Supplementary Information for Molecular Insight into the Protein-Polymer Interactions in N-terminal PEGylated Bovine Serum Albumin

Aravinda Munasinghe[1,4,5,&], Akash Mathavan[1,2,4,&], Akshay Mathavan[1,2,4], Ping Lin[1,4,5] and Coray M. Colina[1,3,4,5,*]

[1]Department of Chemistry, University of Florida, Gainesville, FL 32611, United States

[2]Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611, United States

[3]Department of Materials Science and Engineering, University of Florida, Gainesville, FL 32611, United States

[4]George & Josephine Butler Polymer Research Laboratory, University of Florida, Gainesville, Florida 32611, United States

[5]Center for Macromolecular Science & Engineering, University of Florida, Gainesville, Florida 32611, United States

[&]These authors contributed equally to this work.

[*]Corresponding author: email: colina@chem.ufl.edu, phone: +1 (352) 294-3488

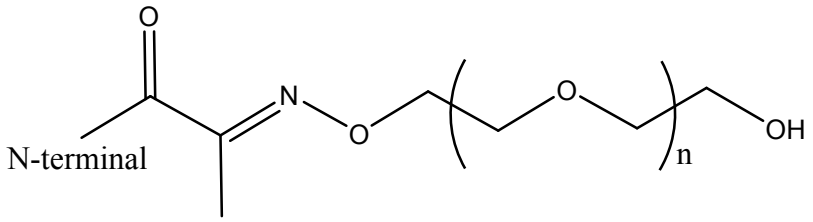
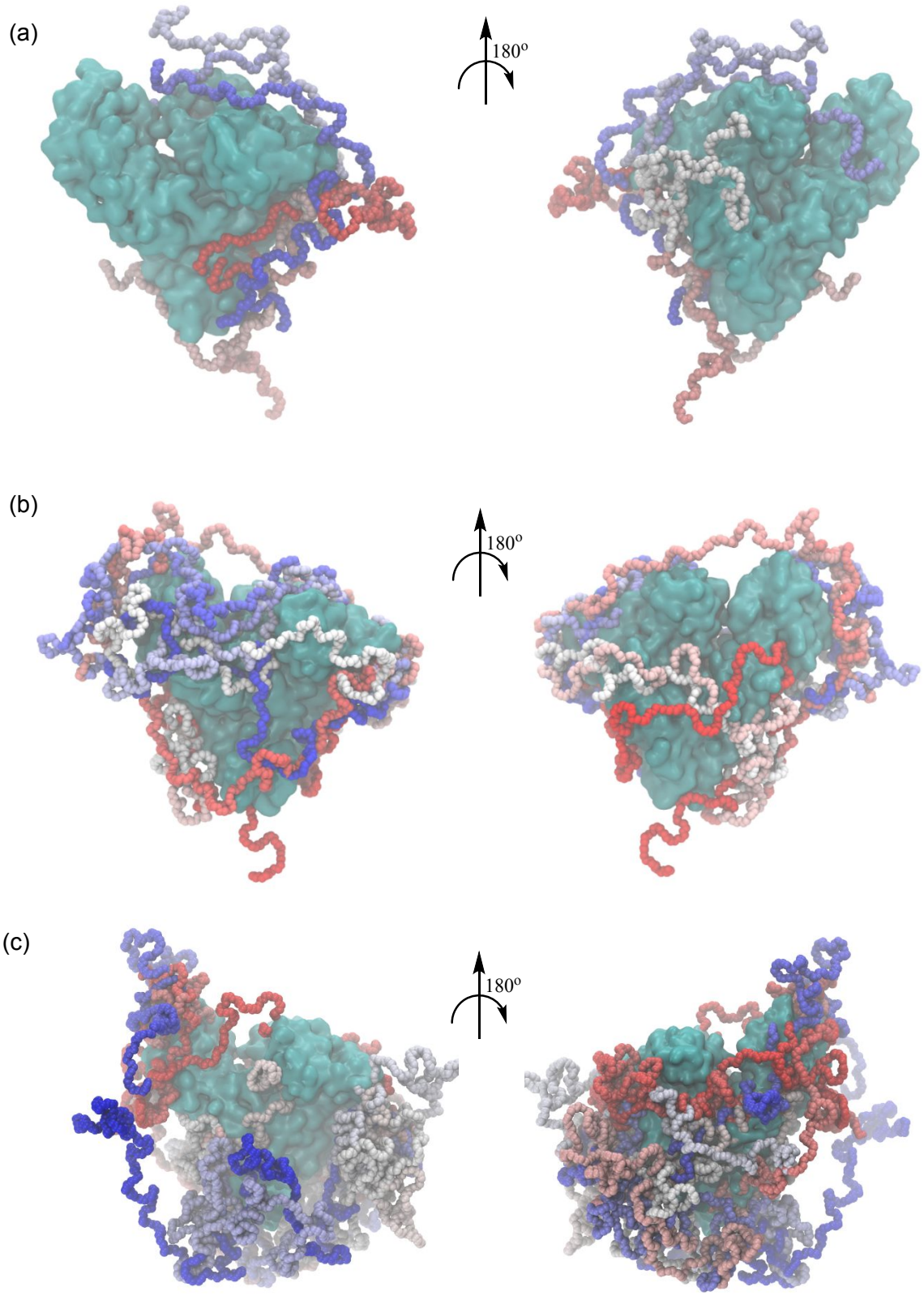


Figure S1. Schematic of the linking molecule of PEG and BSA. Each PEG chain was linked to the N-terminal of the BSA via an oxime link. For each conjugate with 2, 5, 10, and 20 kDa PEG, respective repeating numbers were n = 45, 113, 227, and 454.

Table S1. Simulation system sizes for 2, 5, 10 and 20 kDa PEGylated BSA conjugates.

System	System Size (Number of Atoms)			
	2 kDa	5 kDa	10 kDa	20 kDa
1	226430	207775	448375	407167
2	226017	205563	321594	377747
3	215557	195055	270004	370434
4	214217	192951	248924	366954
5	206679	177645	244456	366046
6	206451	174117	235365	361379
7	203928	173221	228714	347535
8	199962	173221	191483	340579
9	199327	168629	172841	336725
10	193605	158948	170069	320096


(a)
180°
(b)
180°
(c)
180°

(d)

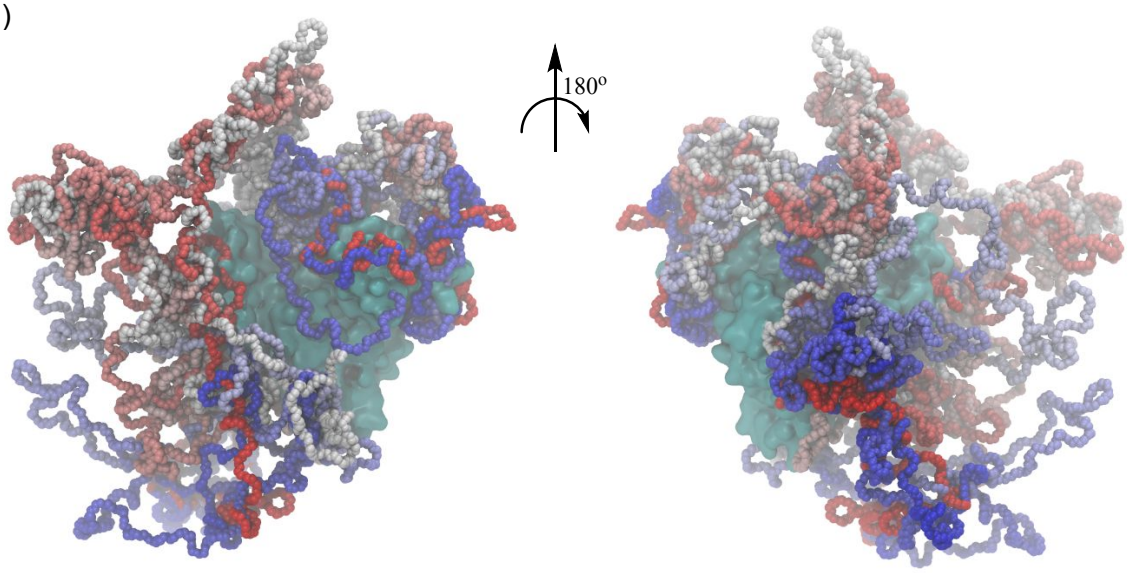


Figure S2. Superimposed PEG conformations used as initial structures in 2 (a), 5 (b), 10 (c), and 20 (d) kDa conjugates. Each color represents a different conformation. Right-hand side figure is generated by rotating the left-hand figure by 180°.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Equation S1. Formula for calculating the average of a data set, where n is the sample size.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} |x_i - \bar{x}|^2}{n-1}}$$

Equation S2. Formula for calculating the standard deviation of a data set, where n is the sample size.

In this work, the last 150 ns of production for the 10 conjugate systems for each PEG molecular weight were collected and aggregated into one “sample”. Therefore, there were four samples of 1.5 μs (one per PEG molecular weight). The average and standard deviation of these samples were calculated according to standard formulas shown above.

Table S2. RMSD calculations for the protein backbone and each domain of BSA for 2, 5, 10 and 20 kDa PEGylated BSA conjugates as well as for free BSA.

System	RMSD (Å)			
	Protein BB	Domain 1	Domain 2	Domain 3
Free BSA	2.8 (0.5)	1.7 (0.2)	1.4 (0.2)	1.9 (0.2)
2 kDa	2.5 (0.4)	1.6 (0.2)	1.3 (0.2)	1.8 (0.2)
5 kDa	2.8 (0.5)	1.8 (0.2)	1.3 (0.3)	1.7 (0.2)
10 kDa	2.6 (0.5)	1.6 (0.2)	1.3 (0.3)	1.8 (0.3)
20 kDa	2.7 (0.5)	1.6 (0.2)	1.3 (0.2)	1.8 (0.2)

To measure changes in secondary structure of BSA over time as a function of PEG M_w, VMD's STRIDE algorithm was employed for all simulations, including an MD simulation of free, native BSA.[1] Overall, no substantial changes in secondary structure were seen across the protein's 583 residues (**Figure S3**). However, small, more frequent changes were detected in domain I (residues 1 to 193), specifically domain IA (residues 1 to 105), of BSA (**Figure S4**). Across both time and simulations for all PEG M_w, residues 80 to 90 of BSA varied between an alpha helix and a 3-10 helix secondary structure motif, with occasional transitions to a pi helix motif. Similar transitions were prevalent in the secondary structure of native BSA, although no pi helix motifs were seen.

The defining distinction between these three motifs involves the pattern of hydrogen bonding; in amino acids of alpha helices, the amine forms a hydrogen bond with the carbonyl four residues earlier. In contrast, hydrogen bonding occurs with the carbonyl 3 residues earlier and 5 residues earlier in the 3-10 helix and pi helix motifs, respectively.[2] It is interesting to note that these frequent changes occur in the domain in which PEG grafting occurred. However, because both 3-10 helices and pi helices encompass short segments of amino acids and frequently occupy an intermediate conformation before transitioning back an alpha helix, these changes cannot be definitively attributed to PEGylation and further analysis may be needed in future work.

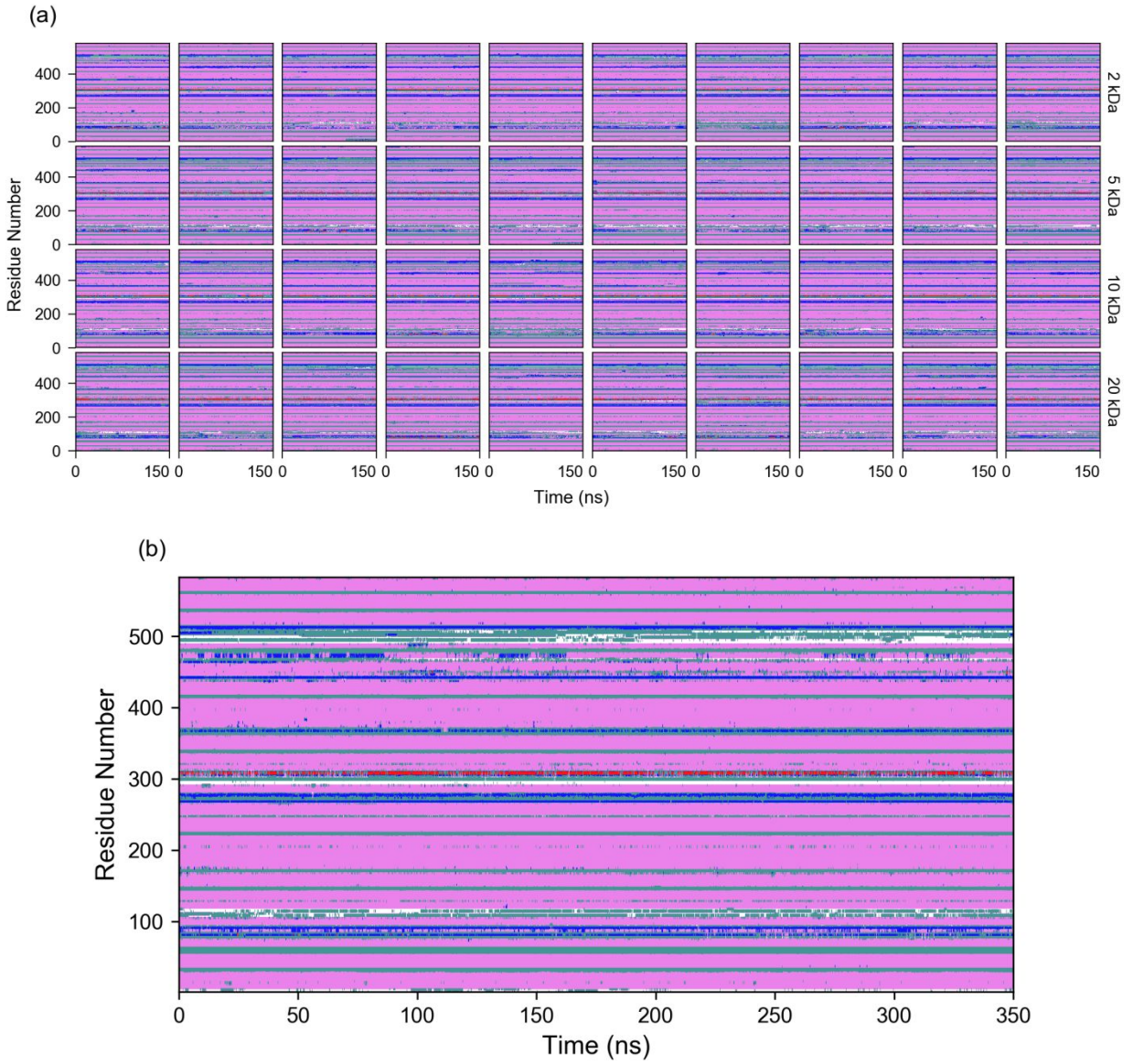


Figure S3. Secondary structure analysis for residues all residues of 2, 5, 10, and 20 kDa PEGylated BSA (a) trajectories using VMD's STRIDE algorithm. Secondary structure analysis was also conducted for free, native BSA (b). Represented secondary structure motifs are: turn (teal), coil (white), alpha helix (pink), 3-10 helix (blue) and pi helix (red).

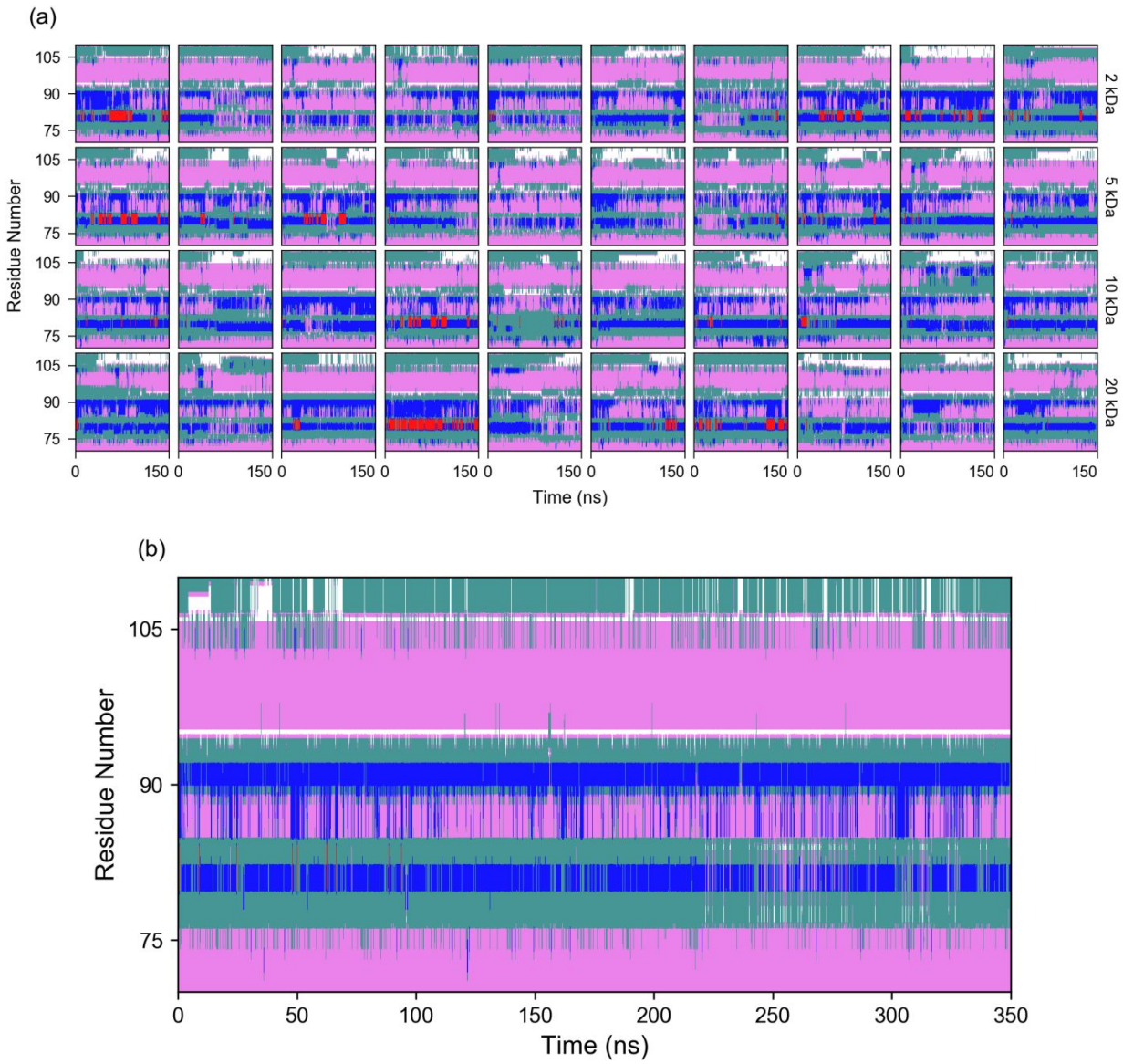


Figure S4. Secondary structure analysis for residues 70 to 110 of 2, 5, 10, and 20 kDa PEGylated BSA (a) trajectories using VMD's STRIDE algorithm. Secondary structure analysis was also conducted for free, native BSA (b). Represented secondary structure motifs are: turn (teal), coil (white), alpha helix (pink), 3-10 helix (blue) and pi helix (red).

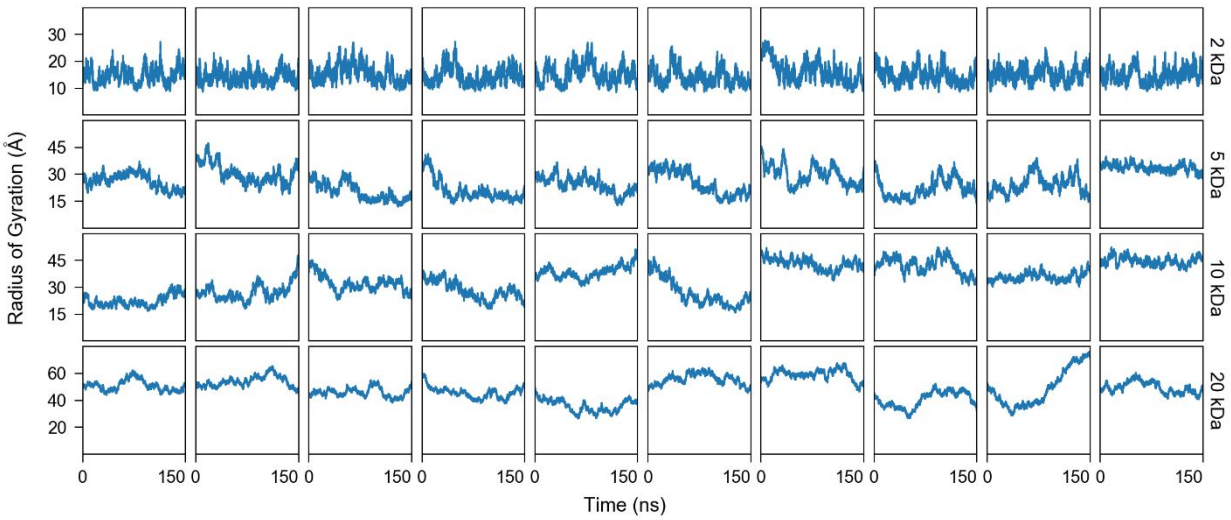


Figure S5. Evolution of radius of gyration over time for 2, 5, 10, and 20 kDa PEGylated BSA trajectories.

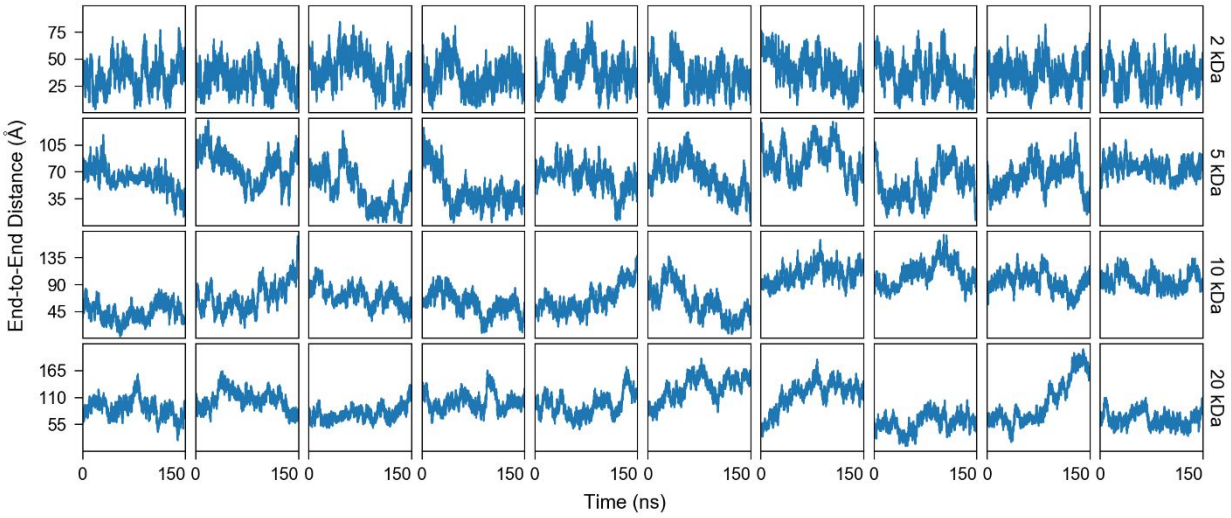
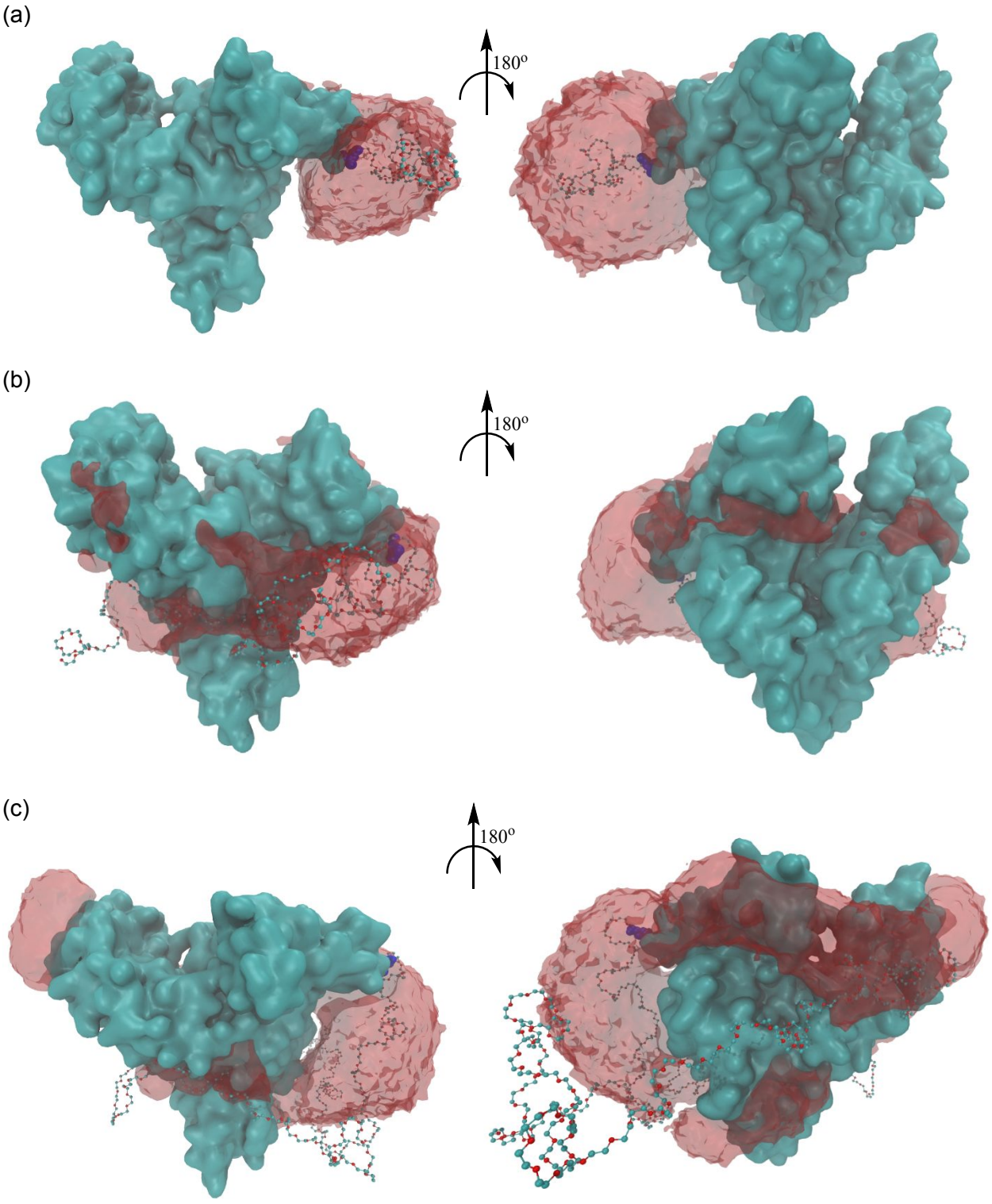


Figure S6. Evolution of end-to-end distance over time for 2, 5, 10, and 20 kDa PEGylated BSA trajectories.


(a)
180°
(b)
180°
(c)
180°

(d)

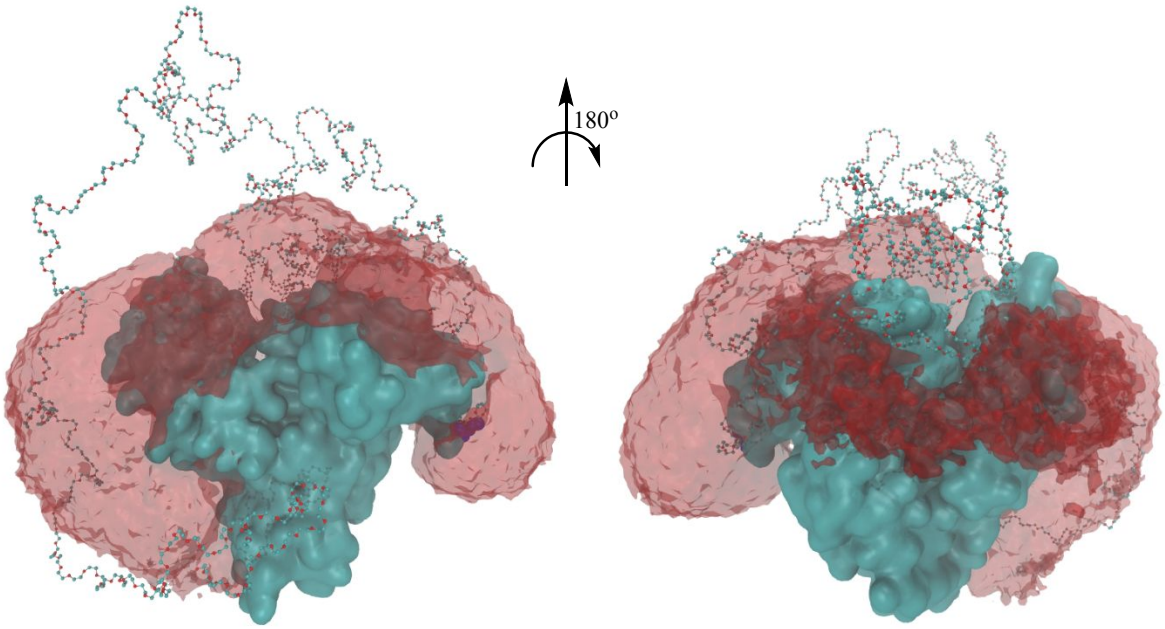


***Figure S7**. Cartoon representations of 2 (a), 5 (b), 10 (c), and 20 (d) kDa PEGylated BSA with transparent red regions marking areas in which PEG density is larger than one-tenth of the highest density in each conjugate. BSA's surface is marked with cyan while a randomly selected PEG chain, in CPK drawing method, shows a conformation in which the chain protrudes away when not interacting with the protein's surface.*

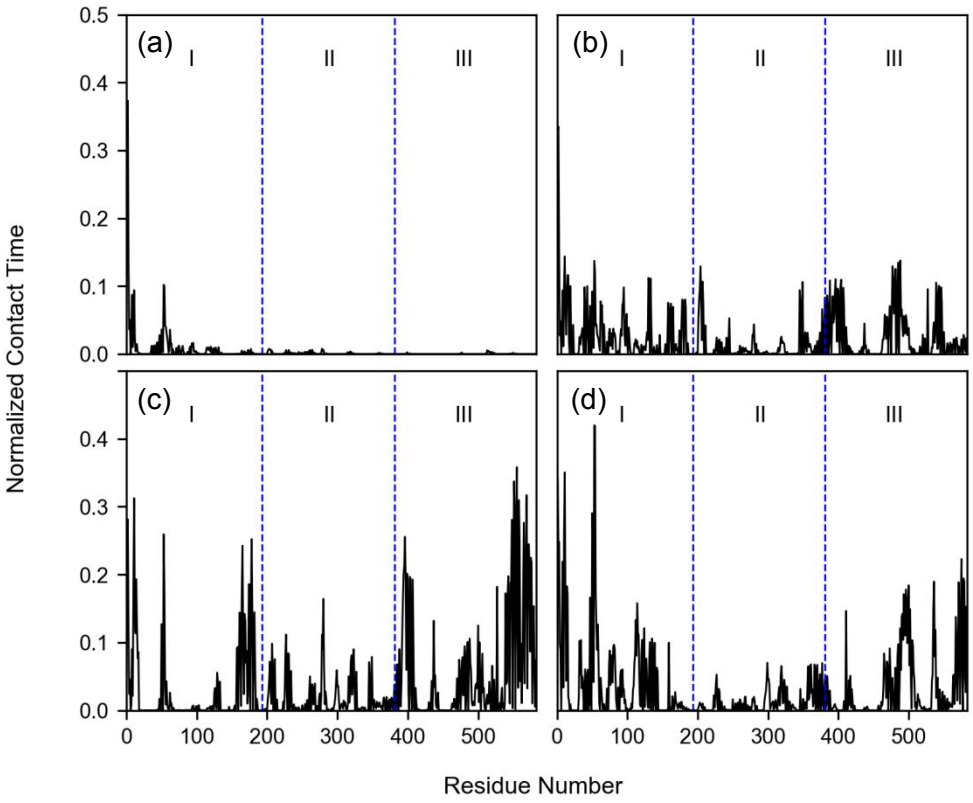


Figure S8. Contact time distribution normalized through simulation time for 2 (a), 5 (b), 10 (c), and 20 (d) kDa PEGylated BSA trajectories. A contact refers to an instance in which the distance between atoms of BSA and the PEG chain was less than 5 Å. Blue vertical lines mark the labeled domains of BSA.

Table S3. Unique lysine residues with top maximum residence times along with respective surrounding hydrophobic surface area (for any single simulation, regardless of PEG M_w).

Residue ID	Residue Name	Maximum Residence Time (ns)	[a]Surrounding hydrophobic surface area (Å^2)
136	Lys	150.00	50(11)
556	Lys	150.00	130(11)
20	Lys	146.50	26(8)
12	Lys	112.25	30(8)
180	Lys	109.30	140(14)
131	Lys	107.30	31(9)
544	Lys	104.21	104(21)
350	Lys	91.63	70(9)
537	Lys	90.28	72(22)
4	Lys	34.32	41(10)

[a] Solvent accessible surface area of hydrophobic residues within 5 Å of each lysine residue was calculated.

References

(1) Heinig, M.; Frishman, D. STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. *Nucleic Acids Res.* **2004**, *32*, 500–502.

(2) Rohl, C. A.; Doig, A. J. Models for the 3(10)-Helix/Coil, Pi-Helix/Coil, and Alpha-Helix/3(10)-Helix/Coil Transitions in Isolated Peptides. *Protein Sci.* **1996**, *5* (8), 1687–1696.