Electronic Supplementary Material (ESI) for Journal of Materials Chemistry B.

Thermo- and Ion-responsive Silk-elastin-like Proteins and Their Multiscale Mechanisms

Haoyuan Shi[1], Ting Ji[2], Chenxi Zhai[1], Junting Lu[2], Wenwen Huang[2,3,4], Jingjie Yeo[1]*

[1]J[2] Lab for Engineering Living Materials, Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, 14853, United States.

[2]The Zhejiang University–University of Edinburgh Institute, Zhejiang University School of Medicine, Zhejiang University, Hangzhou 310058, China.

[3]Department of Orthopedics of the Second Affiliated Hospital, Zhejiang University School of Medicine, Zhejiang University, Hangzhou 310058, China.

[4]Dr. Li Dak Sum & Yip Yio Chin Center for Stem Cells and Regenerative Medicine, Zhejiang University School of Medicine, Zhejiang University, Hangzhou, China.

*Corresponding authors: wenwenhuang@intl.zju.edu.cn, jingjieyeo@cornell.edu.

CHARMM Potential Function and Generalized Born Implicit Solvent (GBIS)

The CHARMM potential function[1] in the fully atomistic molecular simulation is given by:

$$U_{CHARMM} = \sum_{bonds} K_b\left(b_{ij} - b_0\right)^2 + \sum_{angles} K_\theta\left(\theta_{ijk} - \theta_0\right)^2 + \sum_{dihedrals} K_\varphi\left[1 + \cos\left(n\varphi_{ijkl} - \delta\right)\right]^2 + \sum_{improper} K_\phi\left(\phi_{ijkl} - \phi_0\right)^2 + \sum_{Urey-Bradley} K_{UB}(U_{ik} - U_0)^2 + \sum_{nonbonded} \left\{\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi D r_{ij}}\right\}$$

where K_b, K_θ, K_φ, K_ϕ, and K_{UB} are the bond, angle, dihedral angle, improper angle and Urey–Bradley force constants, respectively; b_{ij}, θ_{ijk}, φ_{ijkl}, ϕ_{ijkl}, and U_{ik} are the bond length, bond angle, dihedral angle, improper torsion angle, and Urey–Bradley 1,3-distance respectively; b_0, θ_0, φ_0, ϕ_0, and U_0 are the equilibrium terms for such variables; n is the periodicity and δ the phase of a torsion; ε_{ij} is the well depth of the Lennard-Jones potential; σ_{ij} is the distance at the LJ minimum; q is the partial atomic charge; D is the effective dielectric constant; and r_{ij} is the distance between any atoms i and j.

Generalized Born implicit solvent (GBIS)[2] is used to significantly decrease the computational costs with the approximate method for calculating molecular electrostatics in solvent as described by the Poisson Boltzmann equation (PBE) that models water as a dielectric continuum.[3] The Generalized Born equation is an approximation of the PBE, and the total solvation free energy is given by:[4]

$$\Delta G_{solv}^{GB} = \sum_{i} \Delta G_{ii}^{GB} + \sum_{i} \sum_{i>j} \Delta G_{ij}^{GB}$$

Where ΔG_{ii}^{GB} is the Born radius dependent self-energy of atom i, and ΔG_{ij}^{GB} is the GB energy between atoms i and j.

Temperature Intervals with Global Exchange of Replicas (TIGER2)

TIGER2 in implicit solvent

Compared to traditional REMD, the TIGER2 method[5] significantly improves computing efficiency due to global exchange of replicas, thereby obtaining the global-minimum conformation more efficiently. The sampling cycle is decomposed into (I) heating, (II) sampling, and (III) quenching phases (Fig. S3a). Next, replicas will be compared, selected, and reassigned to higher temperature levels according to their potential energies; i.e., a higher potential energy state is assigned to a higher temperature level. The swap decisions are based on the probability:

$$P = \min\left[1, \quad \exp\left(\frac{E_A - E_B}{k_b \cdot T_{base}}\right)\right]$$

As all the replicas start and end under the baseline temperature, we can freely choose the number of replicas without considering the acceptance ratio, and the distribution of temperatures across the replicas exponentially increase from the lowest to the highest according to the function:

$$T_i = T_{min}\left(\frac{T_{max}}{T_{min}}\right)^{\frac{i-1}{n-1}}$$

where n is the number of replicas.

TIGER2 hybrid solvent with water shell (TIGER2hs)

Structural refinements in explicit solvent and ionic environments will be carried out using the TIGER2hs method. Compared to the implicit TIGER2 method, in TIGER2hs, simulating in the explicit solvent can predict protein conformations more precisely, and exchanging replicas based on potential energies in the implicit solvent with a layer of explicitly modeled water shell (Fig. S3b) avoid statistical noise generated by fully explicit solvent[6] while still accounting for more accurate solvation effects[7] than a purely implicit solvent. The number of water molecules in a shell is based on the proteins' degrees of freedom, such that $N_{shell} = \frac{1}{3} \cdot \frac{3N_{protein}-3}{6}$ where protein has $3N_{protein} - 3$ degrees of freedom and the water molecule in solution has 6 external degrees of freedom.

Martini 3.0 Coarse-grain Potential

Martini coarse-grained scheme uses an approximate 4:1 CG-AA mapping by combining top-down and bottom-up strategies.[8] Among them, Martini 3.0 (version 3.0.b.3.2)[9] is the newest version of Martini's coarse-grain potential, which updates the parameters and improves the accuracy. The potential energy function in the Martini system is described as:

$$U_{Martini} = U_{bonds} + U_{angles} + U_{dihedrals} + U_{constraints} + U_{LJ} + U_{Coulombic}$$

Where U_{bonds}, U_{angles}, and $U_{dihedrals}$ are harmonic bond, angle, and dihedral potentials, $U_{constraints}$ is the constraints in rigid rings and secondary structures, U_{LJ}, and $U_{Coulombic}$ are Lennard-Jones potentials and Coulombic potentials.

Relative Accessible Surface Area (RASA)

RASA of a protein residue is used to measure the residue solvent exposure. In this work, we calculated RASA for dityrosine crosslink sites in tyrosine, which is the ortho and meta carbons in the phenol group. The formula is:[10]

$$\text{RASA} = \frac{\text{SASA}}{\text{MaxSASA}}$$

Where SASA is the solvent-accessible surface area, and MaxSASA is the maximum possible solvent-accessible surface area for the site. The MaxSASA was obtained from Gly-Tyr-Gly

tripeptides with backbone angles of $\phi = -120°$ and $\psi = 140°$, same as Miller's work.[11] The MaxSASA of tyrosine is 228.769 Å^2 based on VMD[12] TCL scripts, which is comparable to the 229.0 Å^2 in Miller's work.[11] Then, we calculate the MaxSASA for four dityrosine crosslink sites in tyrosine, as shown in Table 1. Here, we defined an exposed site that is more than 20 % RASA of the average MaxSASA.[13] The SASA and the number of exposed dityrosine crosslink sites of representative SELP and Azo-SELP structures were shown in Table S2 and Table S3, respectively. However, in Table 1, we calculated the average values based on the most populated cluster obtained by the TIGER2 REMD sampling methods. Here, we did not show the site with a 0 value of SASA.

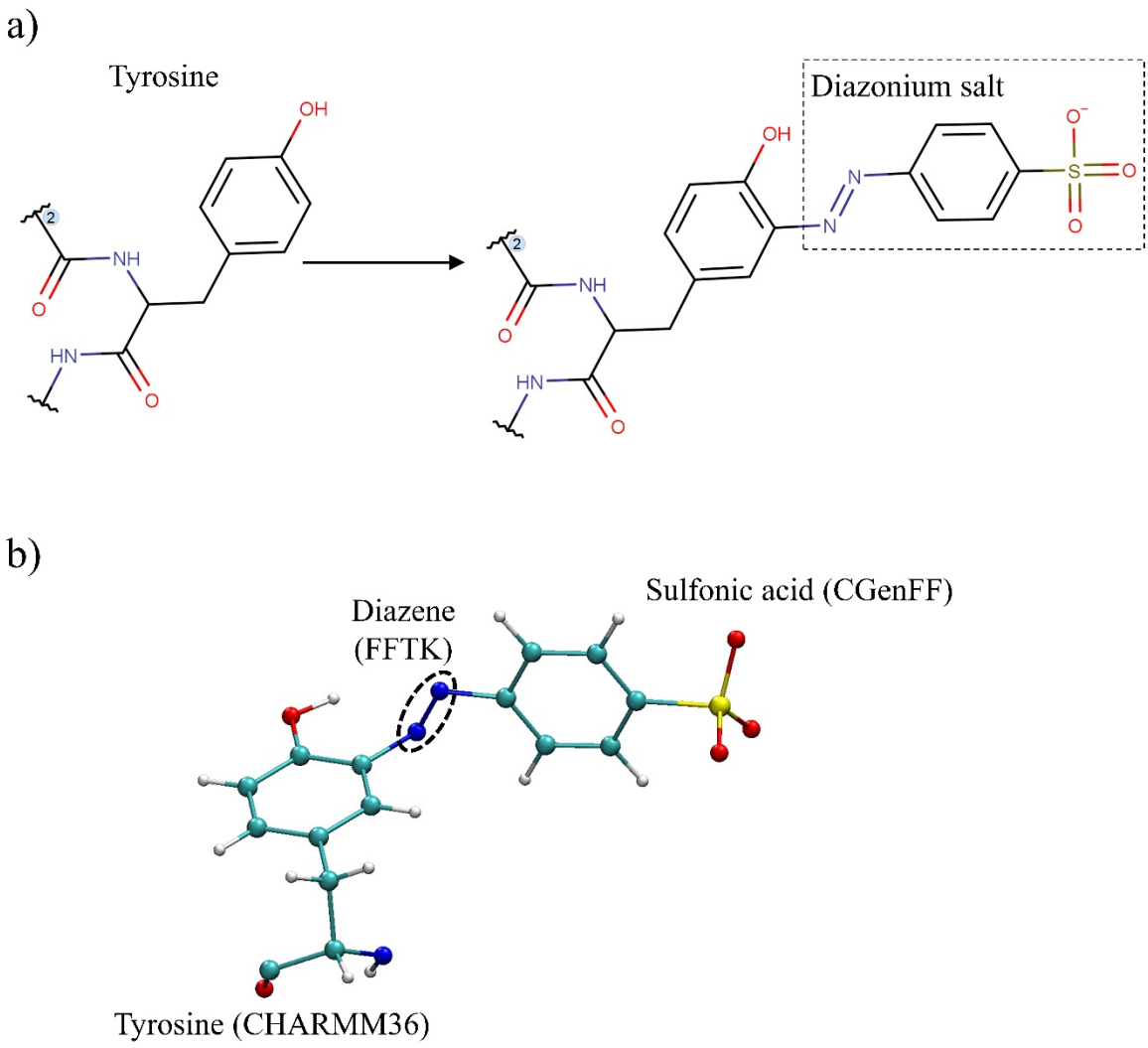


Fig. S1. a) Aryl diazonium salt with sulfonic acid used to modify tyrosine in SELP. b) The forcefield used in modified-tyrosine.

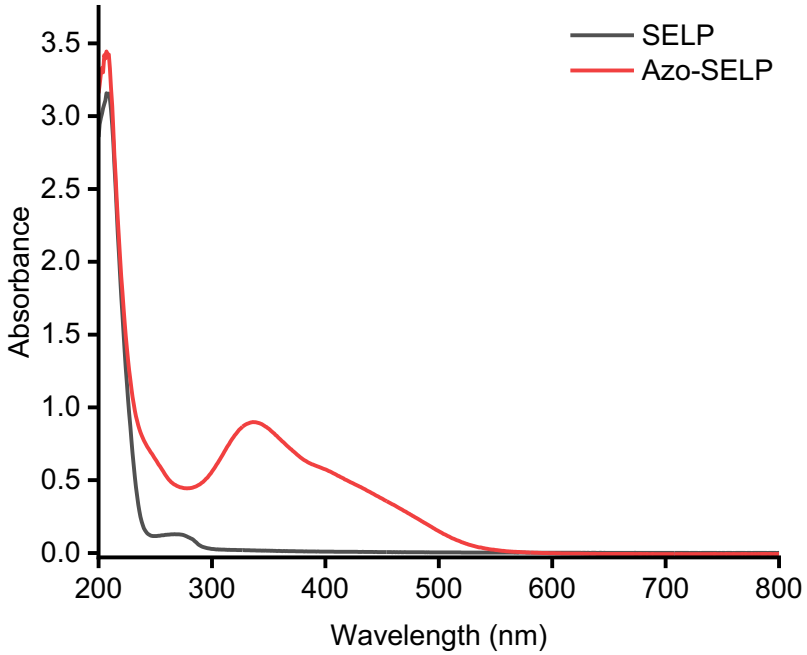
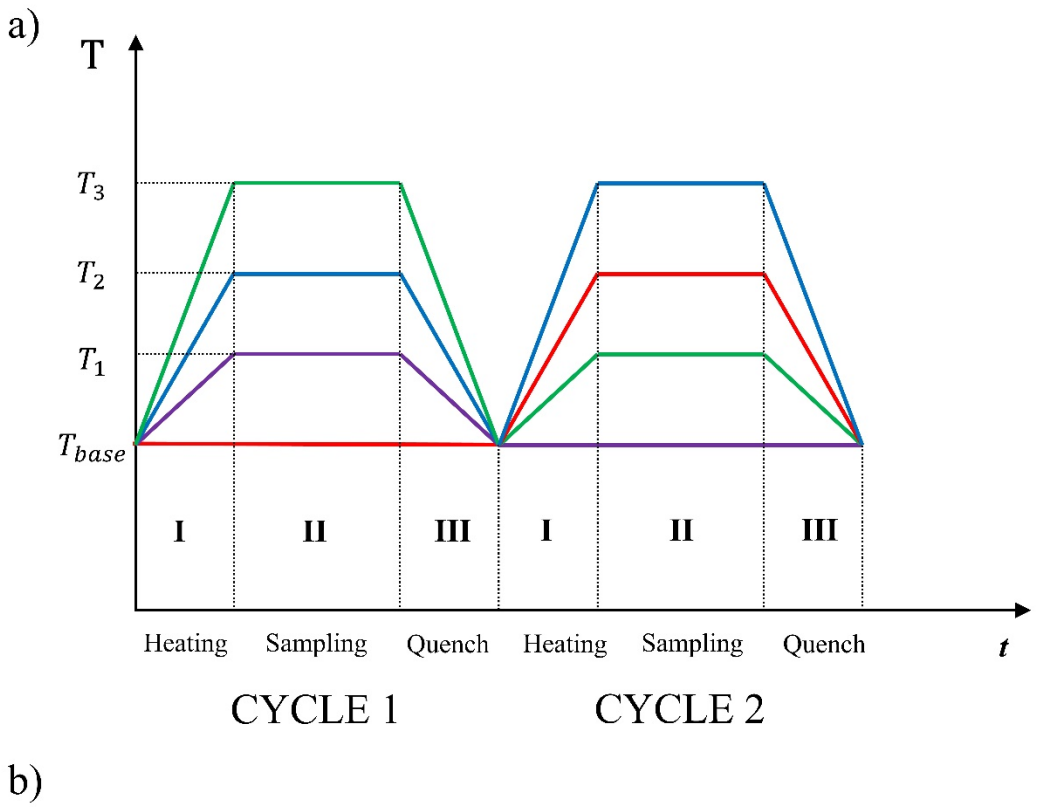


Fig. S2. UV-Vis spectra of SELP (black) and Azo-SELP (red), showing the diazonium modification in Azo-SELP.



100

101 Fig. S3. a) The scheme of the TIGER2 method, each sampling cycle contains (I) heating, (II)
102 sampling, and (III) quenching phases. b) SELPs with water shell.

103

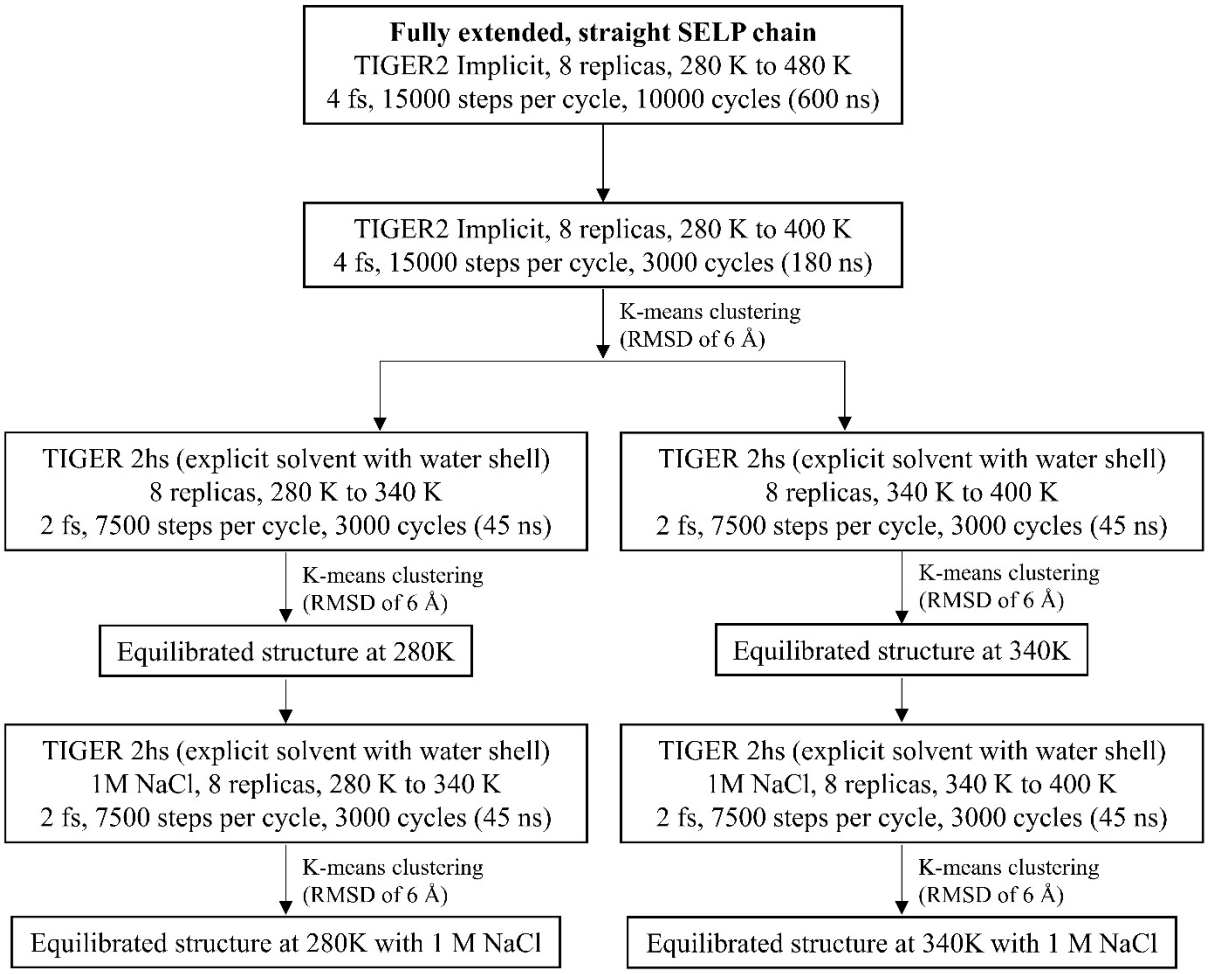


Fig. S4. The scheme of the FAMD simulations with the TIGER2 method for folding SELP at 280 K and 340 K.

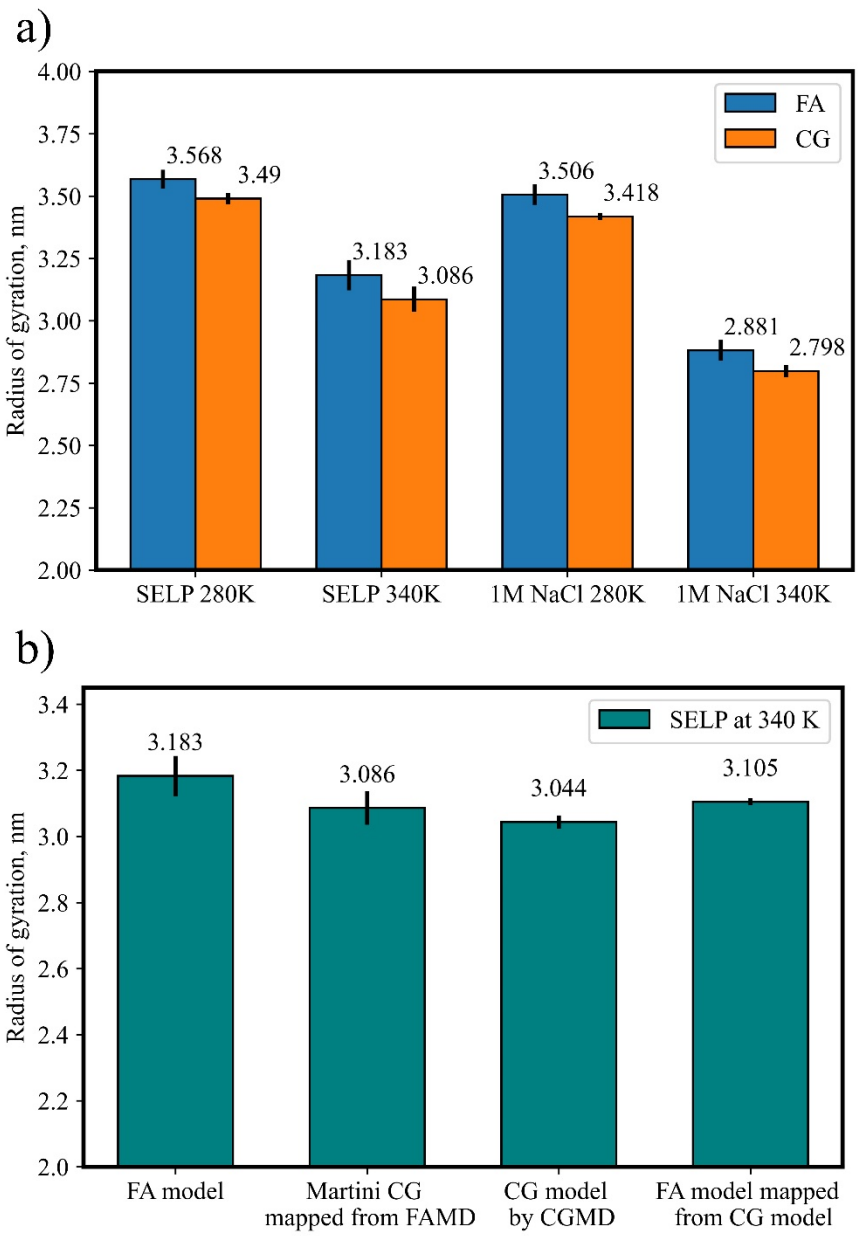


Fig. S5. a) Martini CG mapping from the representative FA SELP model. The CG model is consistent with the FA model regarding the radius of gyration after 20 ns CGMD simulations. b) Four SELP models at 340 K, including FA model, CG model mapped from FA model after 20 ns CG simulation, 500 ns CGMD simulation K as in the Method section from CG SELP model at 280, and FA model mapped from CG model obtained in CGMD simulation. The radius of gyration is consistent for all four models, signifying the reasonability for using CGMD to obtain the structure at different temperatures.

Table S1 The MaxSASA and SASA (RASA = 20 %) for four dityrosine crosslink sites of Tyr in Gly-Tyr-Gly.

Carbon name	MaxSASA/ Å^2	SASA (RASA = 20 %)/ Å^2
CD1	27.292	5.458
CD2	15.458	3.092
CE1	26.085	5.217
CE2	32.364	6.473
Average	25.300	5.060

Table S2 The SASA and the number of exposed dityrosine crosslink sites of the representative SELP structure.

Carbon name	Atom Id	SASA/ Å^2	Exposed site (> 5.06 Å^2)
CD1	282	0.63409913	0
CD2	289	0.42273274	0
CD1	895	8.03192234	1
CE1	897	5.91825819	1
CD2	902	6.76372385	1
CE2	904	6.12962484	1
CE1	1510	3.38186193	0
CD1	2734	4.22732735	0
CE1	2736	7.18645668	1
CD2	2741	8.877388	1
CE2	2743	5.91825819	1
CD1	3347	3.80459476	0
CE1	3349	9.0887537	1
CD2	3354	4.86142635	0
CE2	3356	9.93421936	1
CD1	3960	2.32503009	0
CE1	3962	5.28415918	1
CD2	3967	6.34099102	1
CE2	3969	10.145586	1
CD1	4573	1.69093096	0
CD1	5186	4.65006018	0
CE1	5188	9.30012035	1
CD2	5193	3.80459476	0
CE2	5195	8.24328804	1
CD1	5799	2.9591291	0
CE1	5801	5.28415918	1
CD2	5806	0.21136637	0
CE2	5808	0.42273274	0
Total		145.842795	15

Table S2 The SASA and the number of exposed dityrosine crosslink sites of representative Azo-SELP structure.

Carbon name	Atom Id	SASA/ Å^2	Exposed site (> 5.06 Å^2)
CE1	284	0.21136637	0
CD1	910	0.84546548	0
CE1	912	7.18645668	1
CD1	1538	11.8365164	1
CE1	1540	12.047883	1
CD2	1542	5.70689201	1
CD1	2166	11.413784	1
CE1	2168	9.51148701	1
CD2	2170	0.21136637	0
CD1	2794	7.60918951	1
CE1	2796	4.22732735	0
CD2	2798	4.01596117	0
CD1	3422	8.877388	1
CE1	3424	11.8365164	1
CD1	4050	5.28415918	1
CE1	4052	8.45465469	1
CD1	4678	0.21136637	0
CD2	5938	4.65006018	0
Total		114.13784	11

1 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, 1983, **4**, 187–217.

2 W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *Journal of the American Chemical Society*, 1990, **112**, 6127–6129.

3 A. V Onufriev and D. A. Case, *Annual review of biophysics*, 2019, **48**, 275–296.

4 J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V Kalé, K. Schulten, C. Chipot and E. Tajkhorshid, *The Journal of chemical physics*, 2020, **153**, 44130.

5 X. Li, R. A. Latour and S. J. Stuart, *Journal of Chemical Physics*, 2009, **130**, 1–9.

6 X. Li, J. A. Snyder, S. J. Stuart and R. A. Latour, *The Journal of chemical physics*, 2015, **143**, 144105.

7 N. Geist, M. Kulke, L. Schulig, A. Link and W. Langel, *Journal of Physical Chemistry B*, 2019, **123**, 5995–6006.

8 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *Journal of Physical Chemistry B*, 2007, **111**, 7812–7824.

9 P. C. T. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B. M. H. Bruininks, T. A. Wassenaar, P. C. Kroon, J. Melcr, V. Nieto, V. Corradi, H. M. Khan, J. Domański, M. Javanainen, H. Martinez-Seara, N. Reuter, R. B. Best, I. Vattulainen, L. Monticelli, X. Periole, D. P. Tieleman, A. H. de Vries and S. J. Marrink, *Nature Methods*, 2021, **18**, 382–388.

10 M. Z. Tien, A. G. Meyer, D. K. Sydykova, S. J. Spielman and C. O. Wilke, *PloS one*, 2013, **8**, e80635–e80635.

11 S. Miller, J. Janin, A. M. Lesk and C. Chothia, *Journal of Molecular Biology*, 1987, **196**, 641–656.

12 W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33–38.

13 C. Savojardo, M. Manfredi, P. L. Martelli and R. Casadio, *Frontiers in Molecular Biosciences*, 2021, **7**, 460.