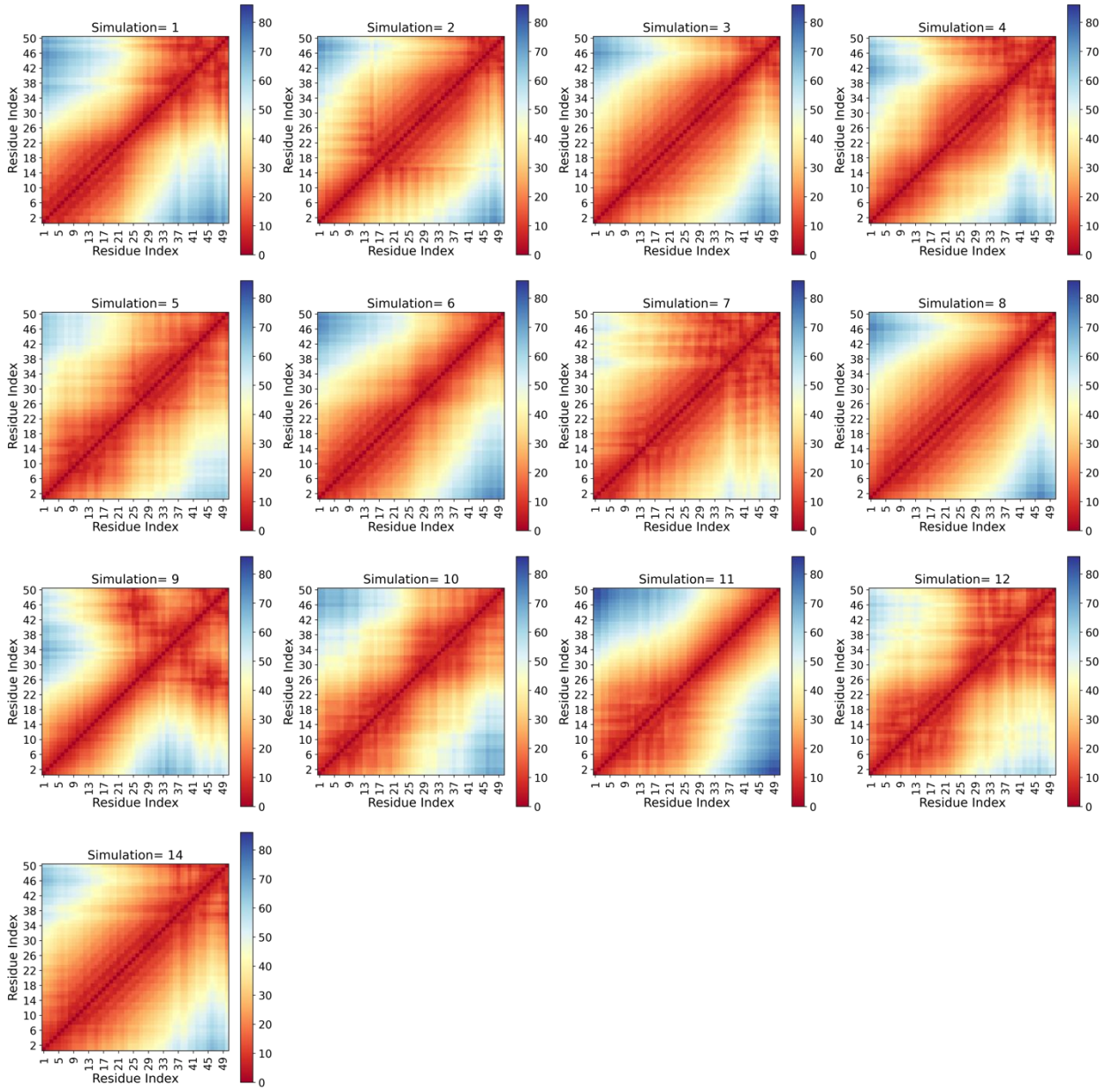

Figure S22: Inter-residue distance heatmap for every human P1 trajectory. Figure 4 has representative conformation images for some of these simulations. For example, simulation #9 is

represented by image #1 in figure 4. Simulation #5 is represented by image #7 in figure 4. The trajectory and conformation images presented in figure 9 are from simulation #4. Distance values are in Angstrom.

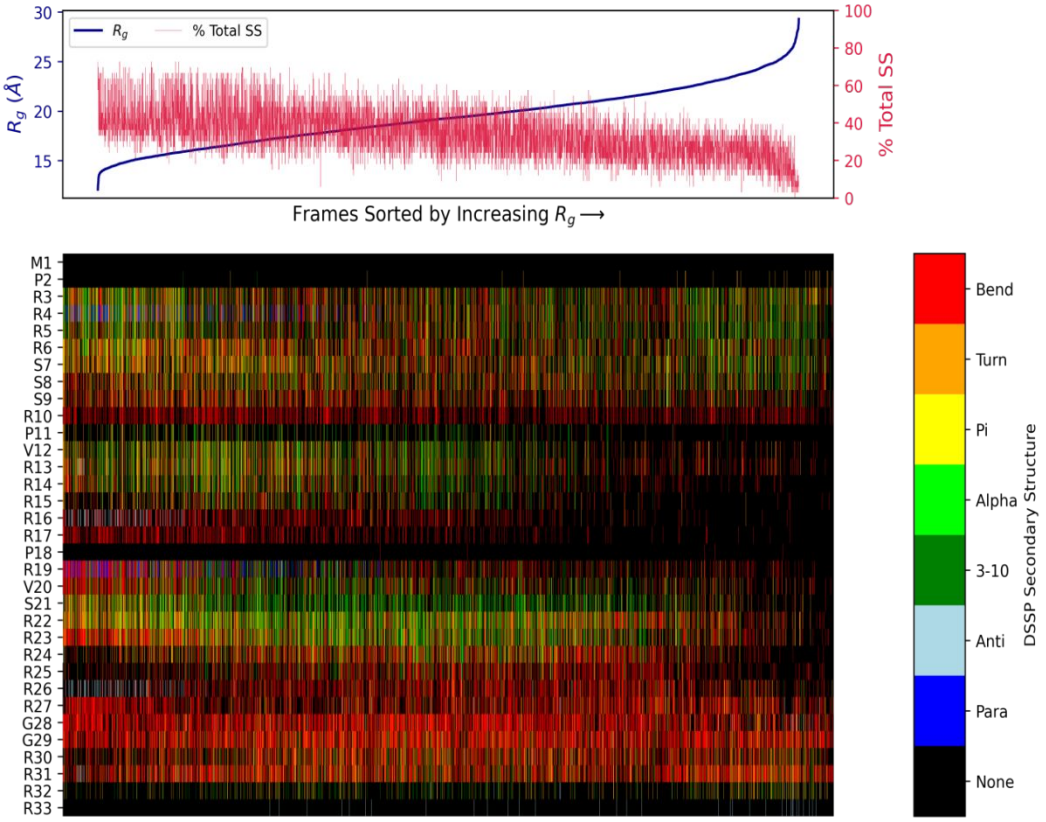

Figure S23: Secondary structure versus radius of gyration values for salmon protamine. Results taken every 200 ps, for all trajectories combined. The % total secondary structure includes all secondary structure categories from the colorbar shown on the right except 'None.' Panel of colors in the bottom left shows the secondary structure of each residue in each frame (with the frames sorted as shown).

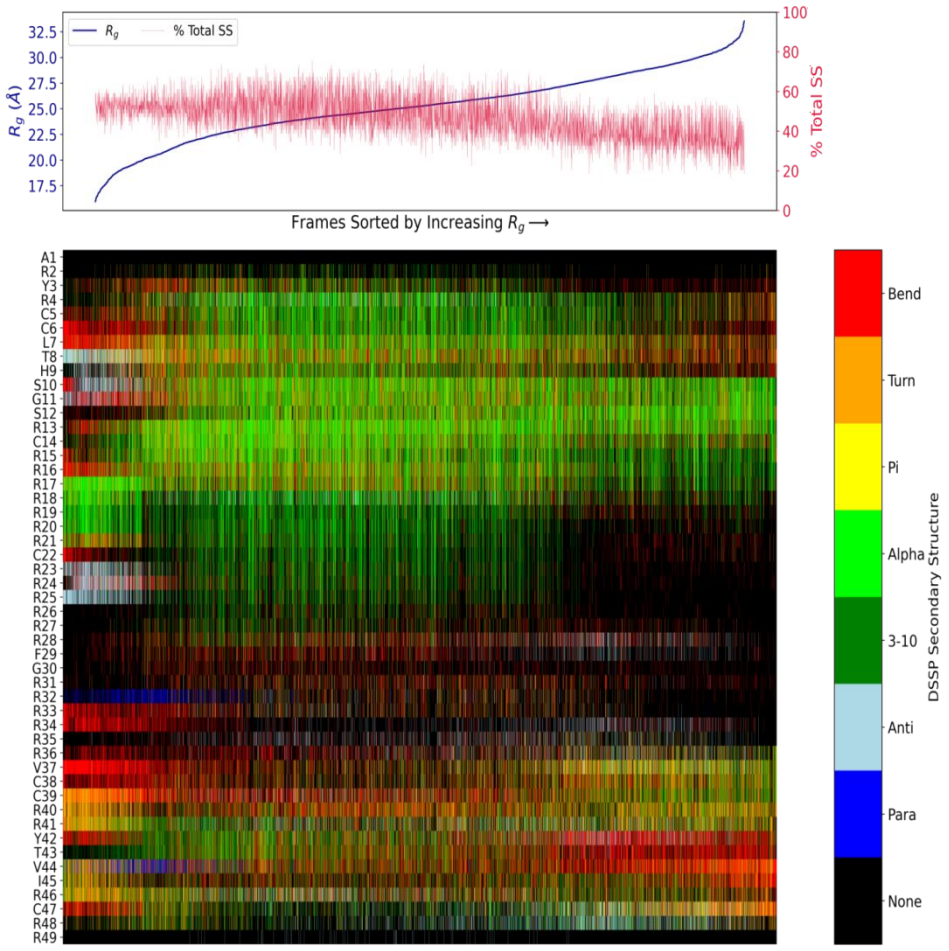

Figure S24: Secondary structure versus radius of gyration values for bull P1 protamine. Results taken every 200 ps, for all trajectories combined. The % total secondary structure includes all secondary structure categories from the colorbar shown on the right except 'None.' Panel of colors in the bottom left shows the secondary structure of each residue in each frame (with frames sorted as shown).

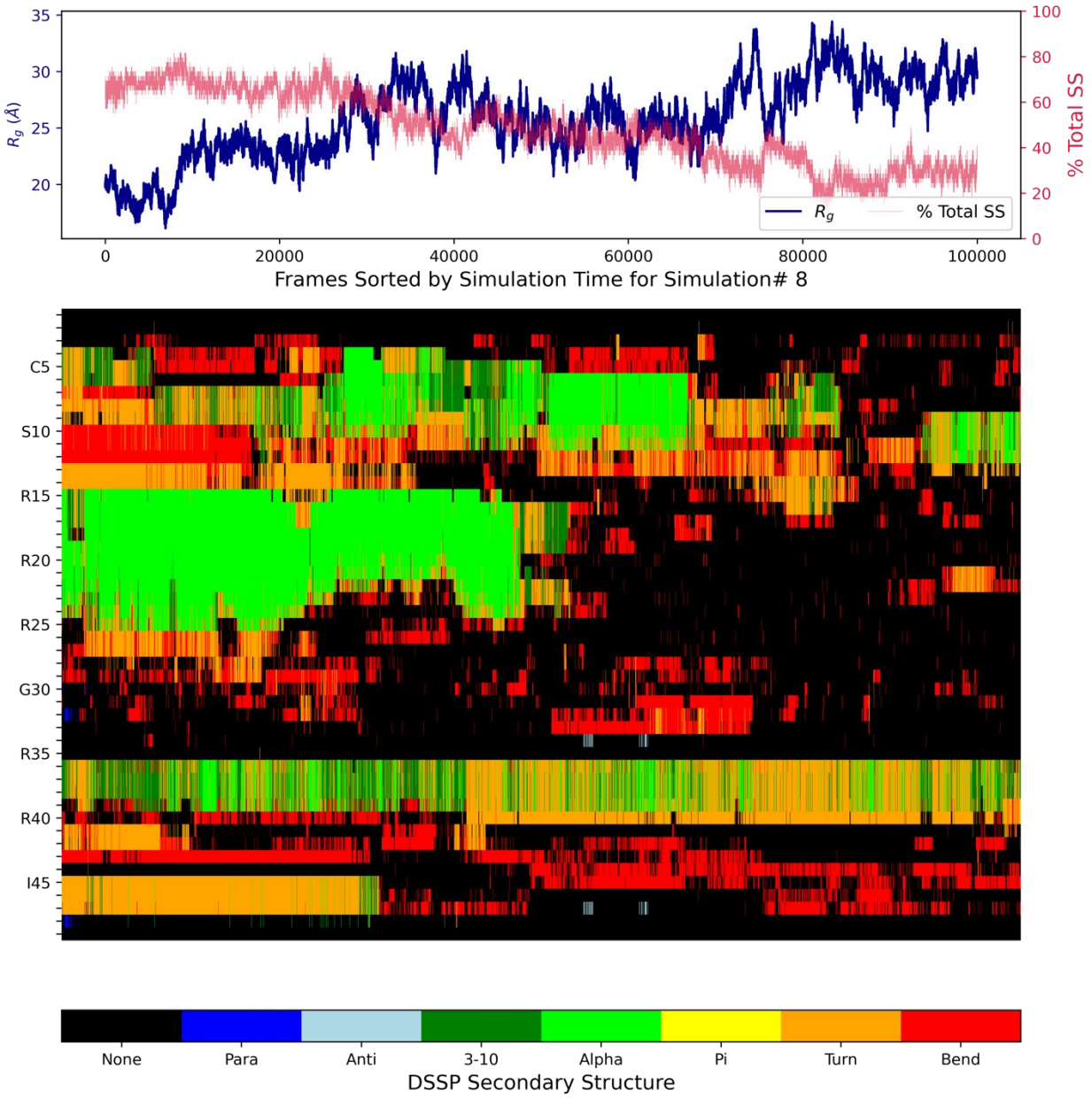

Figure S25: Secondary structure versus simulation time for a single bull P1 protamine explicit solvent trajectory (full 200ns shown for simulation# 8), progressing from left to right. The % total secondary structure includes all secondary structure categories from the colorbar shown at the

bottom except 'None.' Panel of colors in the middle shows the secondary structure of each residue in each frame.

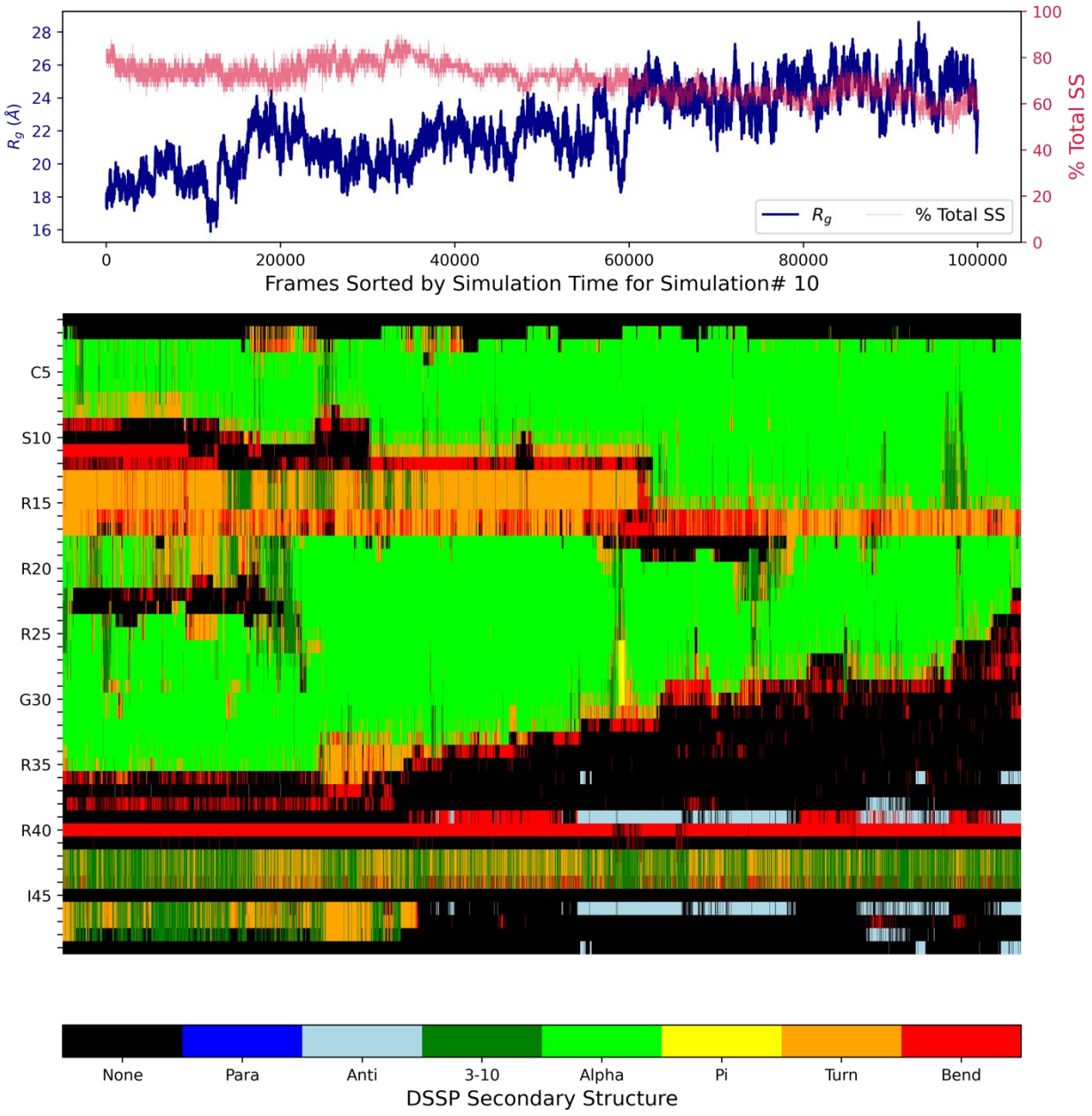

Figure S26: Secondary structure versus simulation time for a single bull P1 protamine explicit solvent trajectory (full 200ns shown, simulation# 10), progressing from left to right. The % total

secondary structure includes all secondary structure categories from the colorbar shown at the bottom except 'None.' Panel of colors in the middle shows the secondary structure of each residue in each frame.

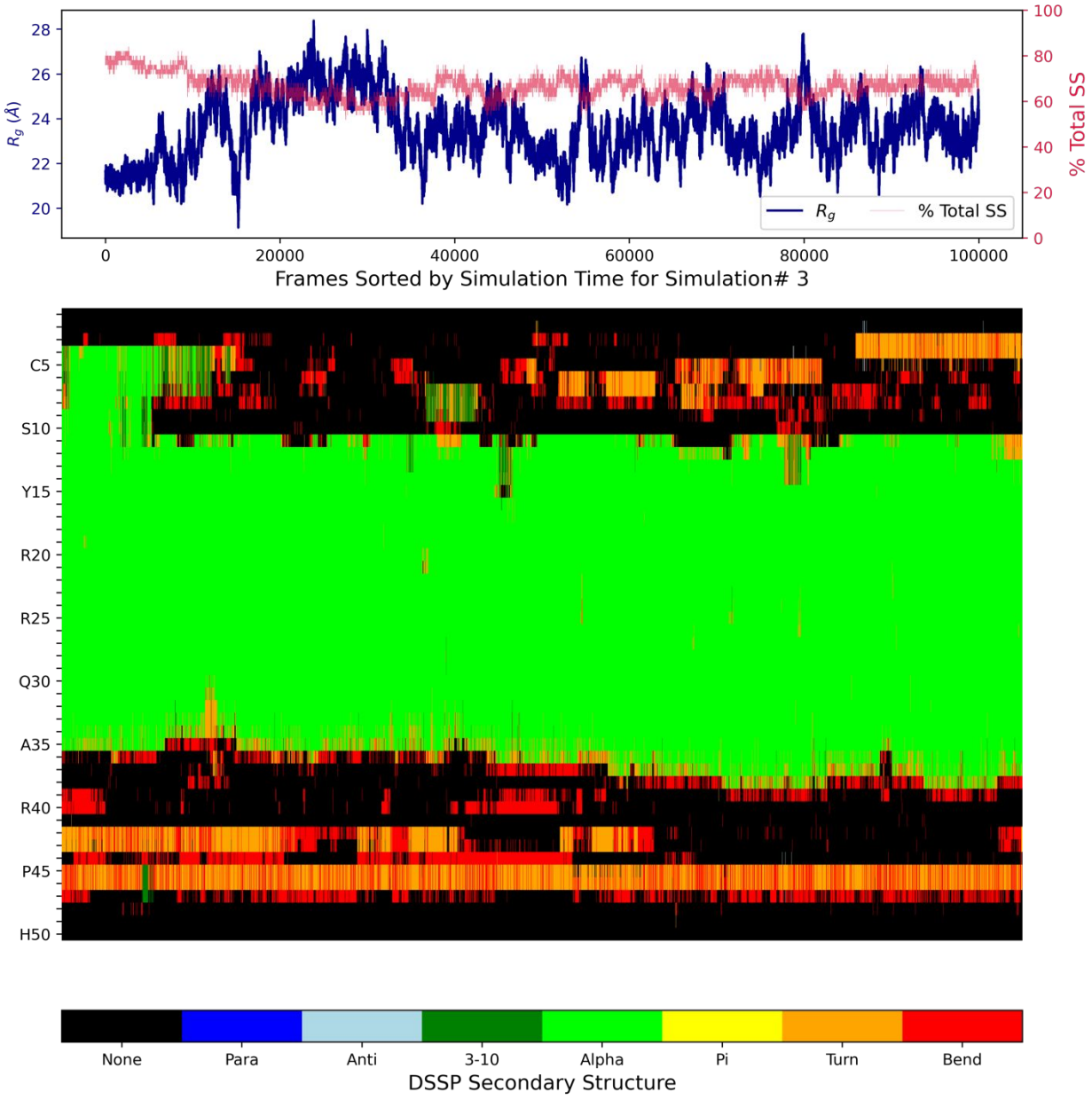

Figure S27: Secondary structure versus simulation time for a single human P1 protamine explicit solvent trajectory (full 200ns shown, simulation# 3), progressing from left to right. The % total secondary structure includes all secondary structure categories from the colorbar shown at the bottom except 'None.' Panel of colors in the middle shows the secondary structure of each residue in each frame.

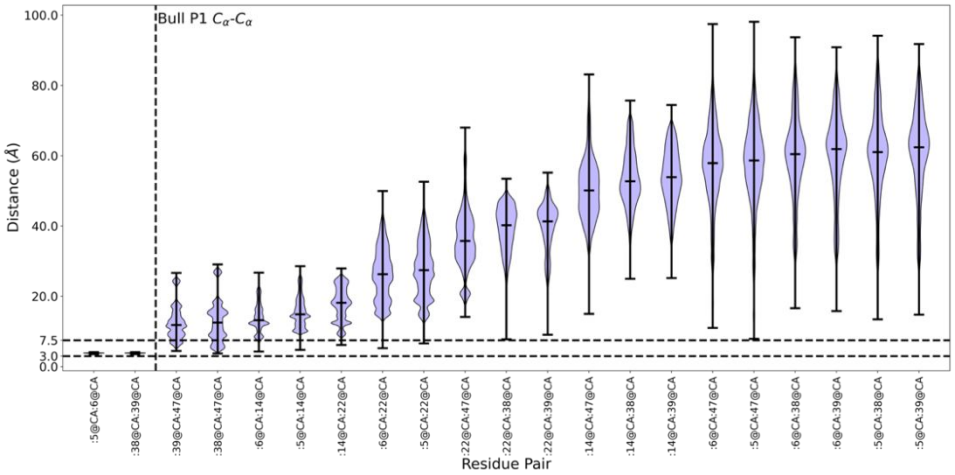

Figure S28: C_α- C_α distances for bull P1 protamine residue pairs. Limits highlighted by dotted horizontal lines are based on data published by Gao *et al.*[2] Dotted vertical line separates non-consecutive residues from consecutive ones. Residue pairs are sorted from low to high mean distance.

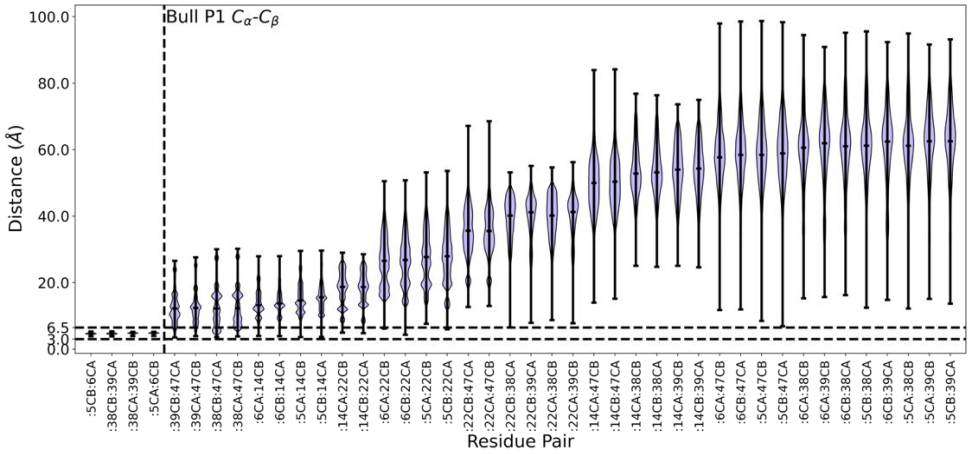

Figure S29: C_α- C_β distances for bull P1 protamine residue pairs. Limits highlighted by dotted horizontal lines are based on figure S7. Dotted vertical line separates non-consecutive residues from consecutive ones. Residue pairs are sorted from low to high mean distance.

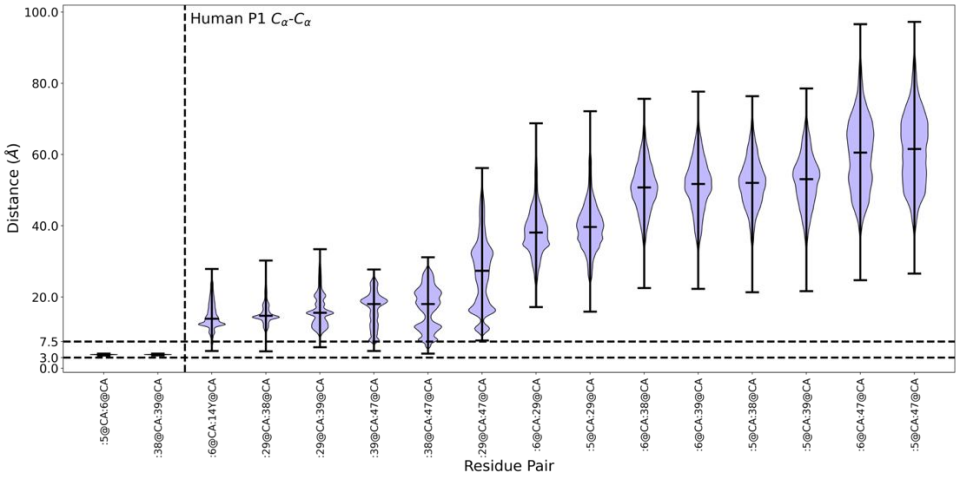

Figure S30: C_α- C_α distances for human P1 protamine residue pairs. Limits highlighted by dotted horizontal lines are based on data published by Gao *et al.*[2] Dotted vertical line separates non-consecutive residues from consecutive ones. Residue pairs are sorted from low to high mean distance.

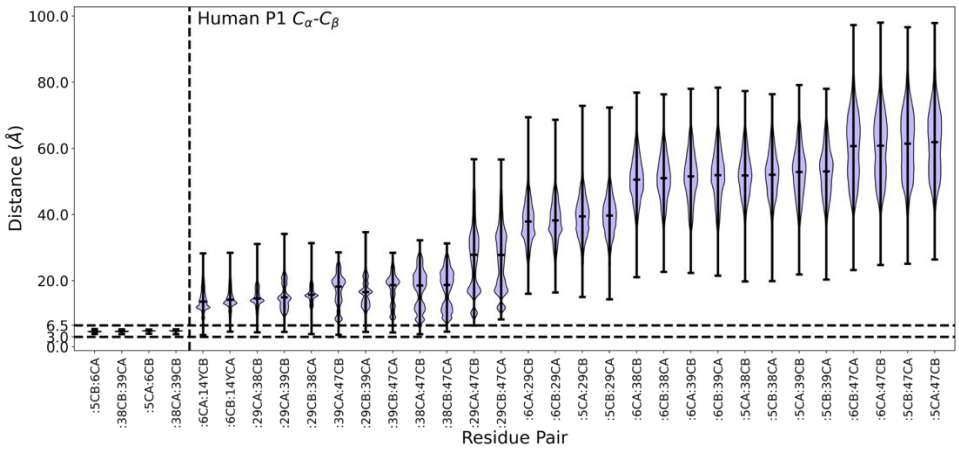

Figure S31: C_α- C_β distances for human P1 protamine residue pairs. Limits highlighted by dotted horizontal lines are based on figure S7. Dotted vertical line separates non-consecutive residues from consecutive ones. Residue pairs are sorted from low to high mean distance.

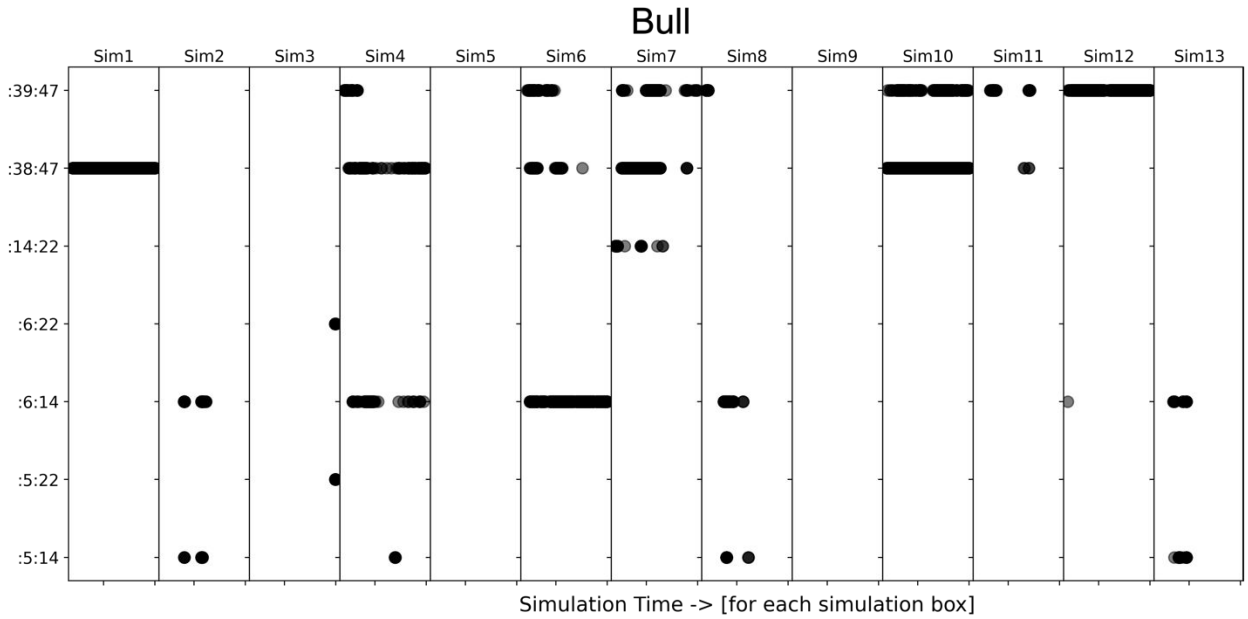

Figure S32: Scatter plot identifying frames when residue pairs meet all distance conditions for potential disulfide bonding (refer to Table 3). X axis represents progress of simulation time for each trajectory separately. Y axis shows residue pairs.

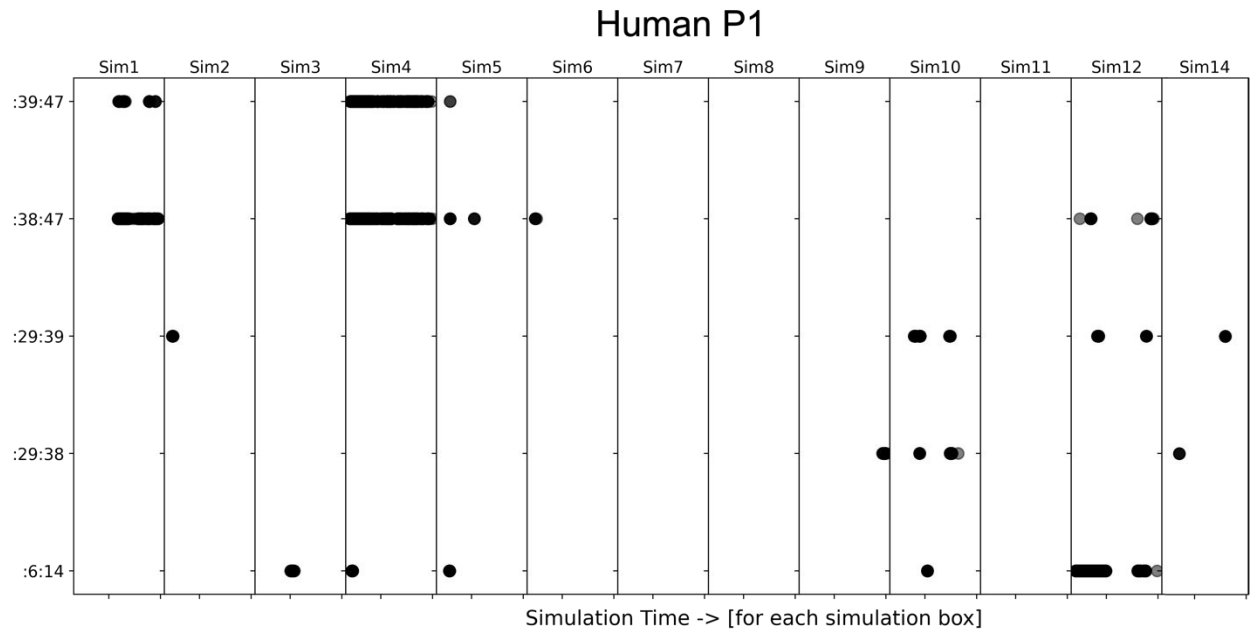

Figure S33: Scatter plot identifying frames when residue pairs meet all distance conditions for potential disulfide bonding (refer to Table 3). X axis represents progress of simulation time for each trajectory separately. Y axis shows residue pairs.

References

(1) Wang, G.; Dunbrack, R. L. PISCES: A Protein Sequence Culling Server. *Bioinformatics* **2003**, *19* (12), 1589–1591. https://doi.org/10.1093/bioinformatics/btg224.

(2) Gao, X.; Dong, X.; Li, X.; Liu, Z.; Liu, H. Prediction of Disulfide Bond Engineering Sites Using a Machine Learning Method. *Sci. Rep.* **2020**, *10* (1), 10330. https://doi.org/10.1038/s41598-020-67230-z.

(3) Powell, C. D.; Kirchoff, D. C.; DeRouchey, J. E.; Moseley, H. N. B. Entropy Based Analysis of Vertebrate Sperm Protamines Sequences: Evidence of Potential Dityrosine and

Cysteine-Tyrosine Cross-Linking in Sperm Protamines. *BMC Genomics* **2020**, *21* (1), 277. https://doi.org/10.1186/s12864-020-6681-2.