Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary Information for: Insights into Substrate Behavior in a Solvent-free Protein Liquid to Rationalize its Reduced Catalytic Rate

Sudarshan Behera and Sundaram Balasubramanian*

Chemistry and Physics of Materials Unit

Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560 064, India

E-mail: bala@jncasr.ac.in

Contents

List of Tables

List of Figures

S1 Supplementary Tables

Table S1: Summary of all the simulations. (Abbreviations used are SFT: Surfactant, P: Protein, PNB: p-nitrophenyl butyrate, c4M2: cationized 4M2 mutant, [c4M2:31S]: cationized 4M2 with 31 surfactants, N_x: Number of x; x being water, SFT, and protein, N_{runs}: No. of independent runs)

Sr. No	Enzyme Form	N_{water}	N_{SFT}	N_P	P: PNB	Cubic box length (nm)	Total no. of atoms	N_{runs}	Trajectory length of each run (ns)
1	Aq. 4M2 (300K, 1bar)	8598	0	1	0	6.58	28,520	3	100
2	Aq. 4M2 with PNB (300K, 1 bar)	8590	0	1	1	6.58	28,522	20	100
3	Aq. c4M2 (300K, 1 bar)	12,672	0	0	0	7.44	41,072	1	100
4	[c4M2:31S] (300K, 1 bar)	66,556	31	1	0	12.82	2,09,203	1	100
5	SFPL (333K, 1 bar)	64x40 =2560	64x31 =1984	64	0	17.87	6,17,920	3	200
6	SFPL-PNB-0.5 (333K, 1 bar)	64x40 =2560	64x31 =1984	64	1:0.5	17.89	6,18,752	3	1000
7	SFPL-PNB-2 (333K, 1 bar)	64x40 =2560	64x31 =1984	64	1:2	17.92	6,21,248	1	1000
8	SFPL-PNB-472 (333K, 1 bar)	32x40 =2560	32x31 =1984	32	1:472	19.4	7,01,664	3	300

S2 Supplementary Text

S2.1 Modeling the 4M2 mutant

The 4M2 is a four site mutant with mutations F17E, A20E, G111D, and M134E (labeled in red in SI Figure S1). One of the mutations (M134E) is close to the active site (catalytic triad is Ser77-His156-Asp133), whereas the other three are located close to the hydrophobic, active site region. The mutant was modeled using the Mutagenesis Wizard of Pymol.[1] It

has the option to choose a standard amino acid residue from a protein and mutate to any other standard amino acid of interest. It provides various possible rotamers for the mutant residue to be selected from, manually. The four site mutations (4M2) were done sequentially, choosing the best rotamer for each mutated residue as the one having the least number of hard contacts with the enzyme.

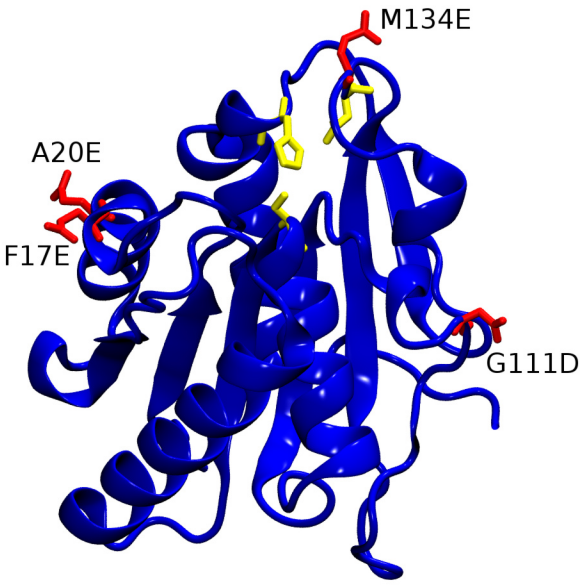

Figure S1: The 4M2 mutant, displaying the location of the four mutations (red licorice representation). The enzyme and the catalytic triad are shown in blue cartoon and yellow licorice representations respectively.

S2.2 Choice of the NVT over the NPT ensemble for production simulations

Prior to the NVT production, the systems were subjected to NPT equilibration. The proper equilibration in the NPT ensemble was checked by the convergence in the box volume (Add a figure). In an aqueous protein solution modelled under constant NPT conditions, the box length fluctuates around the mean to a decent extent; in contrast, the SFPL system is relatively incompressible. For instance, the simulation of aqueous LipA solution under NPT

conditions yields an average box volume of 284.7nm^3 with a standard deviation of 0.878. The simulation of LipA in SFPL form with 32 substrate molecules under NPT conditions yields an average box volume of 5726.6nm^3 with a standard deviation of 2.36; the standard deviation in the box volume (a measure of the box fluctuation amplitude) in the SFPL is around eight times smaller than that in the aqueous solution. Thus, even if one were to do the production run in the NPT ensemble (which we agree is chemically more realistic), within simulation time scales, no significant fluctuation in the box length would take place. Given the large system size and the marginal speed-up one gains by using NVT than NPT, made us do the production runs in the NVT ensemble. The aqueous protein solution simulations were carried out for the sake of comparing the results of the SFPL system and thus, they too were performed under NVT conditions.

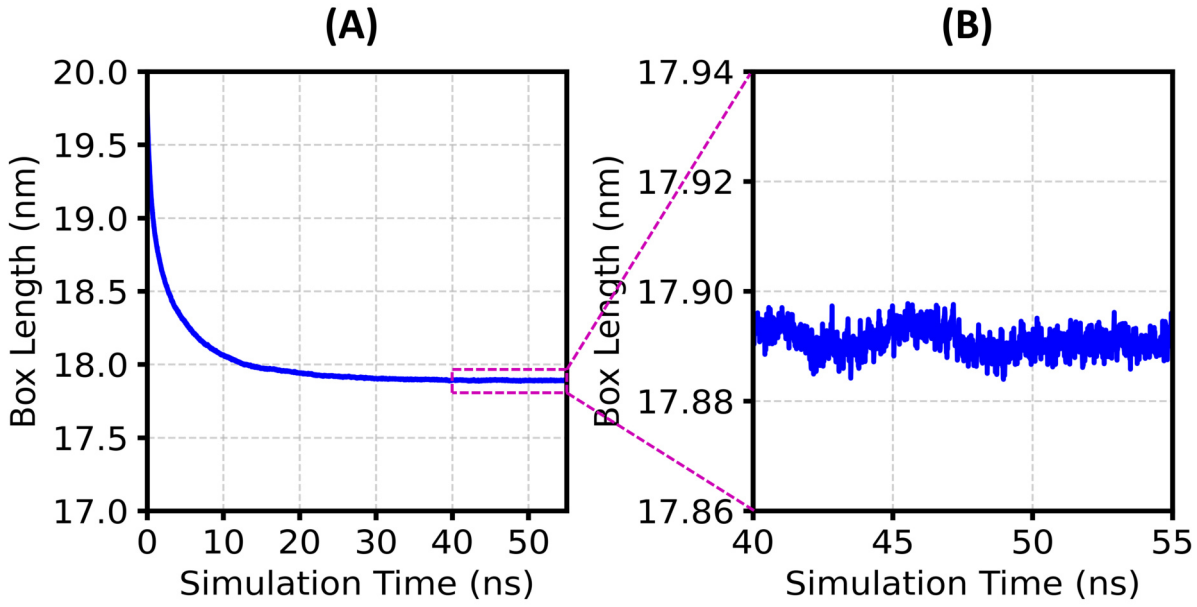

Figure S2: (A) The cubic box length for the SFPL-PNB-0.5 system as a function of simulation time for the last NPT equilibration step at T=333K and P=1bar, (B) the zoomed-in view of panel-A showing the convergence of box length for the last 15 ns.

S2.3 β(t) exponent of MSD

β(t) is calculated from the mean squared displacement (MSD) as follows:

$$\beta(t) = d[log(msd)]/d[log(t)] \tag{1}$$

The value of β(t) provides information on the diffusive nature of the system. A value of less than unity represents a sub-diffusive regime, whereas for diffusive behaviour, it converges to a value of 1.

S2.4 Calculation of the reorientational time autocorrelation function (RACF) of the substrate

The origin of the three vectors is defined to be the COM of the benzene ring of the substrate molecule (Figure 3A and 6A of the main manuscript). Vector X is defined as that from the origin to the C_4 atom of the substrate (Figure 3A and 6A of the manuscript). Y is the vector from the origin to the COM of C_6 and C_8 atoms. Z is the vector perpendicular to the plane formed by X and Y.

The RACF of a vector **v** is defined as follows:

$$C_v(t) = \langle(\mathbf{v}(t_0).\mathbf{v}(t_0 + t))\rangle_{t_0} \tag{2}$$

The vector at time t_0 and $t_0 + t$ are denoted as $\mathbf{v}(t_0)$ and $\mathbf{v}(t_0 + t)$, respectively. Angular brackets represent averaging over time. $C_v(t)$ is then averaged over all PNB molecules.

S2.5 Selecting the six high and low mobile substrate molecules

Individual MSD of the COMs of all the PNB molecules (32x3 = 96 from 3 independent trajectories of SFPL-PNB-0.5 case) was calculated. The MSD values at 100 ns time were sorted, and the six PNB molecules having the highest MSD values and the six PNB molecules

having the least MSD values were selected as the six high and six low mobile substrate molecules, respectively.

S2.6 Dynamical heterogeneity of PNB in the SFPL

Why are some substrate molecules almost still, while some are highly mobile? Given the highly sluggish behavior of the substrates, calculating their dynamical heterogeneity is ruled out. Thus, to answer this question, we examined the environment of the six high mobile and six low mobile PNB molecules (enzymes and surfactant molecules). The low mobile PNB molecules have significantly more interactions with enzymes than the high mobile ones (SI Figure S22). PNB molecules which interact most with the enzymes move less. The case is the opposite for those substrate molecules interacting with the surfactant (SI Figure S23). The high mobile PNBs interact mostly with the polar hydrophilic part of the surfactant (SI Figure S24). To summarize our observation, PNB molecules are predominantly located in the polar part of the surfactant layer around the enzymes and a minor fraction of them which are present on the enzyme surface are much less mobile than the ones located in the surfactant layer.

S2.7 Results and Discussion for the SFPL-PNB-472 system

The convergence of structure of the SFPL-PNB-472 was verified by the evolution of the RDF between the COM of Protein and COM of PNB (P-PNBRDF, SI Figure S3). The change in the RDFs become very small for the last three intervals (140-160ns, 160-180ns and 180 to 200ns in the SI Figure S3B). Hence, the 200-300 ns segment of the simulation trajectories (for all the three independent runs) were used for all the analyses reported in this study.

S2.7.1 Location of the substrates

With a very high concentration of PNB (P:PNB ratio of 1:472 as reported in the experiment), the system can be likened to a dipersion of PSW-complexes in liquid PNB. This can be

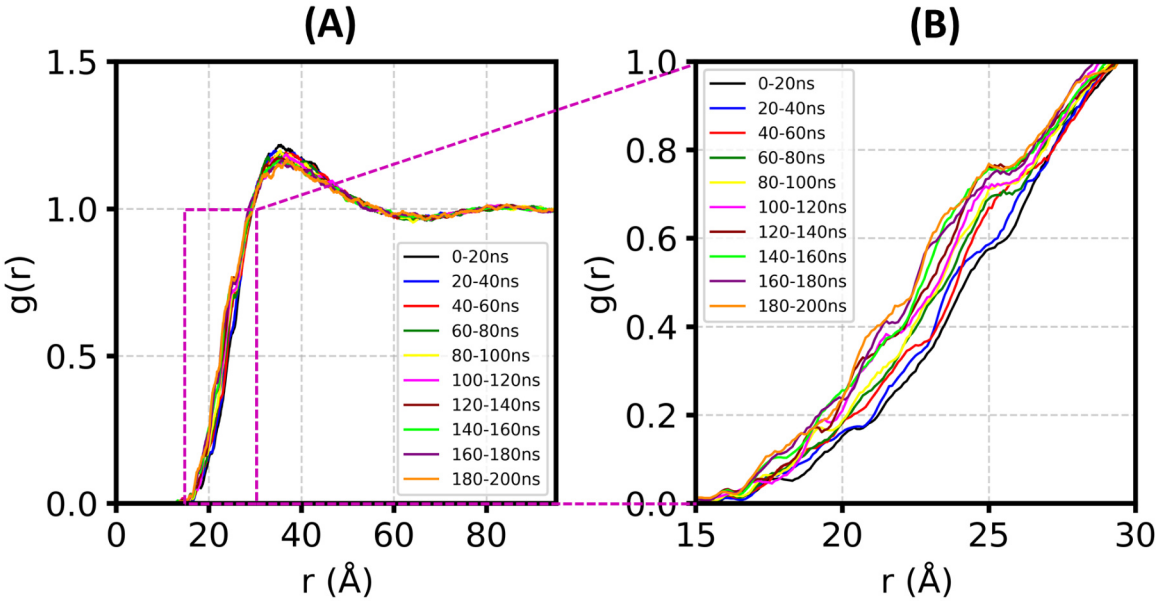

Figure S3: (A) The evolution of P-PNB RDF (RDF between Protein COM and PNB COM), (B) the zoomed-in view of panel-A. The change in the RDF becomes very small for the last three intervals (140-160ns, 160-180ns and 180 to 200ns).

discerned from the simulation box (SI Figure S4), where many regions of the box appear like bulk liquid PNB, depicted in green. It has been reported earlier that the SFPL of LipA (with Rg value of $\sim$ 15 Å) has a surfactant layer of thickness $\sim$ 13.5 Å.[2] The running coordination number (SI Figure S5B) of PNB with respect to radial distance from the COM of the enzymes (calculated from P-PNB g(r), SI Figure S5A) is $\sim$100 at 28.5 Å (the radius of Protein+surfactant layer), which is $\sim$20% of the total number of PNB molecules per enzyme in the system. Thus, on average, only $\sim$20% of the PNBs are present close to the PSW-complexes and nearly $\sim$80% of the substrate molecules are present in a bulk liquid PNB-like environment.

Water molecules are present close to the enzyme (either buried or on the surface of the enzyme) even in the SFPL-PNB-472 system, as depicted from the number density profile of water (SI Figure S5C). Very few PNB molecules are present close to the enzymes, while the closest moieties are the head groups of the surfactants. However, in regions where surfactants are not present (such as near the non-positively charged residues), PNB molecules

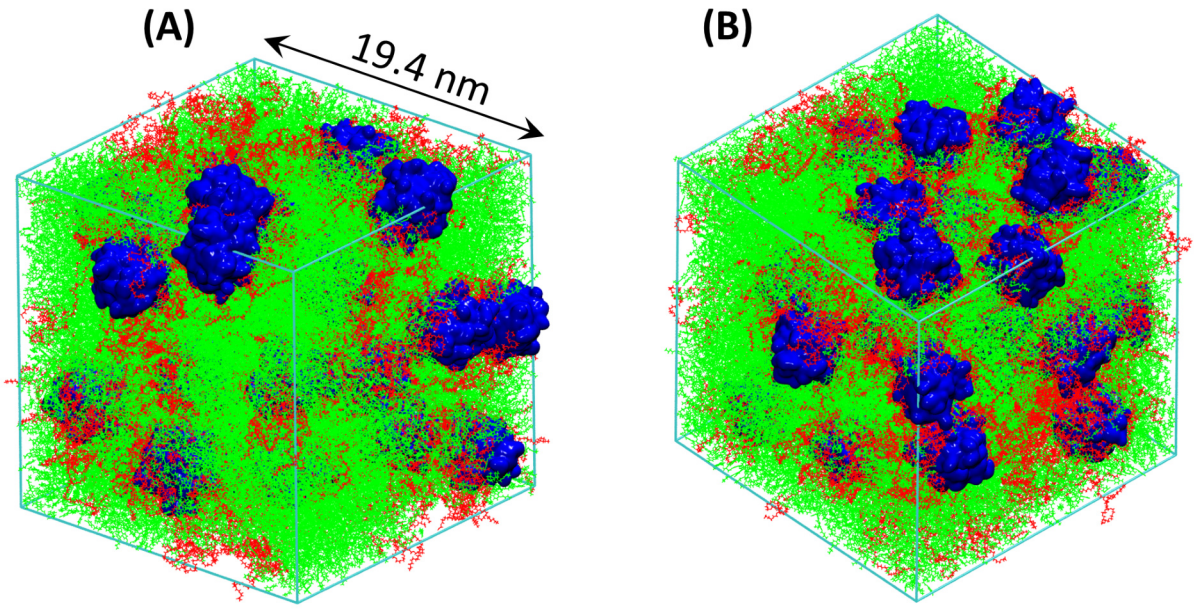

Figure S4: Two snapshots (A and B) of equilibrated configurations from two independent SFPL-PNB-472 simulations. The color scheme: blue surfaces for enzymes (32 enzymes), red lines for surfactants and green lines for PNB molecules (32 x 472 = 15,104 PNBs).

can be expected to be present near the enzyme surface. An example is the enzyme-substrate complex presented in the main MS. Unlike the SFPL-PNB-0.5 and SFPL-PNB-2 systems, PNB molecules tend to interact more with the alkyl part of the surfactant than the PEG parts. The PNB-Ca g(r) has a larger peak heights (both 1st and 2nd peaks) than the PNB-Cp and PNB-Op RDFs. This is because of the fact that a larger fraction of PNB molecules are located far from the enzyme. Since the alkyl parts of the surfactants are projected out from the enzymes, they have more PNB molecules to interact with.

S2.7.2 Conformations of the substrates

A substrate molecule, in the SFPL-PNB-0.5 and SFPL-PNB-2 systems, is observed to exhibit various conformations with similar propensities as in the aqueous protein solution (refer to Figure 3 of the manuscript, and SI Figure S17). It is no surprise to notice a similar conformational behavior of PNB in the SFPL-PNB-472 system (SI Figure S6), as ∼80% of the substrate molecules are in a bulk liquid PNB like medium and expected to have higher

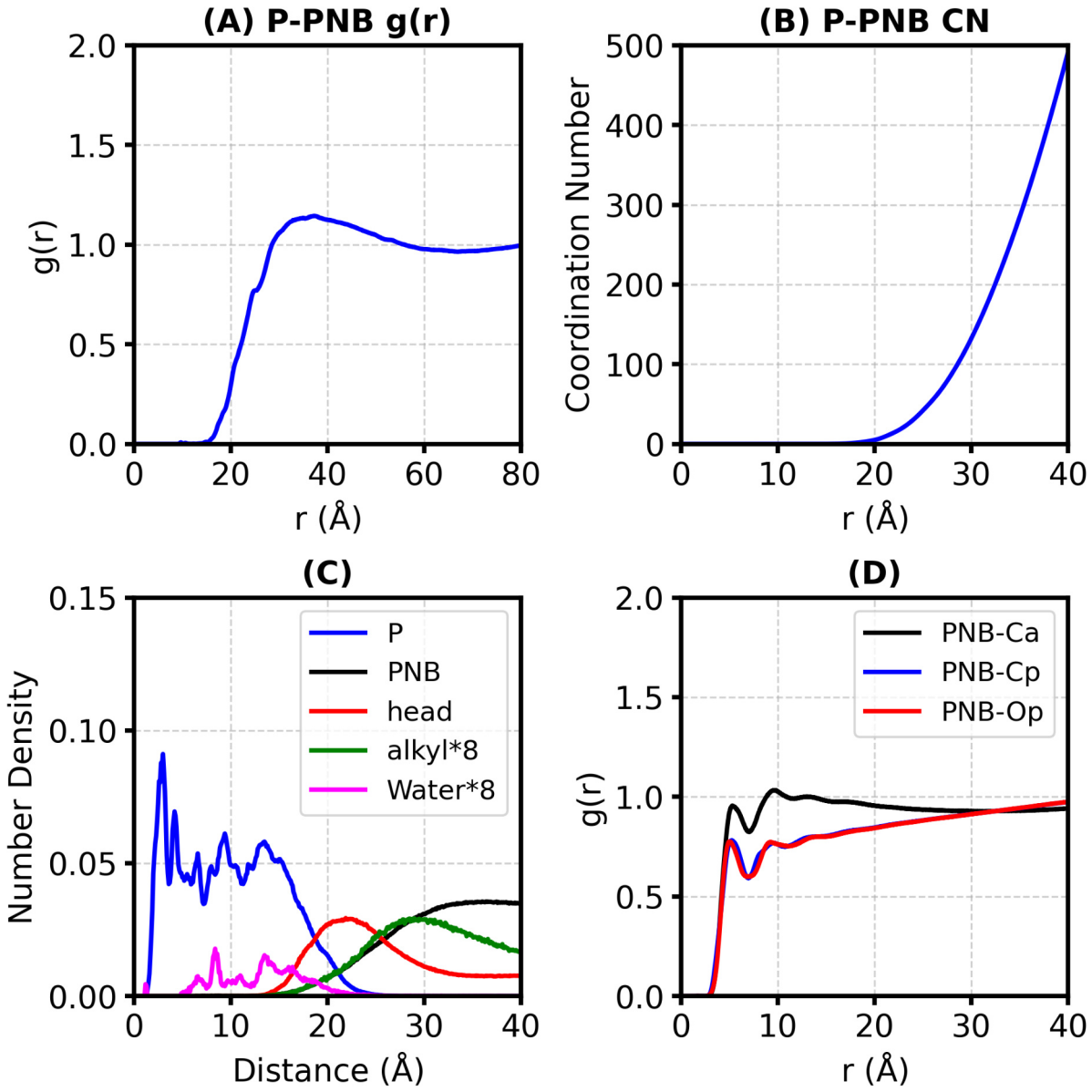

Figure S5: (A) Radial distribution function (g(r)) between the centers of mass of the protein (P) and PNB in the SFPL-PNB-472, and (B) the running coordination number derived from the g(r) of Panel-A. (C) The number density profile (nDP, number of non-hydrogen atoms per Å^3) of different components of SFPL-PNB-472. Keywords used are "P" for Protein, "PNB" for p-nitrophenyl butyrate (the substrate), "head" for the hydrophilic part of the surfactant (PEG and carboxylate part), "alkyl" for the alkyl tail of the surfactant. The label "alkyl*8" means the number density of the alkyl tail is multiplied by a factor of 8 so that it appears on the same scale as that for protein. (D) Radial distribution function between the center of mass of PNB and different types of heavy atoms of the surfactant in the SFPL-PNB-472. Ca, Cp and Op are the carbon atom of the alkyl tail, the carbon atom of PEG and the oxygen atom of PEG part of the surfactant, respectively.

conformational flexibility than in the SFPL-PNB-0.5 and SFPL-PNB-2 cases.

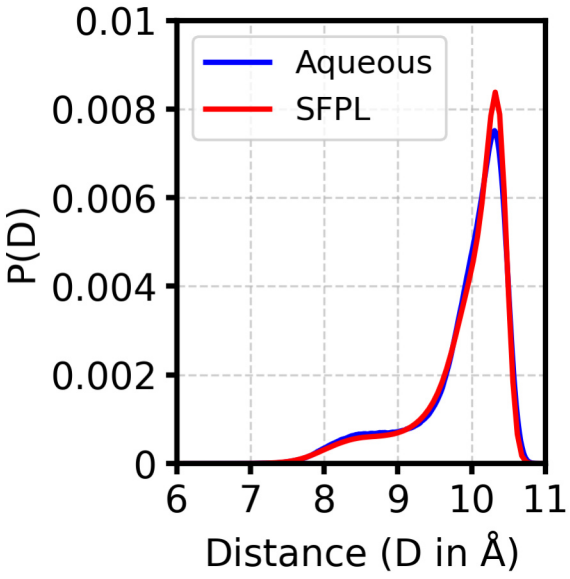

Figure S6: The end-to-end distance of the substrate (same as explained in Figure 3 of the manuscript) of SFPL-PNB-472 system compared with that of the aqueous solution.

S2.7.3 Translational dynamics of the substrates

The substrate molecules are significantly more mobile in the SFPL-PNB-472 than in the SFPL-PNB-0.5 and SFPL-PNB-2 systems. The MSD value at 50 ns is 30 Å^2 and 300 Å^2 for SFPL-PNB-0.5 (and SFPL-PNB-2, Figure 4 of the manuscript) and SFPL-PNB-472 (SI Figure S7), respectively. The one order of magnitude higher value of MSD in the SFPL-PNB-472 can be attributed to the higher number of PNB molecules (∼80%) in a bulk liquid PNB like environment. However, the MSD values are substantially smaller than the aqueous solution (SI Figure S18). The nature of translational dynamics of the substrate for the SFPL-PNB-472 remains sub-diffusive in the simulation time scale (SI Figure S7B).

S2.7.4 Rotational dynamics of the substrates

The rotational dynamics of the substrate molecules has also become substantially faster in the SFPL-PNB-472 (SI Figure S8B) than in the SFPL-PNB-0.5 and SFPL-PNB-2 (Figure

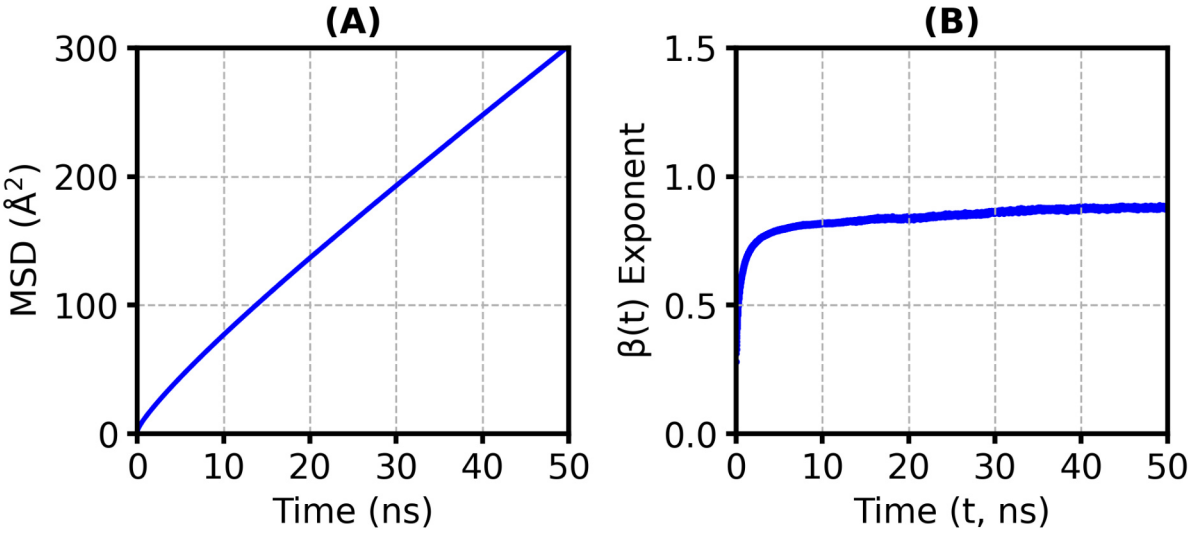

Figure S7: (A) The MSD of the COM of PNB molecules, averaged over three independent trajectories of the SFPL-PNB-472 system, and (B) the β-Exponent as a function of simulation time.

6B), but remains slower compared to that in aqueous solution (SI Figure S8A). The faster rotational motion of substrate in the SFPL-PNB-472 than in the SFPL-PNB-0.5 and SFPL-PNB-2 can be attributed to the same reason as described for the increase in the translational dynamics, i.e, a larger fraction of the PNB molecules are present in a bulk liquid PNB like medium and hence have a faster rotational and translational dynamics.

S2.7.5 Interactions of the substrate molecules with the enzymes

At the highest concentration of PNB molecules (SFPL-PNB-472 system), a greater number of substrates interact with the protein surface. ∼100 substrate molecules per enzyme are found to be present on the protein surface and surfactant layer (CN of PNB from COM of protein is ∼100 at 28.5 Å, SI Figure S5B). The P($R_i d$) values of the amino acid residues show that a large number of residues (145 out of 181) interact with substrate molecule(s) throughout the entire course of simulation time (with P(R_{id}) values of more than 0.002; the P(R_{id}) for residues interacting with PNB in all MD frames (10000 frames) = (10000/10000)*(1/472) = 0.002118). The relatively large number of PNB molecules, present close to the enzyme, makes

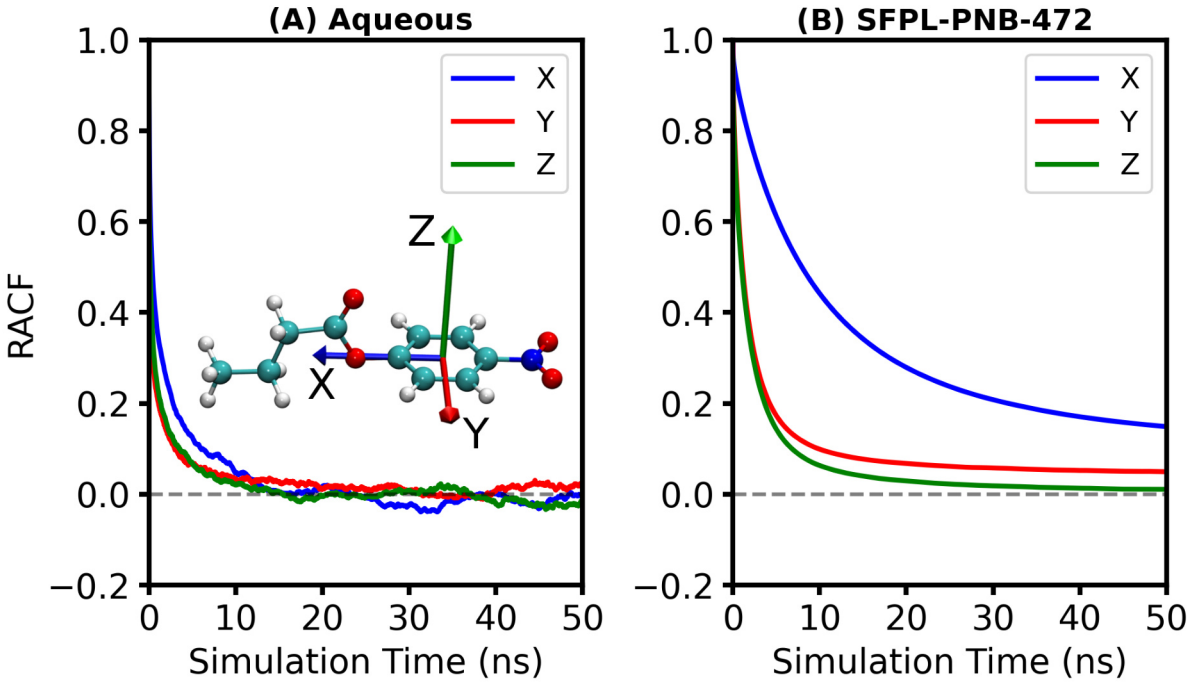

Figure S8: The reorientational autocorrelation functions (RACF) of the three molecular vectors for (A) aqueous solution, and (B) SFPL-PNB-472 systems.

them available for almost all the amino acid residues to interact with. Another fascinating observation is the formation of the Michaelis complex in the SFPL-PNB-472 system, which is elaborated in Section-3.6 of the manuscript.

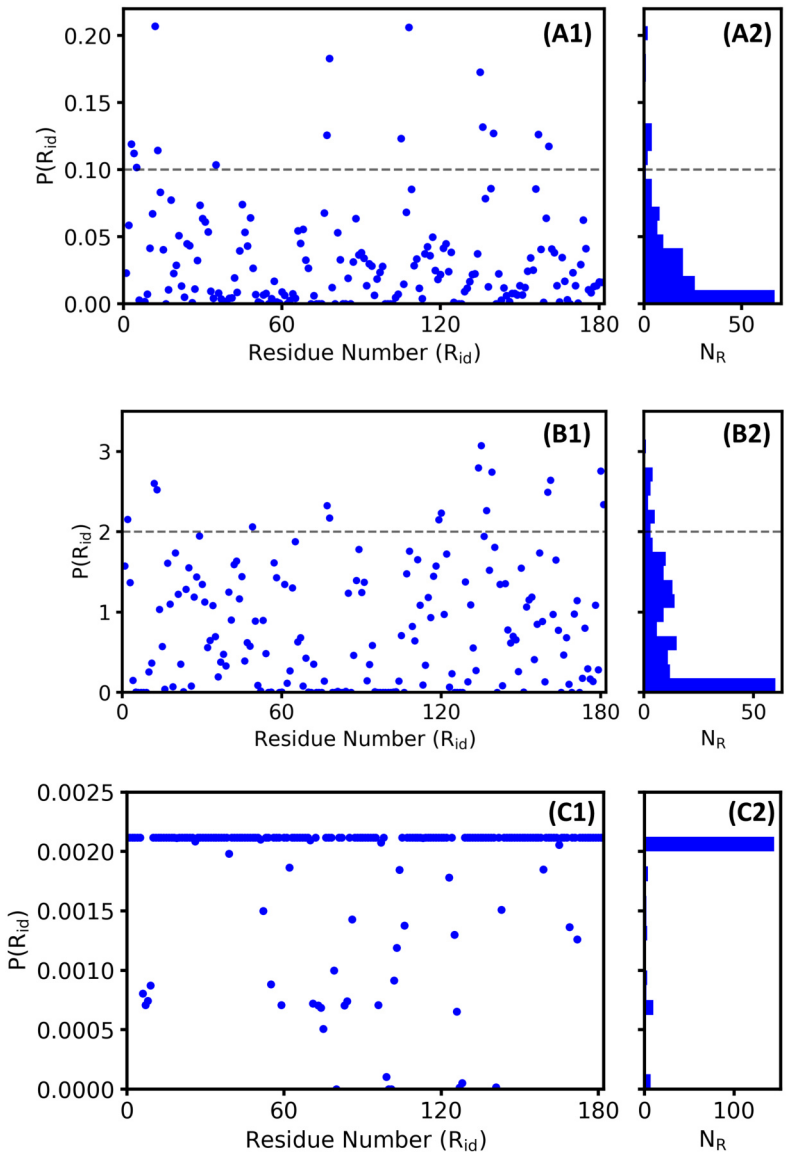

Figure S9: P(R_{id}) of all the amino acid residues (standard and cationized) for the (A1) aqueous, (B1) average of SFPL-PNB-0.5 and SFPL-PNB-2, and (C1) SFPL-PNB-472 case. P(R_{id}) of a residue (R_{id}) is defined as the number of frames the residue interacts with PNB molecule(s) divided by the P:PNB stoichiometry (0.5, 2, and 472 for SFPL-PNB-0.5, SFPL-PNB-2 and SFPL-PNB-472, respectively) and the total number of frames. The distribution of P(R_{id}) are plotted in (A2), (B2) and (C2), respectively. Panel (A2), (B2) and (C2) share the same Y-axes as (A1), (B1) and (C1), respectively and N_R is the number of residues. The horizontal dashed lines at 0.1 and 0.033 in the top two panels are the cutoffs used to identify the residues which interact with a substrate molecule the most. For the SFPL-PNB-472, ∼80% of the total number of residues (145 out of 181) are interacting with PNB molecules throughout the entire course of simulation time, with P(R_{id}) values of more than 0.002.

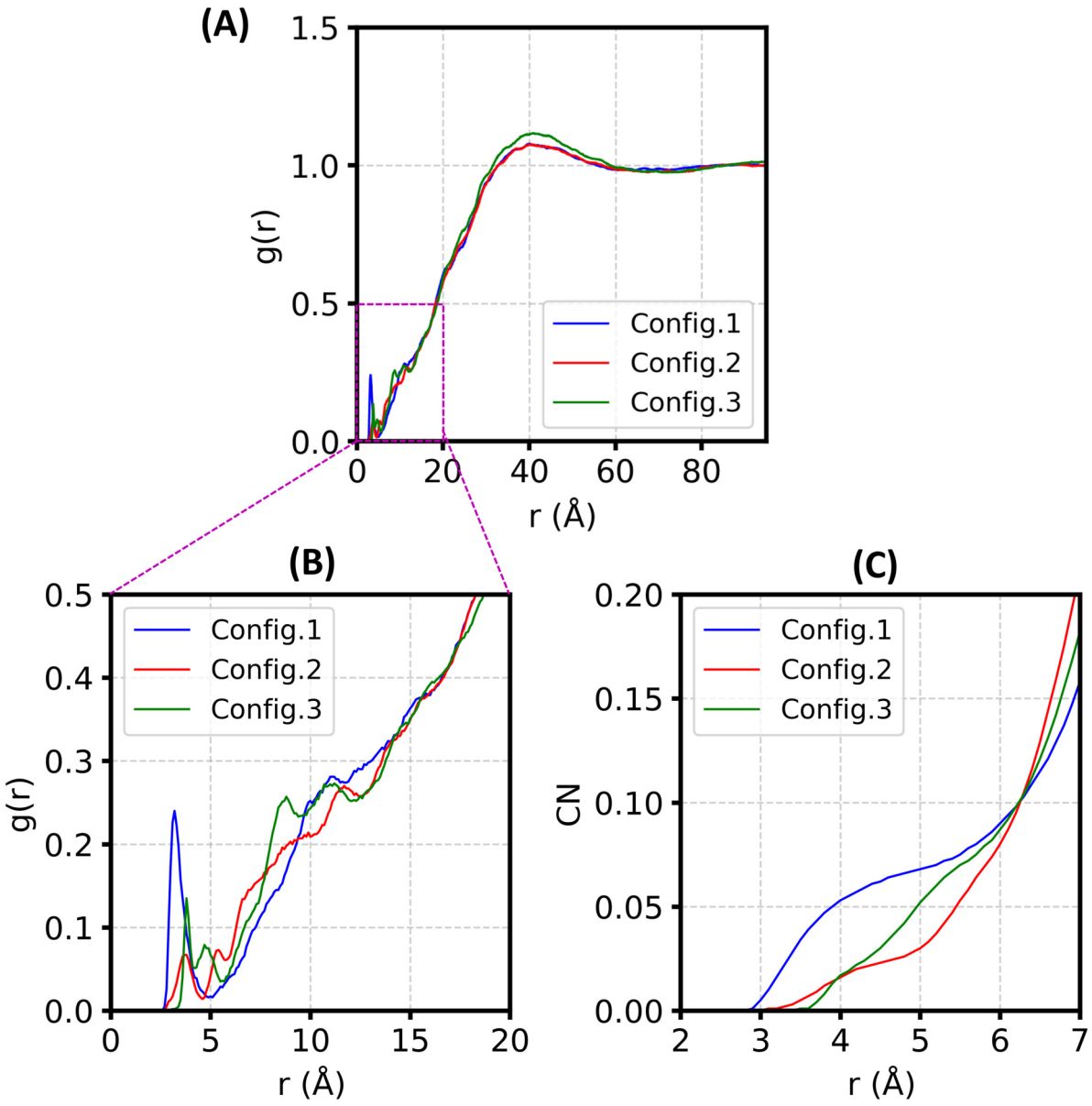

Figure S10: Pair correlation between active site residue Ser77 and the substrate. (A) The OG-Ca (refer to Figure 8 of the main manuscript) g(r), (B) the zoomed-in view of panel-A focusing on the small peaks at ∼3.5Å, and (C) the running coordination number derived from the OG-Ca g(r).

S2.8 Supplementary Movies

S2.8.1 Trajectory of an SFPL simulation (SM1.mp4)

The full trajectory (1 μs) of an SFPL-PNB-0.5 simulation is presented. Enzymes and surfactants are shown in blue cartoon and red transparent surface, respectively. Substrates are displayed as green spheres. The movie is focused on the mobilities of substrate molecules in the SFPL-PNB-0.5. Water molecules are not shown.

S2.8.2 Binding of substrate to LipA active site in the aqueous solution (SM2.mp4)

Four trajectories (100 ns each) of substrate and LipA in the aqueous medium are presented, where the substrate binds (and later leaves) the active site region of LipA. The enzyme and the catalytic triad are shown in orange cartoon and yellow licorice representations. Substrate molecules are shown in sphere representation. Water molecules are not displayed.

S3 Other Supplementary Figures

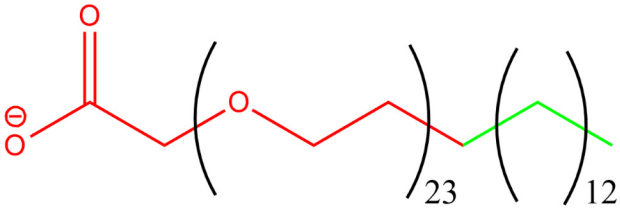

Figure S11: Molecular structure of the carboxylated Brij-L23 surfactant. The hydrophilic head part (PEG and carboxylate group) is shown in red, while the hydrophobic alkyl part is shown in green.

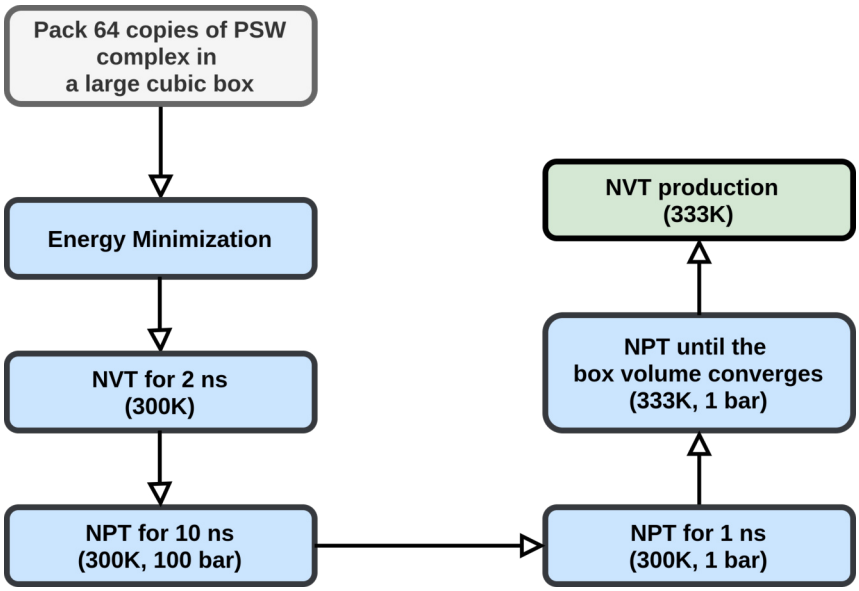

Figure S12: Flowchart of the protocol followed to simulate the SFPL. PSW refers to the Protein-Surfactant-Water complex. Configurations at Stage VI of the flowchart were also visually examined for any pockets of locked-in vacuum regions. If found, the configurations were further relaxed.

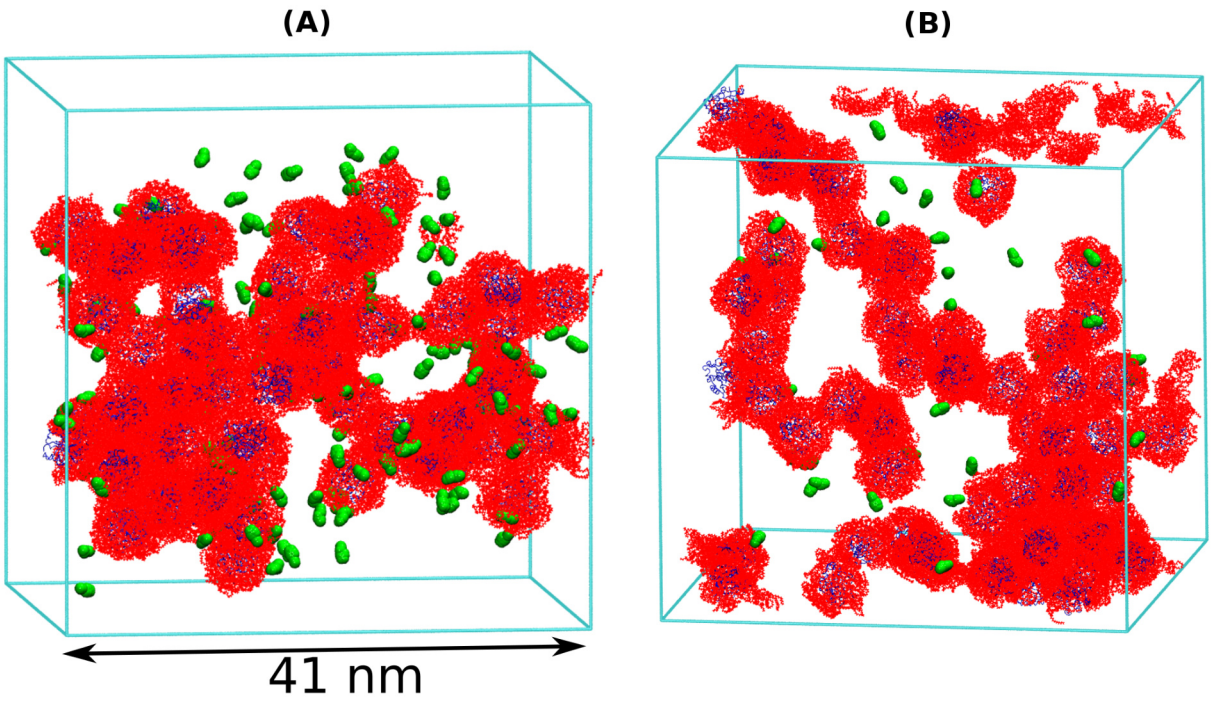

Figure S13: Snapshots showing the locations of the initial positions of PNB molecules in (A) the SFPL-PNB-2 and (B) the SFPL-PNB-0.5 systems. Color scheme: blue cartoon for enzymes, red line for surfactants, and green ball for PNBs.

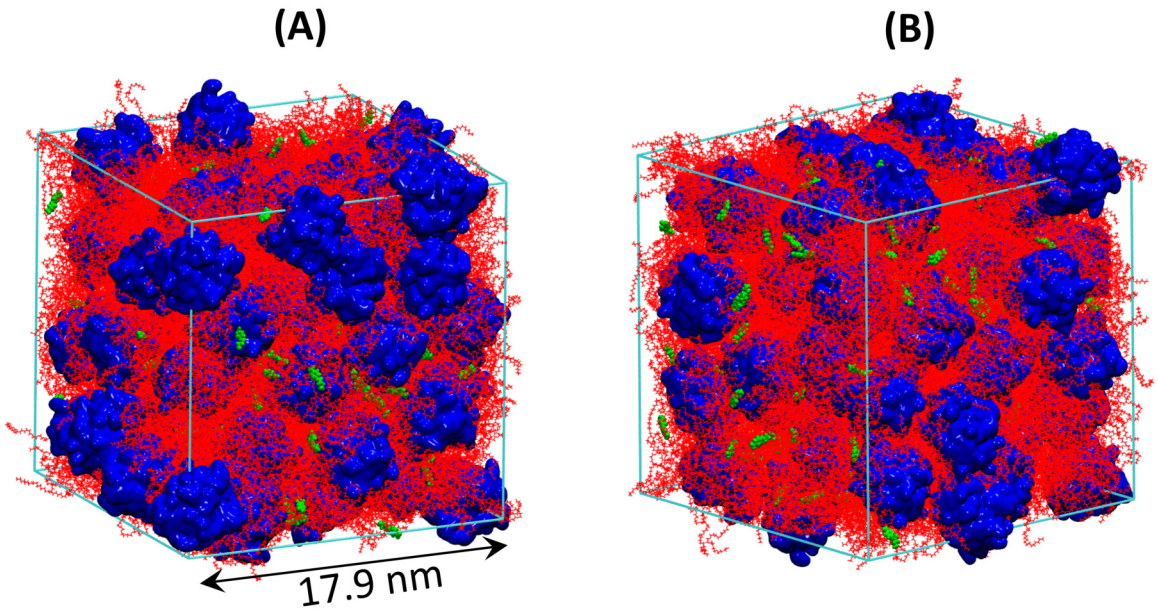

Figure S14: Two different views (A and B) showing the final configuration of SFPL-PNB-2, with color scheme: blue surface for the enzymes (64 enzymes), red line for the surfactants, and green ball for PNBs (128 PNBs).

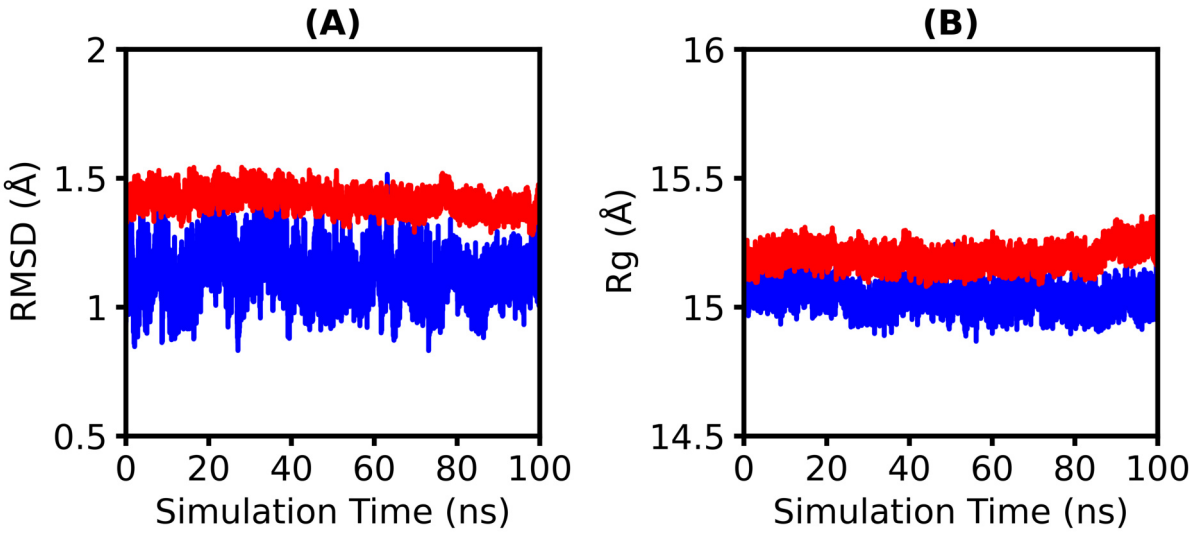

Figure S15: (A) RMSDs of backbone atomic positions of enzymes in aqueous (blue) and SFPL (red) media considering the crystal structure as the reference, (B) Radii of gyration of the enzymes. These results are from one independent trajectory and averaged over all the proteins in the simulation box (one for aqueous and sixty-our for SFPL-PNB-0.5 system).

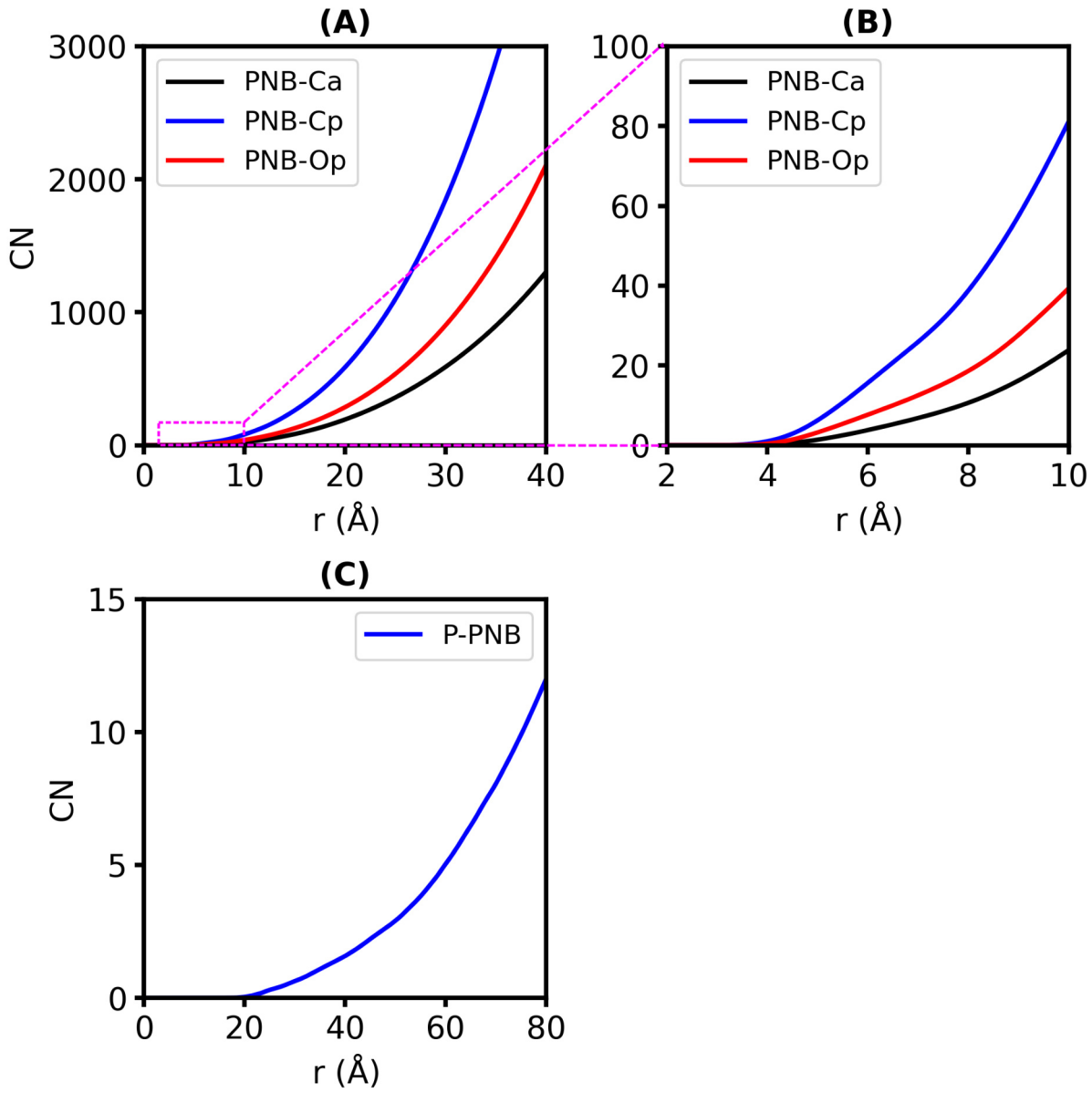

Figure S16: (A) Running coordination number derived from the RDFs with the same naming convention as in Figure 2C of the main manuscript, (B) the zoomed-in view of panel A focusing on the coordination numbers at the first coordination shell (the first minimum in the RDF plots, which is 7.4 Å). The coordination number at the first minimum in the RDF are 8, 16 and 32 for PNB-Ca, PNB-Op and PNB-Cp, respectively. There are two ethylene glycol groups (-CH2-CH2-O-) present for one alkyl carbon (Ca) around a substrate molecule. (C) Running coordination number derived from P-PNB RDF (Figure 2D of the manuscript).

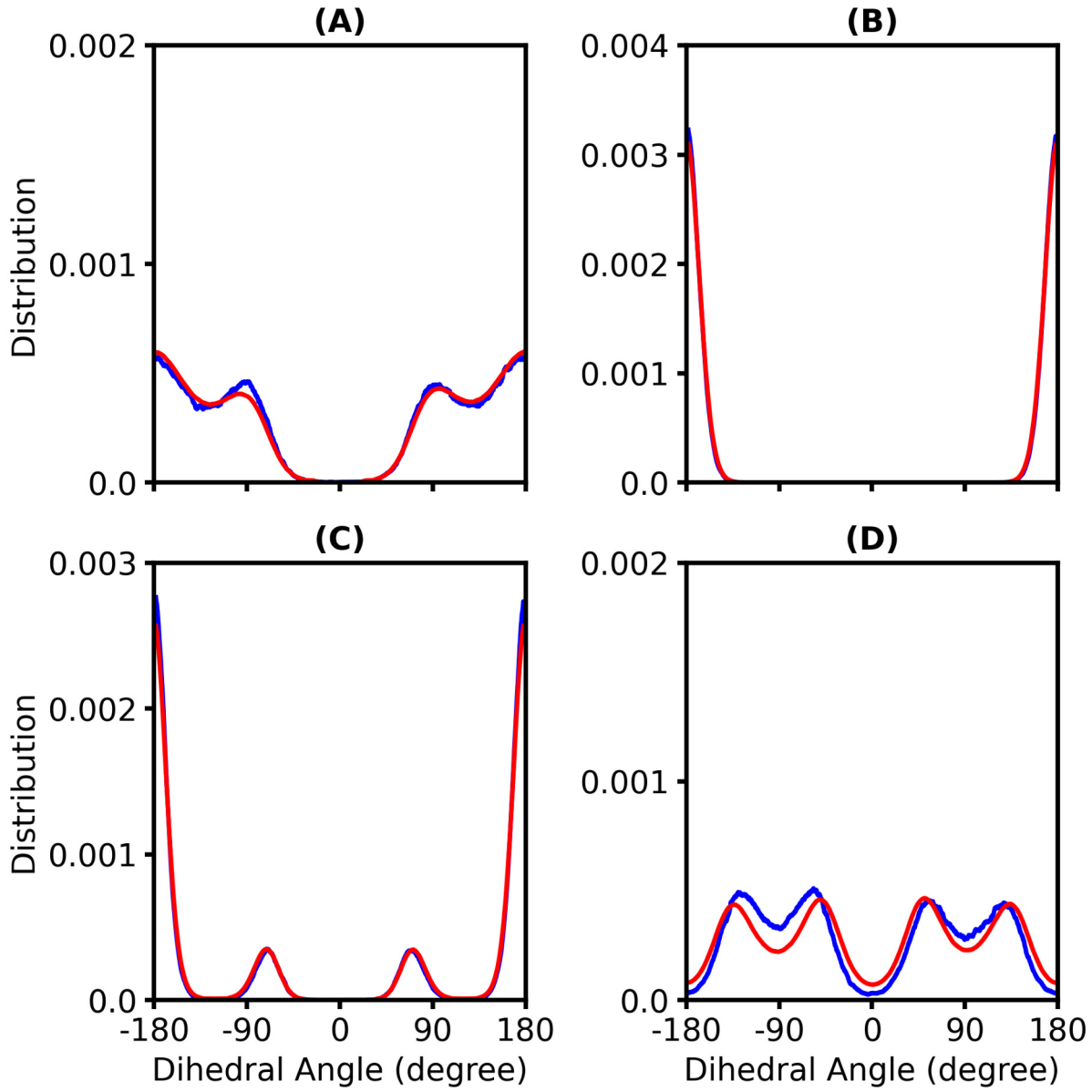

Figure S17: Dihedral distributions of the four flexible dihedrals of PNB (A) C_1-C_2-C_3-O_1, (B) C_2-C_3-O_1-C_4, (C) C_3-C_2-C_1-C, and (D) C_3-O_1-C_4-C_5. Refer to Figure 3A of the manuscript for the molecular structure and atom names. Blue and red colors are for the substrate present in the aqueous and SFPL environments, respectively.

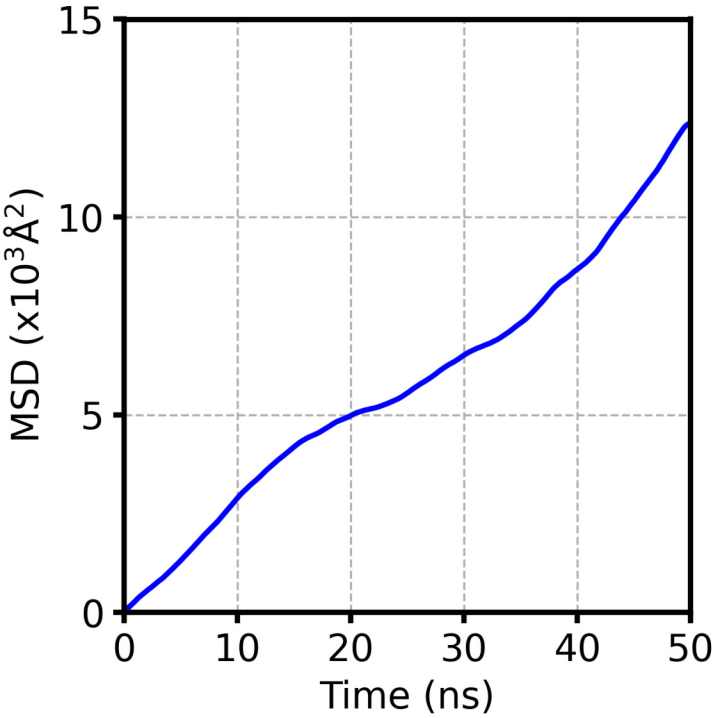

Figure S18: Mean-squared-displacement (MSD) of the COM of substrate in the aqueous system averaged over 20 independent trajectories.

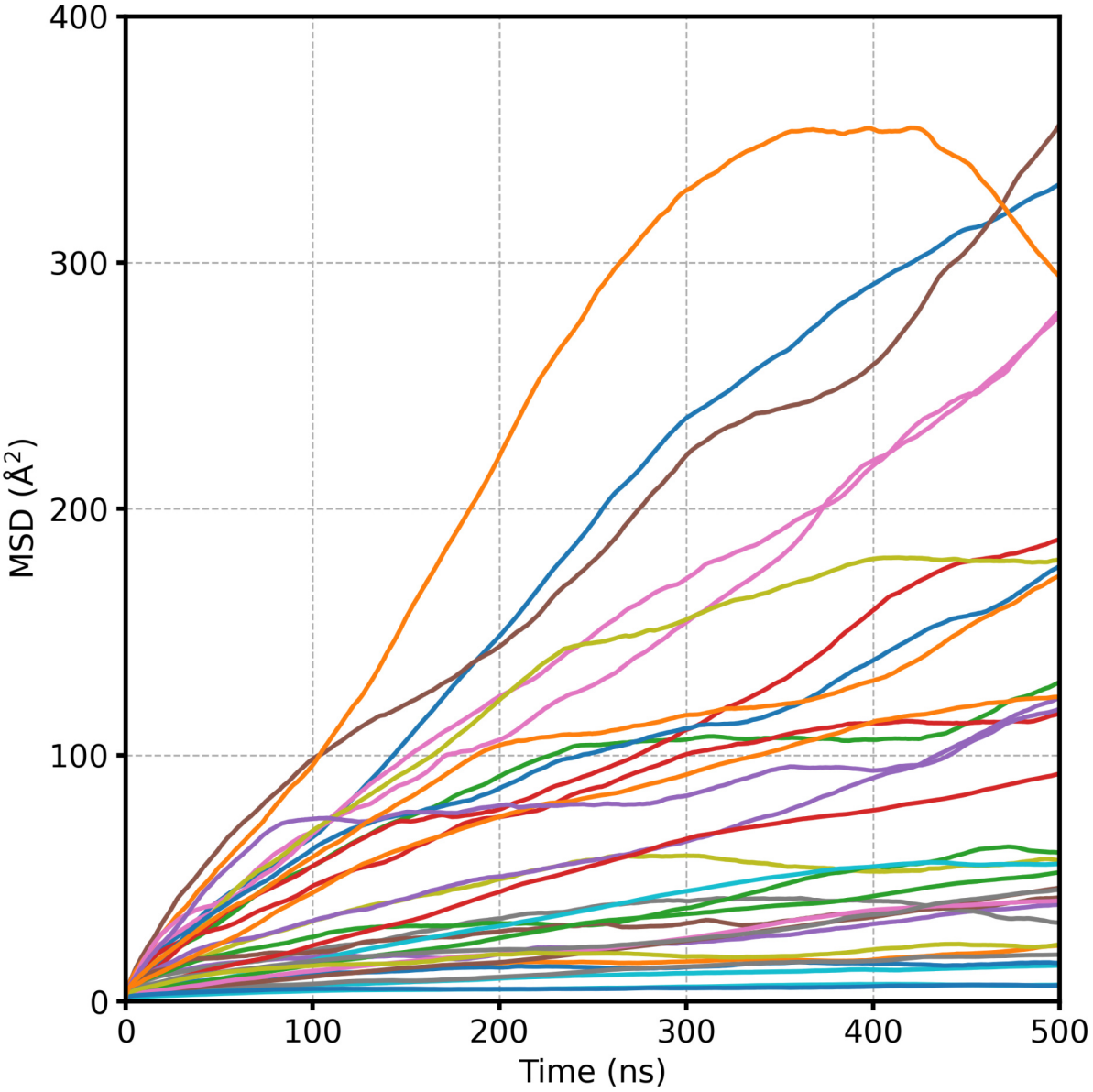

Figure S19: Individual MSD of the COMs of the substrate molecules (overall 32 in number) in an SFPL-PNB-0.5 simulation (out of the three independent trajectories) as a function of time. The substrate molecules show a high range of translational motion, some being almost immobile while some exhibiting relatively large displacements.

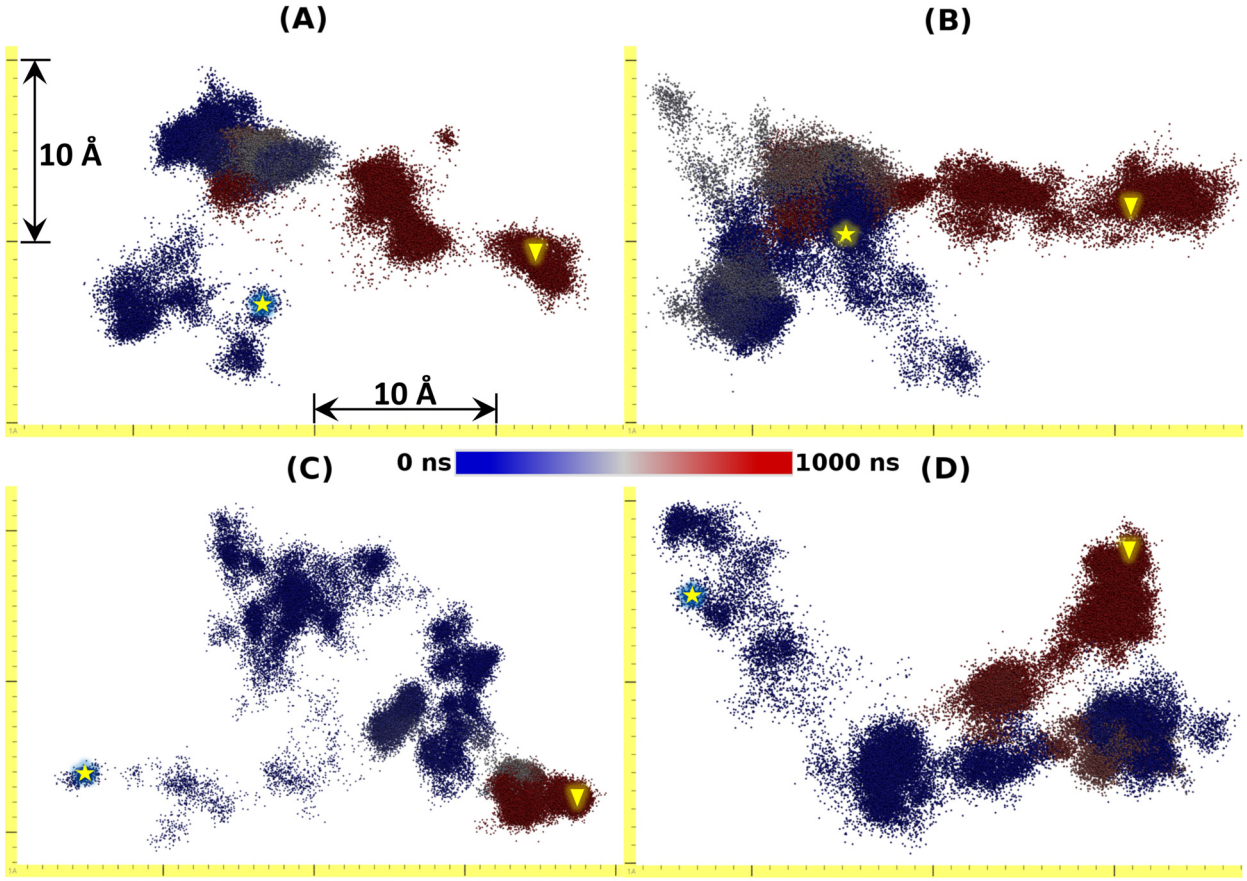

Figure S20: Motion of the center of mass of four high mobile PNB molecules with color scale from blue (early time) to red (later time) over 1 μs of simulation time. Each dot represents the position of the COM of the PNB (projected into a 2D space) at a particular time frame. Two adjacent dots are separated by 10 ps. The ruler with yellow color depicts the length scale, with major and minor ticks representing 10 Å and 1 Å, respectively. The starting and end points of the trajectory are marked by yellow star and yellow inverted triangle, respectively.

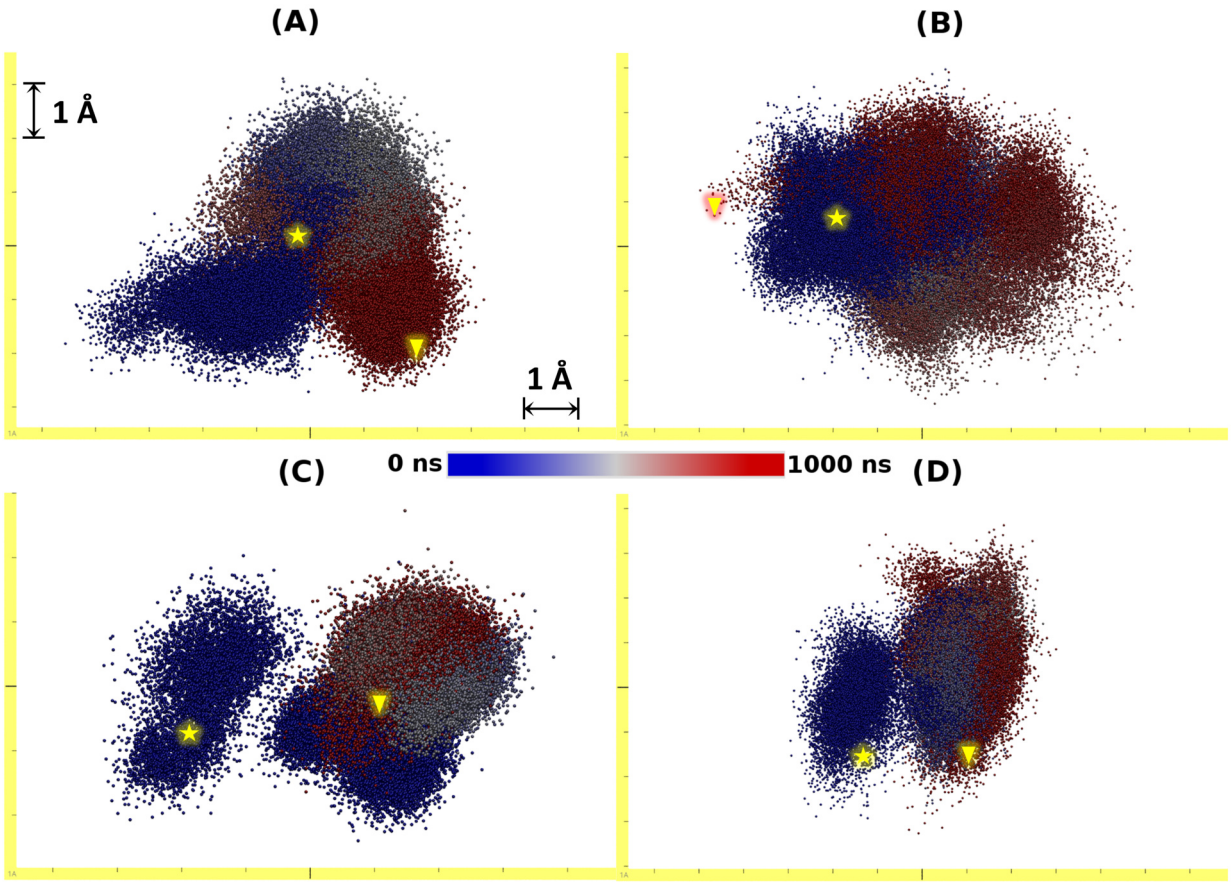

Figure S21: Motion of the center of mass of four PNB molecules having low mobilities, with color scale from blue to red over 1 μs of simulation time, where each dot represents the position of the COM of the PNB (projected into a 2D space) at a particular time frame. Two adjacent dots are separated by 10 ps. The ruler with yellow color depicts the length scale, with major and minor ticks representing 10 Å and 1 Å, respectively. The starting and end points of the trajectory are marked by yellow star and yellow inverted triangle, respectively.

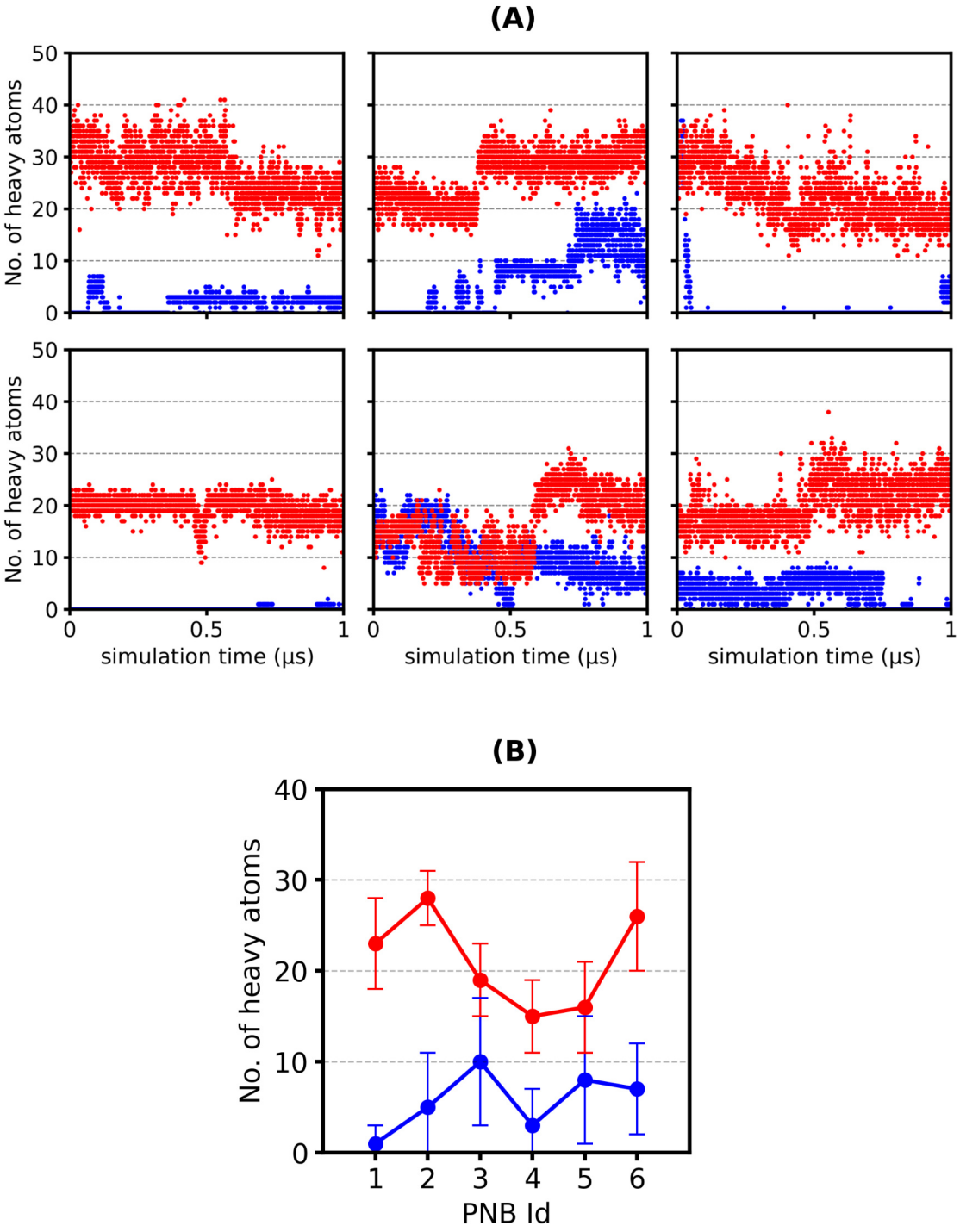

Figure S22: The number of non-hydrogen atoms of the enzymes, which interact (within 5 Å of non-hydrogen atoms of PNB) with the six low mobile (red dots) and six high mobile (blue dots) PNB molecules in the SFPL system, (A) as a function of simulation time and (B) their mean (circles) and standard deviation (error bars). PNB molecules which interact more with the enzymes move less than the ones which interact less with the enzymes.

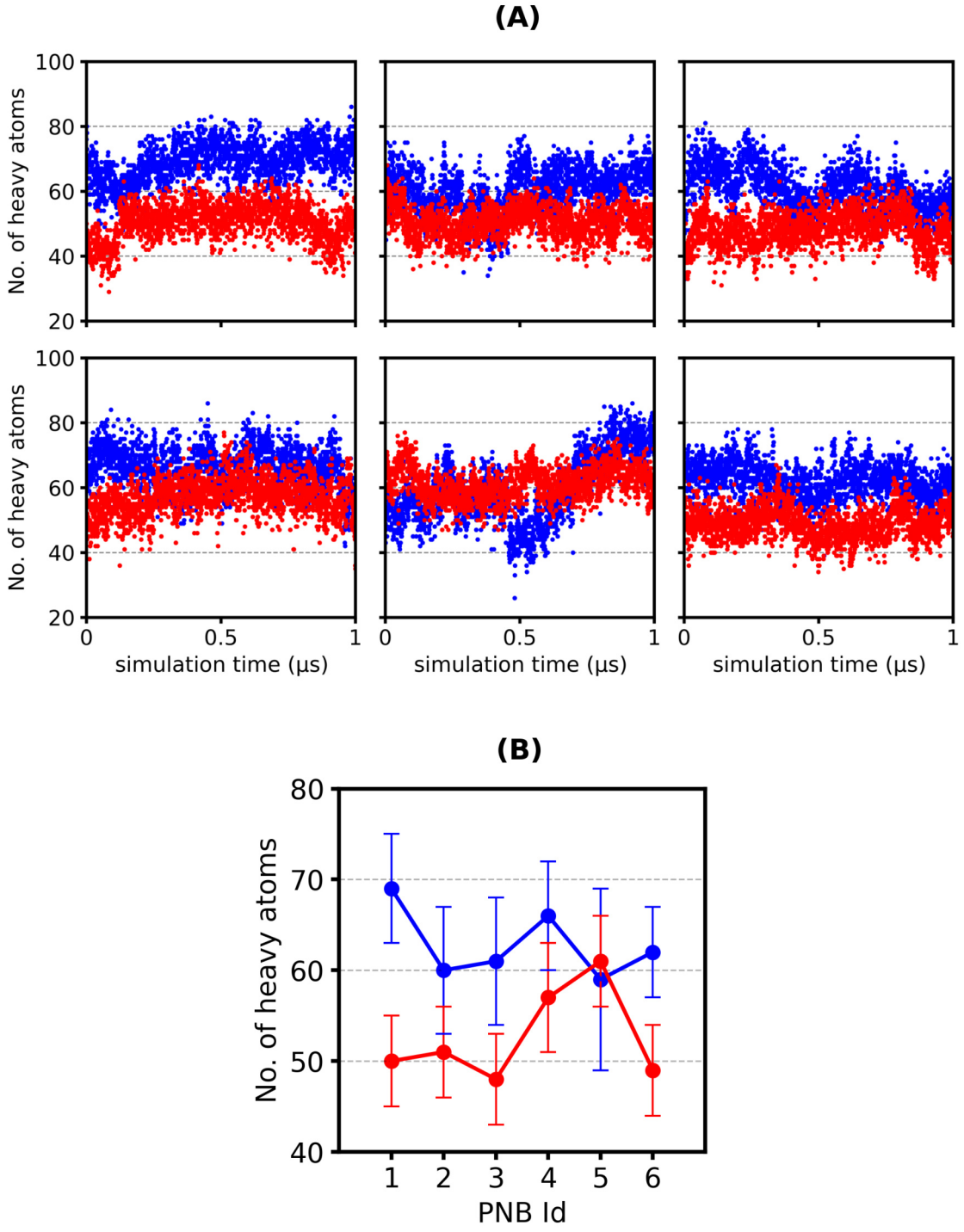

Figure S23: The number of non-hydrogen atoms of the surfactants, which interact (i.e., present within 5 Å of non-hydrogen atoms of PNB) with the six low mobile (red dots) and six high mobile (blue dots) PNB molecules in the SFPL system, (A) as a function of simulation time and (B) their mean (circles) and standard deviation (error bars). PNB molecules which interact more with the surfactants move more than the ones which interact less with the surfactants.

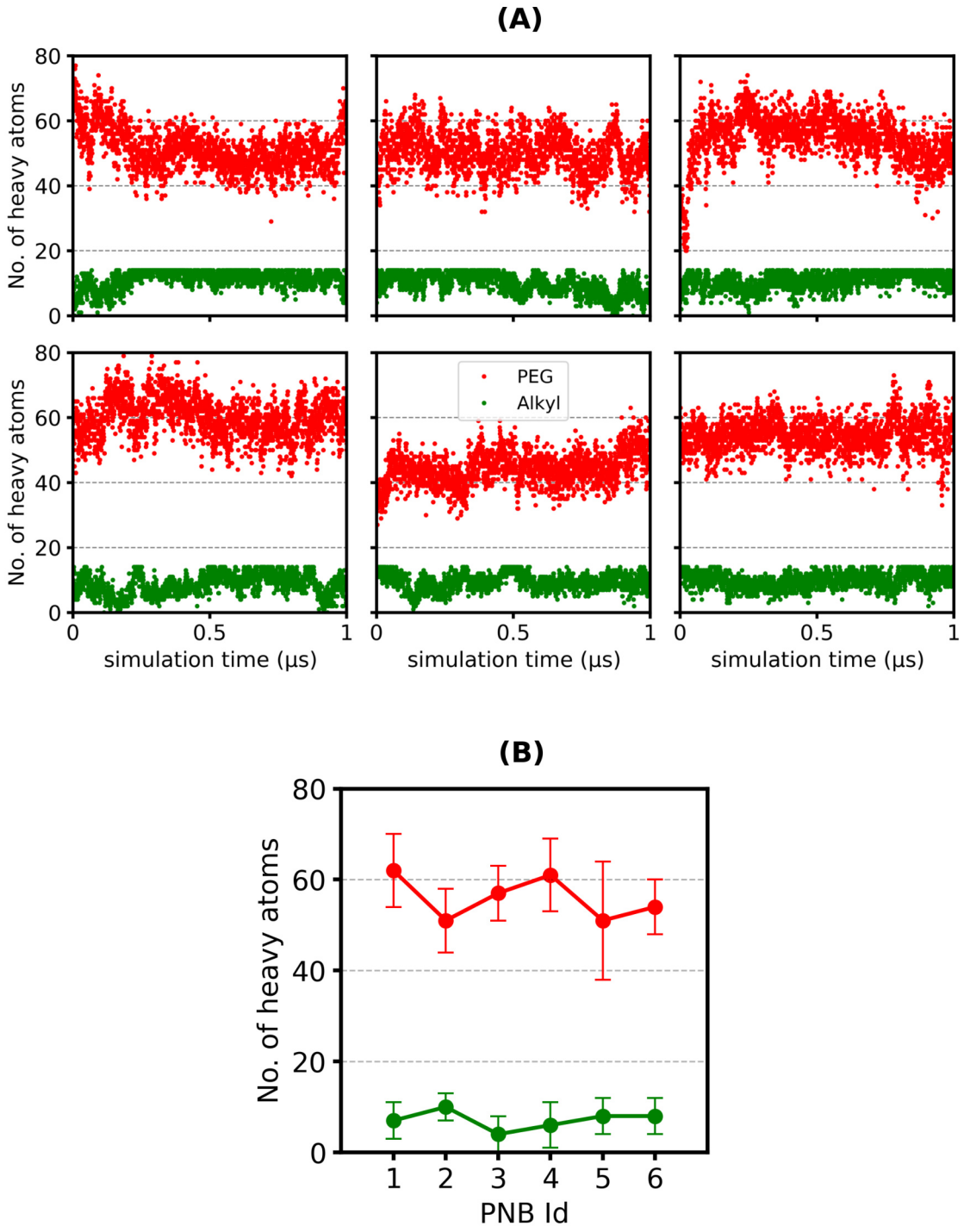

Figure S24: The number of non-hydrogen atoms of the polar part (PEG, red dots) and non-polar (Alkyl, green dots) of the surfactants, which interact (i.e., lie within 5 Å of non-hydrogen atoms of PNB) with the six PNB molecules showing high mobilities in the SFPL system, (A) as a function of simulation time and (B) their mean (dots) and standard deviation (error bars). Substrate molecules largely move through the polar part of the surfactants.

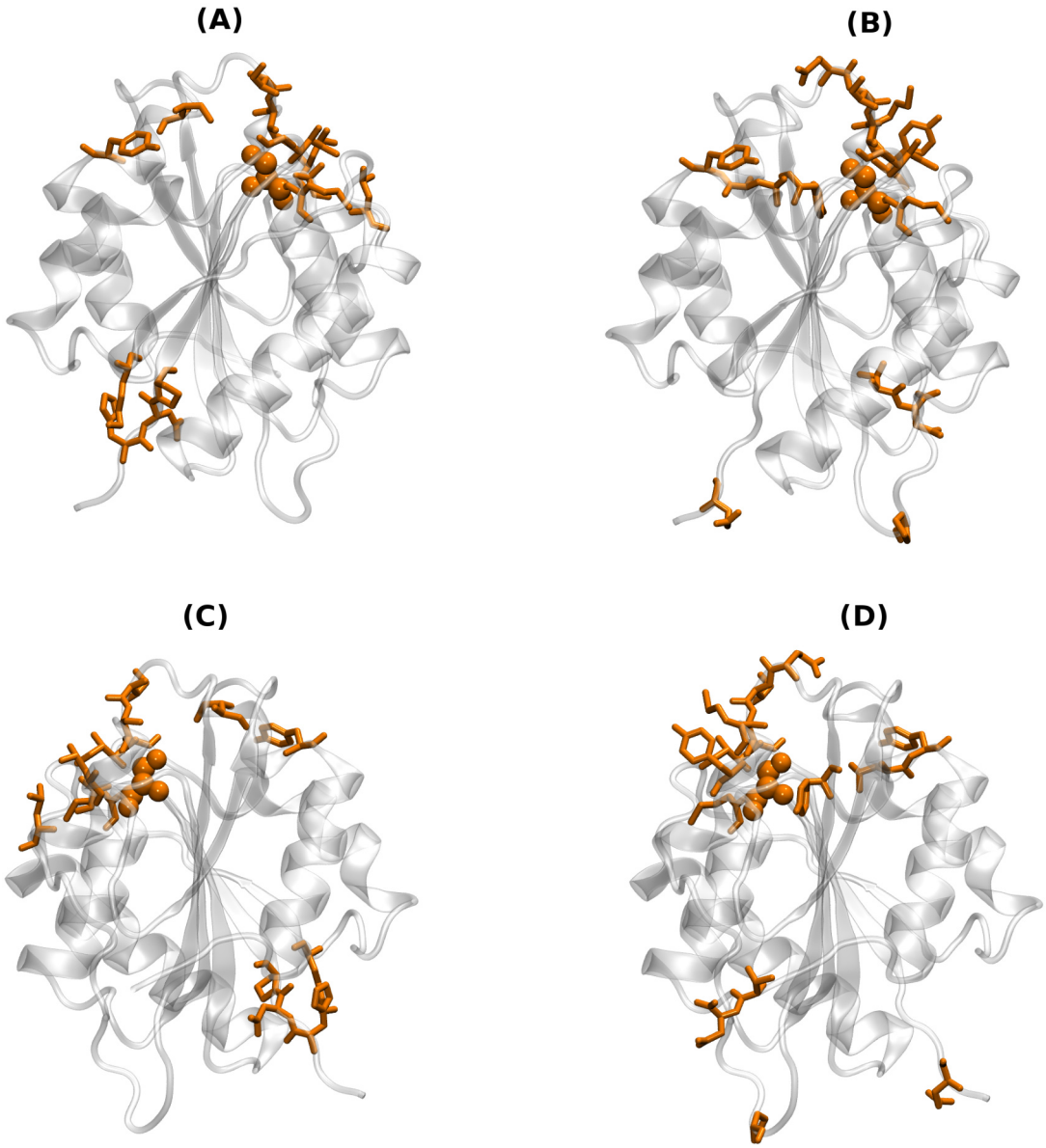

Figure S25: The same residues and enzyme as in Figure 7 of the manuscript shown from two different angles. A and B are for the aqueous and SFPL systems, respectively, shown from a fixed angle. C and D are for the aqueous and SFPL systems, respectively, shown from another fixed angle.

References

(1) Martinez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009**, *30*, 2157–2164.

(2) Behera, S.; Das, S.; Balasubramanian, S. An atomistic view of solvent-free protein liquids: the case of Lipase A. *Phys. Chem. Chem. Phys.* **2021**, *23*, 7302–7312.