

Supporting Information: Control of Polymer-Protein Interactions by Tuning Composition and Length of Polymer Chains

Menghan Xie,¹ Xu Jia,¹ and Xiao Xu^{1,2}

¹*School of Chemistry and Chemical Engineering,*

Nanjing University of Science and Technology,

200 Xiao Ling Wei, Nanjing 210094, P. R. China;

²*The State Key Laboratory of Molecular Engineering of Polymers,*

Fudan University, Shanghai 200433, China

I. STRUCTURE SCHEMATIC

We use the Martini force field to model the NIPAm, TBAm, AAc, and the mapping relationship is shown in Fig. S1. The composition of various polymers during the simulation process is shown in Fig. S2.

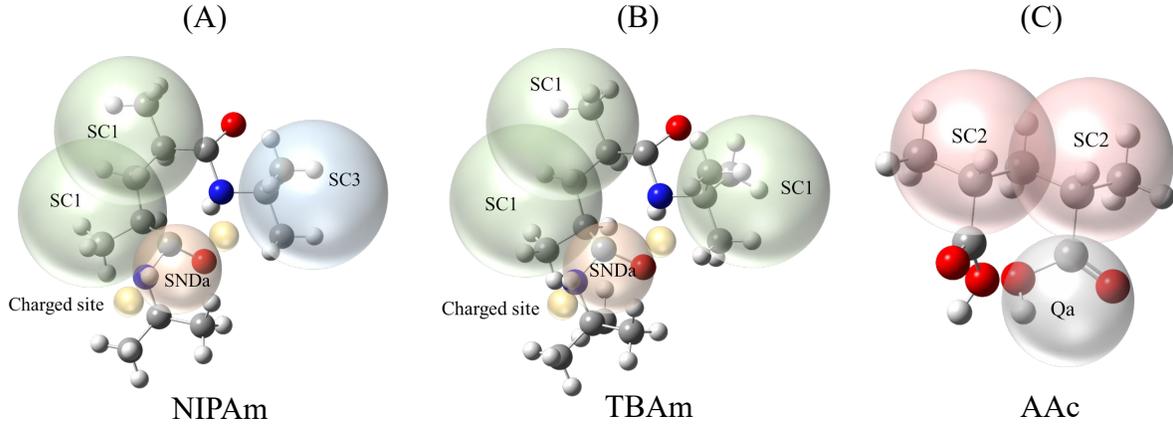


FIG. S1. (A) Mapping of all-atom to coarse-grained for NIPAm. (B) Mapping of all-atom to coarse-grained for TBAm. (C) Mapping of all-atom to coarse-grained for AAc.

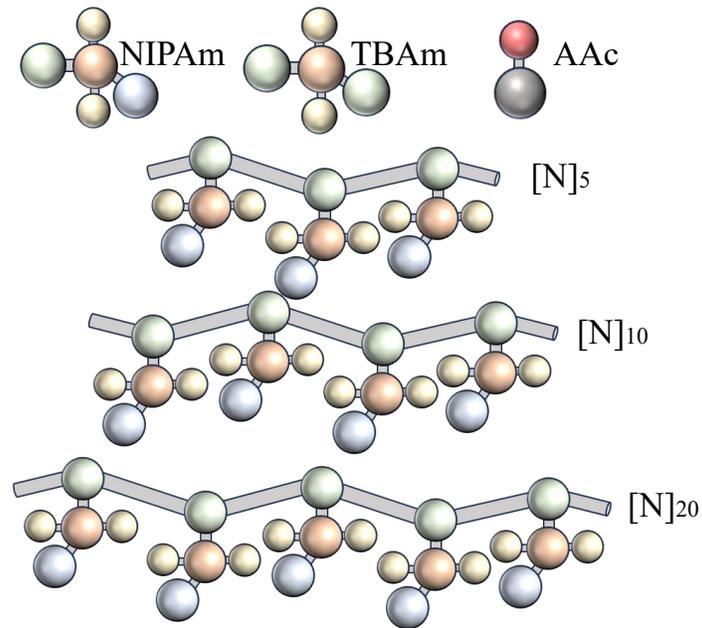


FIG. S2. Schematic diagram of monomers and a set of polymers $[N]_x$ with an increasing degree of polymerization.

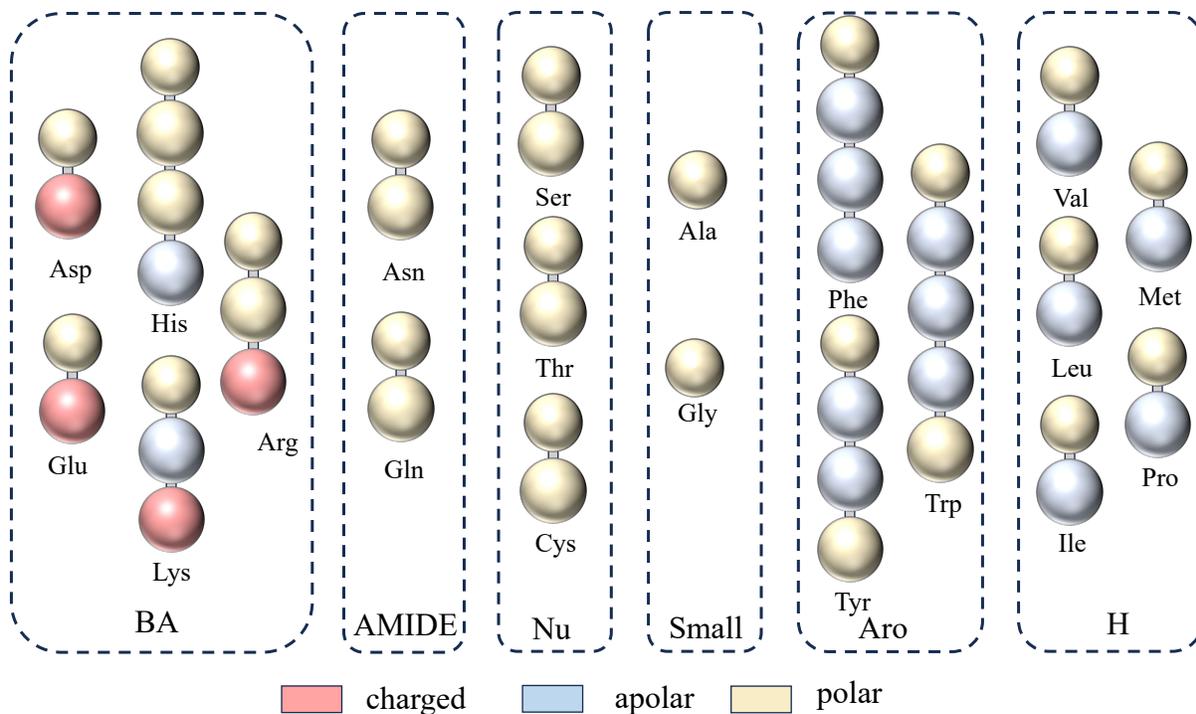


FIG. S3. The 6 classes of categorized AAs

II. AMINO ACID CLASSIFICATION

The precise classification of the 20 amino acids is shown in Fig. S3.

III. SCHEMATIC DISTRIBUTION OF BASIC AMINO ACIDS

The location of the basic amino acids on the protein is the blue area marked in Fig. S4.

IV. THE CONTACT TIMELINE OF THE PROTEIN-OLIGOMER BINDING COMPLEX

Figure S5 and S6 show the temporal evolution of each EpCAM AA contact in the structure of the protein complexed with ligand $[T]_1, [T]_{10}, [T]_{20}$, $[N]_{10}, [A]_5$ and $[A]_{10}$. In order to facilitate the observation of the binding site, we add a color bar specifying the protein subunits along with the protein AA index. Moreover, in Fig. S7, the AA type is specified together with the protein AA index.

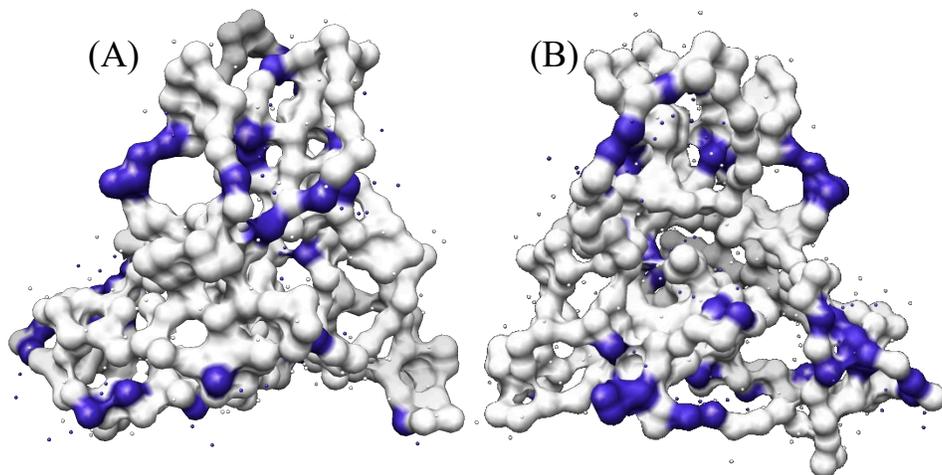


FIG. S4. The front (A) and back (B) side of the protein with a highlighted area of distributed basic AAs.

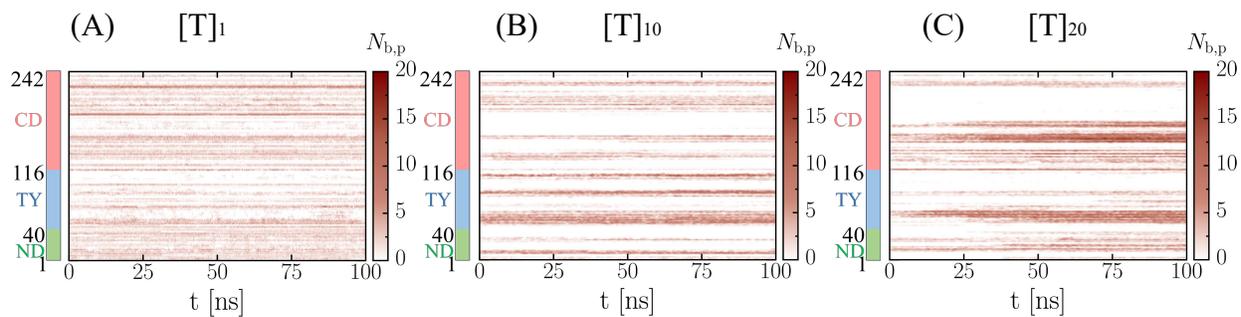


FIG. S5. The contact timeline of the protein complexed with the $[T]_1$ (A), $[T]_{10}$ (B) and $[T]_{20}$ (C) on the protein. The color bar in the y-axis depicts protein AAs belonging to a specific protein subunit.

V. THE COMPLEX STRUCTURE OF THE PROTEIN BOUND WITH ONE LIGAND

In Fig. S8, the complex structure of the EpCAM protein complexed with $[N]_1$, $[N]_5$, $[N]_{10}$, and $[N]_{20}$ are presented. The PMF profile for these bound ligands are depicted in Fig. 6(A) in the manuscript.

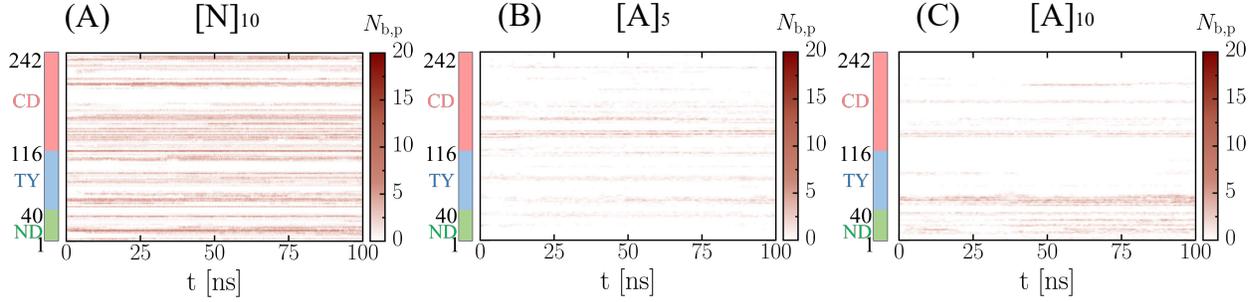


FIG. S6. The contact timeline of the protein complexed with the $[N]_{10}$ (A), $[A]_5$ (B) and $[A]_{10}$ (C) on the protein. The color bar in the y-axis depicts protein AAs belonging to a specific protein subunit.

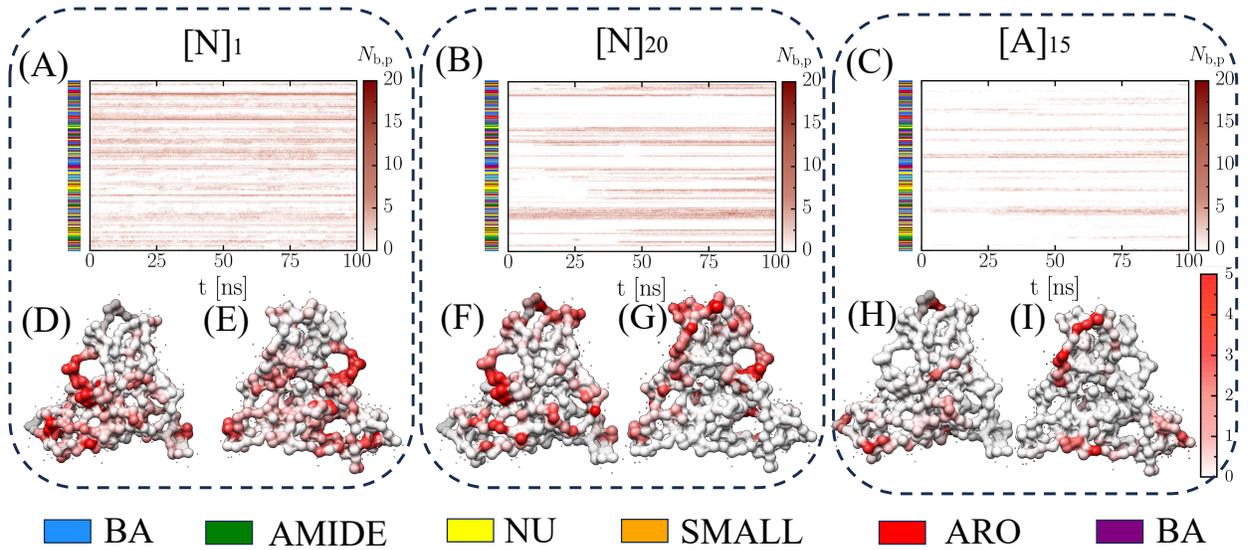


FIG. S7. Contact timeline of the protein complexed with the NIPAm monomer $[N]_1$ (A), PNIPAm $[N]_{20}$ (B) and copolymer chain $[A]_{15}$ mixed with AAc monomers (C) trajectories. Color bar (left) depicts the distribution of six types of the AA. The blue, green, yellow, orange, red and purple color indicate the BA, AMIDE, NU, SMALL, ARO and BA protein AAs, respectively. color map intensity (right) reflects the number $N_{b,p}$ of polymer CG beads that are in contact with the protein (Within 5 \AA). The CG structure of the EpCAM protein with the heatmap coloring scheme showing the explicit regions that are bound with the NIPAm monomer $[N]_1$ (D-E), PNIPAm $[N]_{20}$ (F-G) and charged copolymer $[A]_{15}$ (H-I). Note that, panel (D,F,H) and panel (E, G, I) individually present the adsorption on the two sides of the protein, respectively.

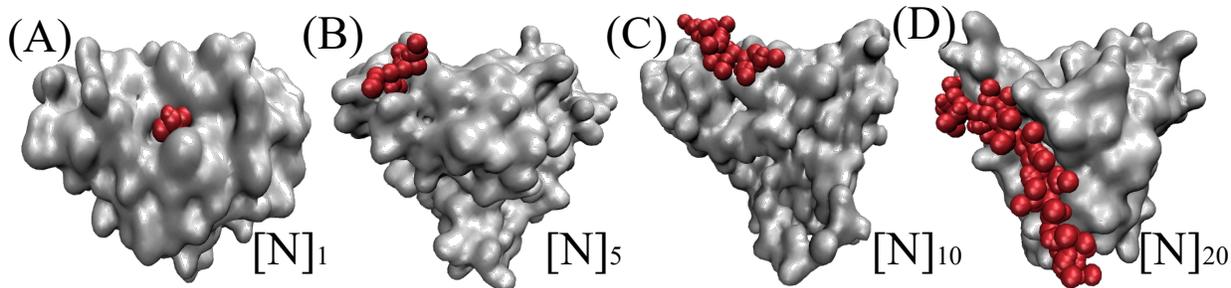


FIG. S8. The snapshot of the simulated complex of the protein EpCAM with single ligand $[N]_1$ (A), $[N]_5$ (B), $[N]_{10}$ (C) and $[N]_{20}$ (D).

VI. THE AROMATIC AAS IN THE BINDING PROCESS

For a specific binding region on the protein left wing, the distribution of aromatic AAs is displayed in Fig. S9. That is found coincide well with the binding spots for $[N]_1$, while those aromatic AAs are almost ignored by the oligomer $[N]_{20}$.

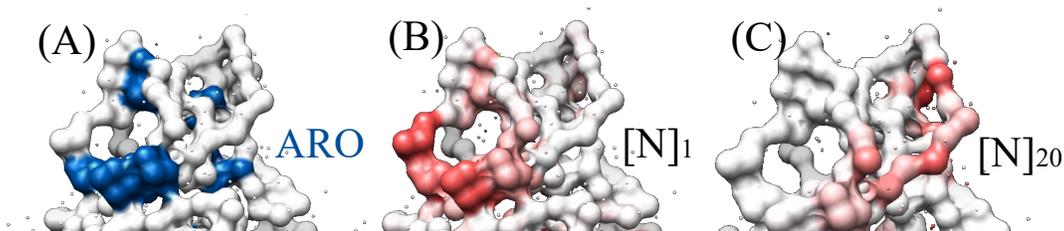


FIG. S9. For a specific binding region on the protein left wing, the distribution of aromatic AAs (A), which is compared with the heatmap coloring of the protein areas bound with the NIPAm monomers (B) and the PNIPAm oligomer $[N]_{20}$ (C).

VII. ALL ATOM SIMULATIONS

To further justify our approach, we perform additional all-atom (AA) simulations to compare with the CG runs. The drastic increase of the atom number in AA simulation means that the equivalent simulation is significantly slower than the Martini runs. Thus, we select the system with the least structural complexity where the unbound protein is equilibrated only with one polymeric ligand. The snapshot of the simulation system is as follows. The snapshot of the simulation system is presented in Fig. S10. We employed the OPLS force

field to model the protein and the oligomer. The simulation comprises roughly 40000 water molecules treated with the SPCE model. To fully equilibrate the binding structure, the simulation lasts for 100 ns. In the following plot, we compare the AA simulations with the CG runs in terms of the equilibrated protein-oligomer binding complex.

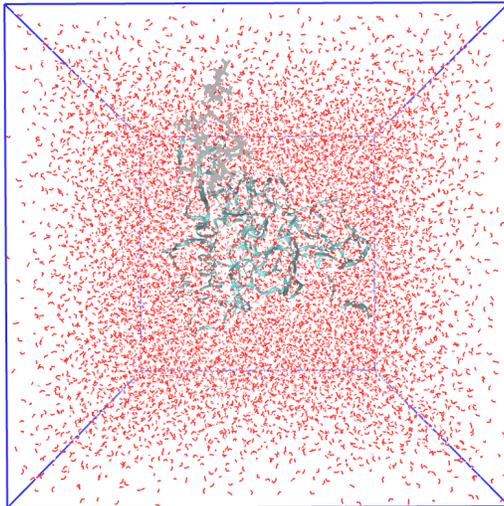


FIG. S10. The snapshot of the AA simulation system containing the oligomer ligand $[N]_{20}$ and the EpCAM protein solvated in water.

For AA runs, the ligands are selected as the single NIPAm monomer ($[N]_1$) and the NIPAm oligomer ($[N]_{20}$). The simulation temperature and salt concentration are set to the same between the AA and CG simulations. The complex structure attained from the simulation is displayed in Fig. S11. In Fig. S11(A-D), we depict the binding complex attained from the AA simulations with the ribbon and surface representation, while the results from the CG simulations are presented in Fig. S11(E-F). From this plot, we are satisfied with the degree of structural prediction accuracy of the CG simulation, as the binding site for both $[N]_1$ and $[N]_{20}$ ligand are found basically the same between two approaches. For instance, the CG and AA simulation both determine the TY loop of the protein as an important binding site for the long oligomer $[N]_{20}$. More importantly, in both simulations we find the switch of the binding site from the protein interior to exterior when increasing the degree of polymerization for ligands, which is an important argument in this work and is now fully supported by AA simulations.

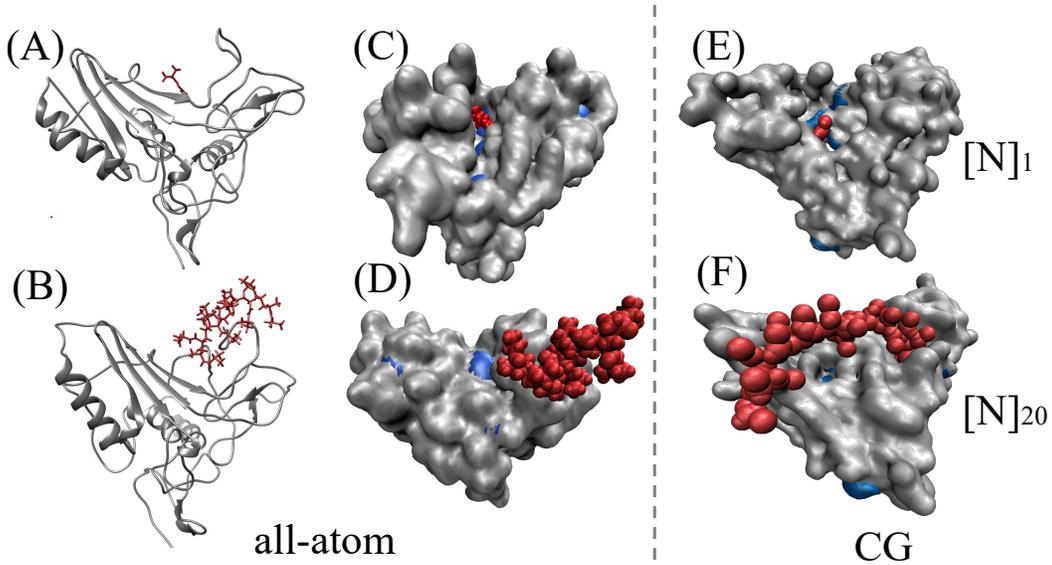


FIG. S11. Results from AA simulations of the structure of the protein complexed with a single monomer $[N]_1$ (A-C) and oligomer $[N]_{20}$ (B-D). Protein is displayed with the ribbon and surface representation in panel (A-B) and panel (C-D), respectively. To compare with the AA simulations, snapshots from Martini simulations of the protein complexed with the single NIPAm monomer $[N]_1$ (E) and the single PNIPAm chain $[N]_{20}$ (F) are presented. Ligands are colored in red and protein Tyr AAs are highlighted in blue.

VIII. PMF CALCULATION

In Fig. S12, regarding the oligomer $[N]_{20}$ (upper) and $[N]_{10}$ (lower), we calculate the PMF profile for the first three oligomer ligands bound with the protein. That essentially monitors the sub-sequential loading of copolymers onto the protein. For all cases, it appears that the binding affinity for the ligand decays with more ligands being adsorbed onto the protein, as a result of the increasing steric effect and the interaction among all bound ligands. That result signals a negative binding cooperativity of the binding process. Based on this result, the oligomer ligands in this work are characterized by its largest binding affinity to the protein, that is, the binding free energy calculated for the firstly bound ligand.

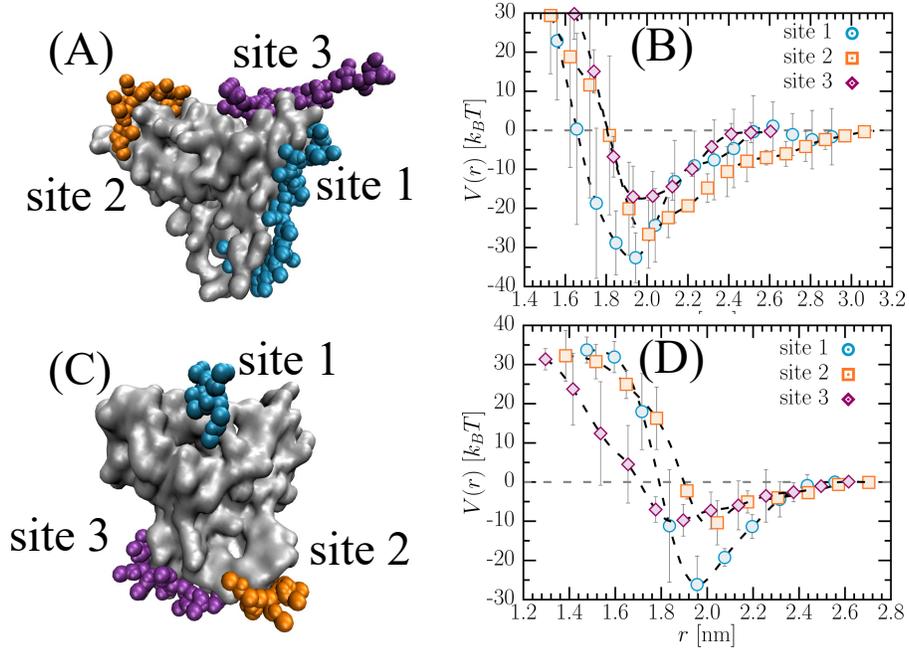


FIG. S12. The potential of mean force (PMF) profile $V(r)$ as a function of the protein-ligand center of mass distance r for the oligomer ligand $[N]_{20}$ (A-B) and $[N]_{10}$ (C-D). Panels (A,C) depict the complex structure of the protein bound with three ligands. Site 1/2/3 indicates the first, second and third ligand being adsorbed with the protein, respectively. The corresponding PMF profiles for those bound ligands are presented in panel (B,D).