

---

# Supporting Information

## Zero-Shot Generation of Protein Conformational Ensembles Through AlphaFold Latent Flooding

---

 **Runtong Qian**<sup>1,2,3</sup>     **Rui Zhan**<sup>2,3</sup>     **Zilin Song**<sup>2,3,\*</sup>     **Jing Huang**<sup>2,3,\*,‡</sup>

<sup>1</sup>*School of Life Sciences, Fudan University, Shanghai 200433, China;* <sup>2</sup>*State Key Laboratory of Gene Expression, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310030, China;* <sup>3</sup>*Westlake AI Therapeutics Lab, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China.*

\***Correspondence:** songzilin@westlake.edu.cn (Z.S.) and huangjing@westlake.edu.cn (J.H.).

‡**Lead contact.**

### Contents

<b>1</b>	<b>Supporting Methods</b>	<b>2</b>
1.1	Molecular Dynamics Simulations . . . . .	2
1.2	Latent Dynamics Visualizations . . . . .	2
<b>2</b>	<b>Supporting Figures</b>	<b>3</b>
	Figure S1 . . . . .	3
	Figure S2 . . . . .	4
	Figure S3 . . . . .	5
	Figure S4 . . . . .	6
	Figure S5 . . . . .	7
	Figure S6 . . . . .	8
	Figure S7 . . . . .	9
	Figure S8 . . . . .	10
	Figure S9 . . . . .	11
	Figure S10 . . . . .	12
	Figure S11 . . . . .	13
	Figure S12 . . . . .	14
	Figure S13 . . . . .	15
	Figure S14 . . . . .	16
	Figure S15 . . . . .	17
	Figure S16 . . . . .	18
<b>3</b>	<b>Supporting Tables</b>	<b>19</b>
	Table S1 . . . . .	19
	Table S2 . . . . .	20
	Table S3 . . . . .	21
	Table S4 . . . . .	22

# 1 Supporting Methods

## 1.1 Molecular Dynamics Simulations

Molecular dynamics simulations was used as the orthogonal reference for the structural fluctuations of the ubiquitin protein. The native folded structure (PDB id: 1UBQ) was solvated in a 70 Å cubic box. The CHARMM36m and the CHARMM-modified TIP3P force fields were used to model the protein and the solvent, respectively. The short-range nonbinding interactions were explicitly treated within 10 Å and were smoothed out at 12 Å. The periodic boundary conditions were employed in all three dimensions and the particle mesh Ewald summation was applied for long range electrostatic and dispersion interactions. All covalent bonds to hydrogen atoms were constrained. The solvated system was then subjected sequentially to an initial energy minimization, a 4 ns canonical (NVT) equilibration, and a 4 ns isothermal-isobaric (NPT) equilibration. The production run was carried out for 500 ns with temperature maintained at 300 K using the Nosé-Hoover chain thermostat and pressure coupled to 1 bar using the Monte Carlo barostat. The dynamical trajectory was integrated at 2 fs time steps and was recorded every 100 ps.

## 1.2 Latent Dynamics Visualizations

The latent dynamics driven by AFLF can be quantified by the evolution of the massive activations (the top and bottom 1% tensor elements) in the latent states. We computed the set overlap of the massive activations by the Jaccard Index (JI). Formally, let  $\mathbf{B}_h(\tau)$  be the binary matrix of identical dimensions that flags the massive activations and  $\mathcal{B}_h(\tau)$  the subset of linearized indices of the nonzero entries in  $\mathbf{B}_h(\tau)$ , the JI is defined,

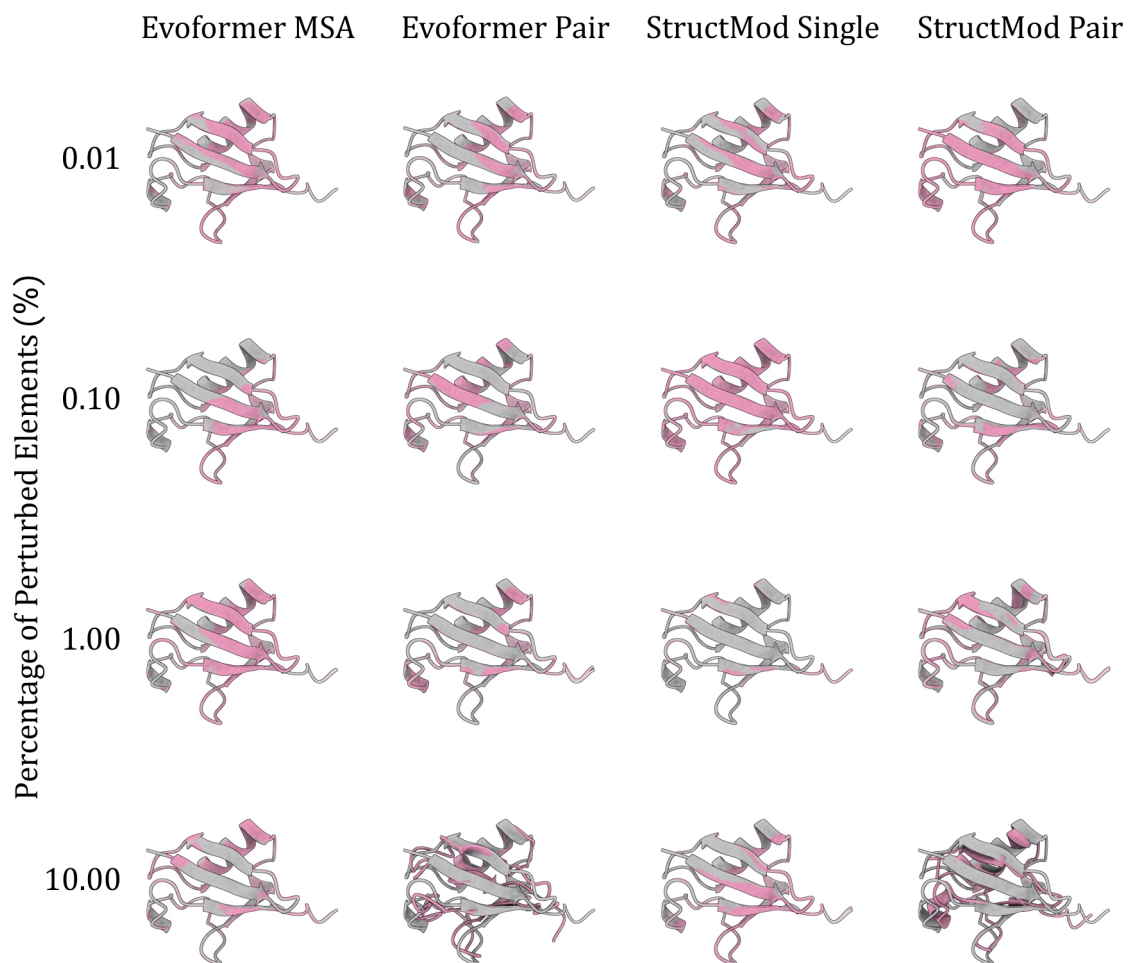
$$\text{JI}(\tau, \tau') := \frac{\#(\mathcal{B}_h(\tau) \cap \mathcal{B}_h(\tau'))}{\#(\mathcal{B}_h(\tau) \cup \mathcal{B}_h(\tau'))}, \quad (1)$$

such that it quantifies the overlap of massively activated entries between latent states at different times. The UMAP dimensionality reduction was performed by using the pairwise JIs as the distance metric. The optimal transport distance in one dimension quantifies the cost of mutating two balanced matrices into each other and is obtained by the monotone coupling,

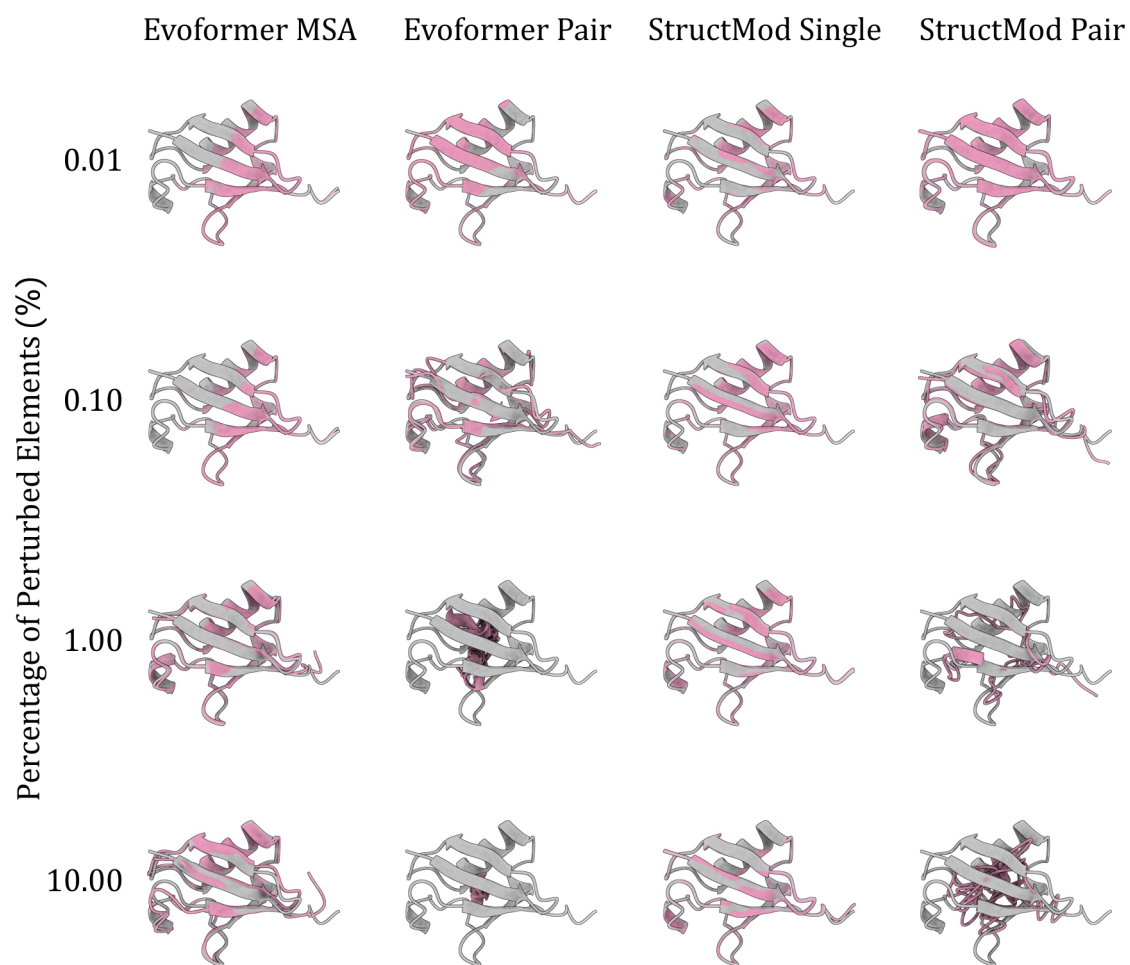
$$d_{OT}(\tau, \tau') := \frac{1}{\sum \mathbf{B}_h(\tau)} \sum \|\text{sort}(\mathbf{B}_h(\tau)\mathbf{h}(\tau)) - \text{sort}(\mathbf{B}_h(\tau')\mathbf{h}(\tau'))\|_1. \quad (2)$$

Empirically, we used JI to monitor the dynamical evolution in positions of the massive activations and  $d_{OT}$  in value distributions.

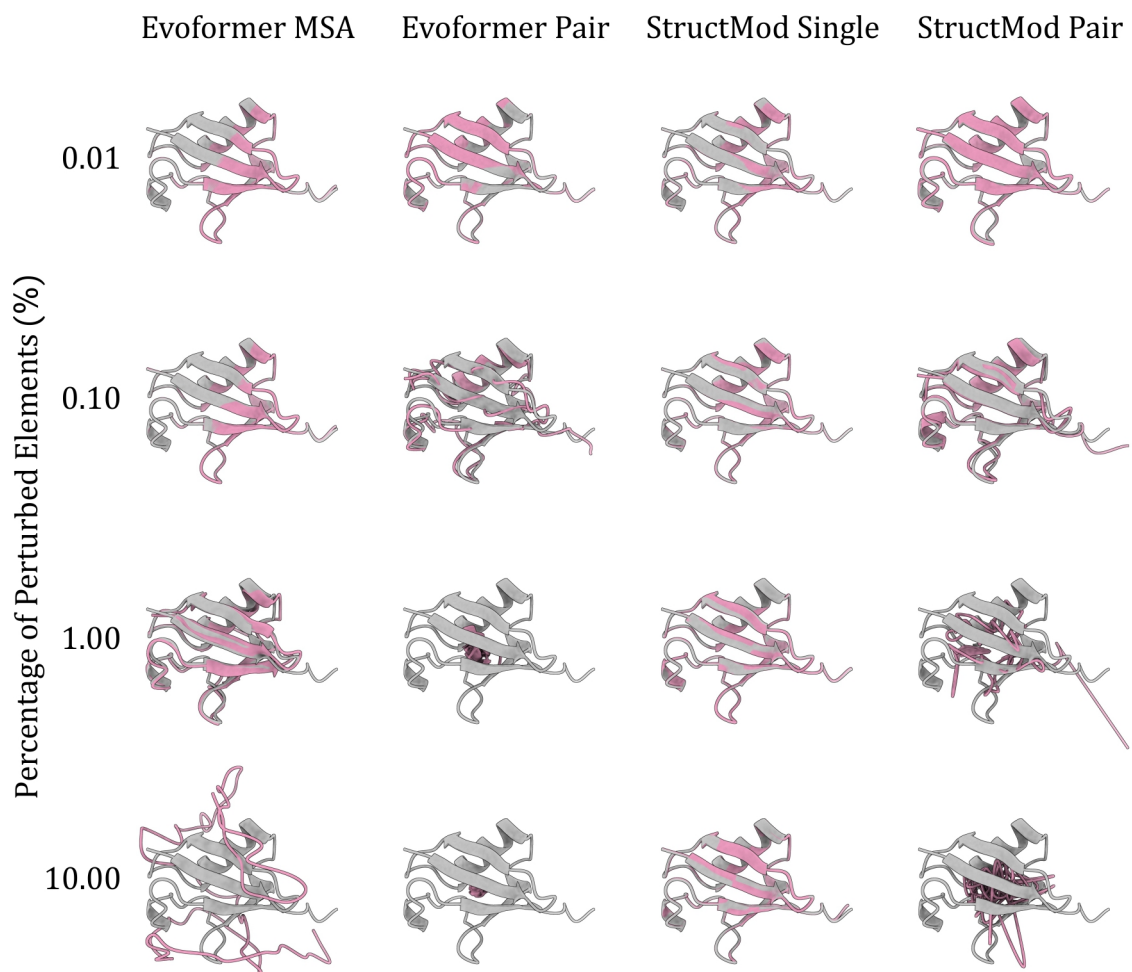
## 2 Supporting Figures



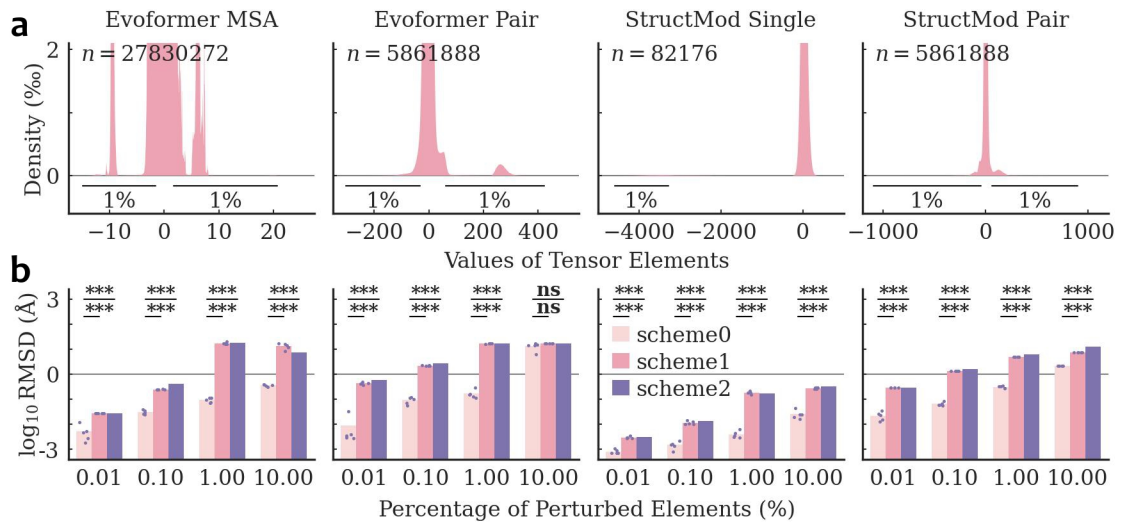
**Figure S1.** Ubiquitin structures predicted from scheme0 latent state perturbations. The structure predicted from perturbed latent states (pink) are superposed with the unperturbed predictions (gray). All structures were rendered at the same scale.



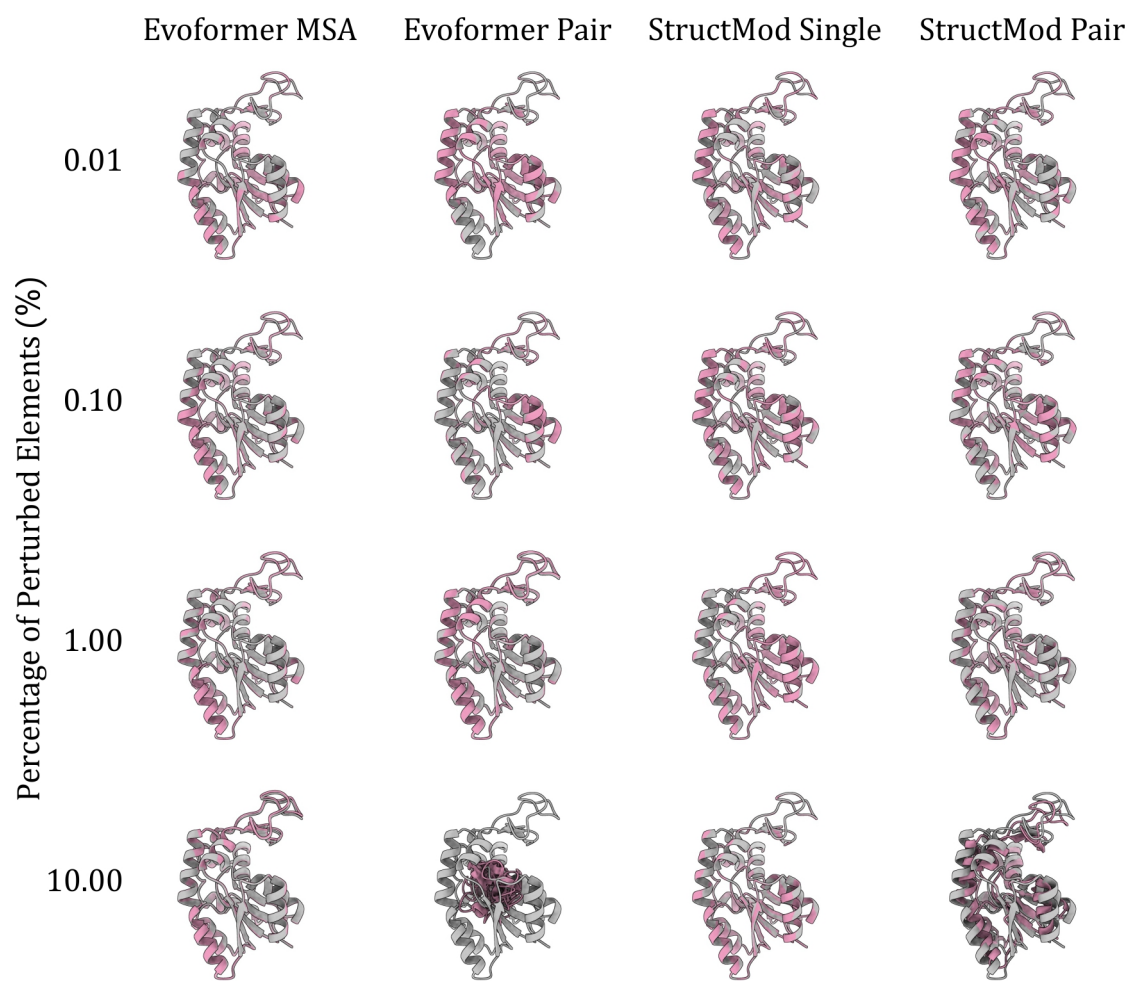
**Figure S2.** Ubiquitin structures predicted from scheme1 latent state perturbations. The structure predicted from perturbed latent states (pink) are superposed with the unperturbed predictions (gray). All structures were rendered at the same scale.



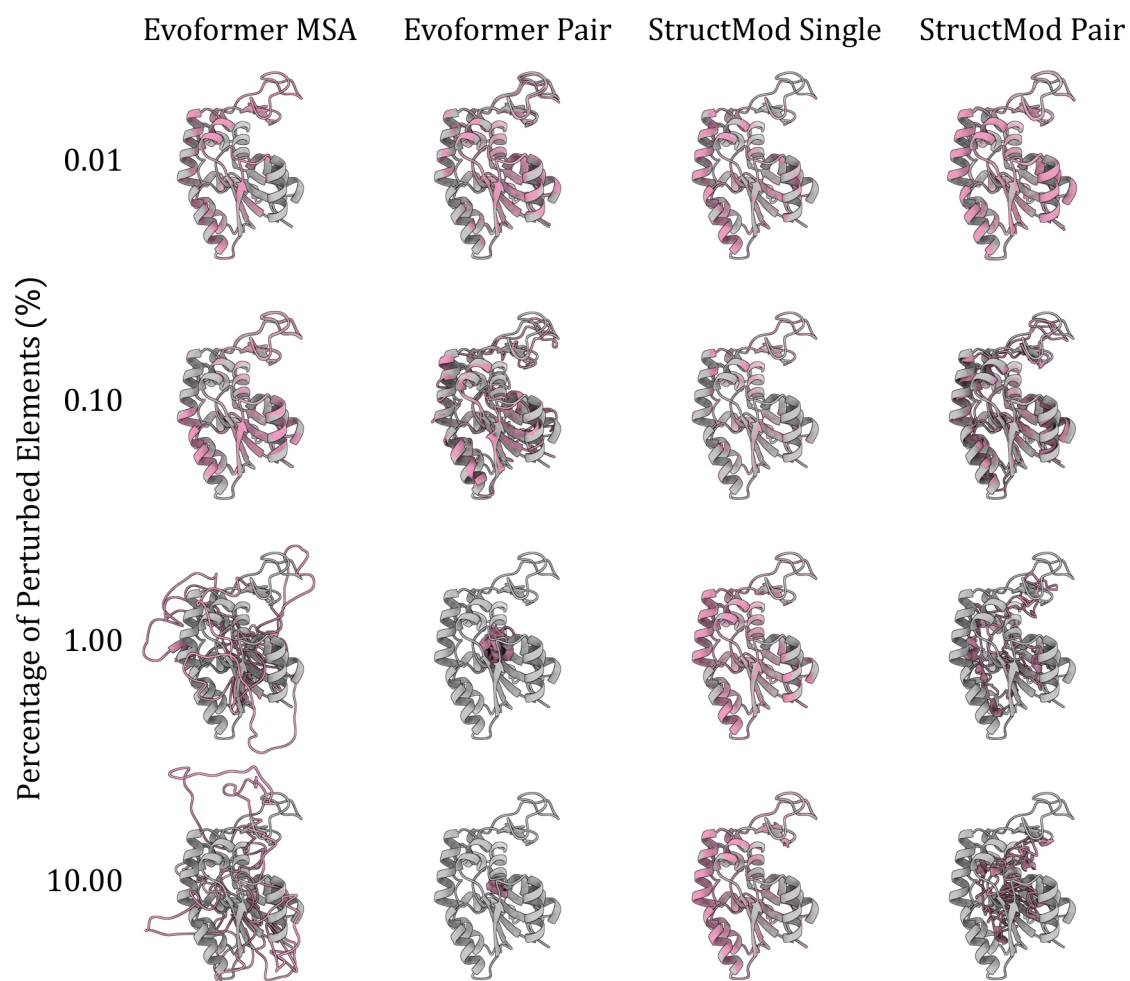
**Figure S3.** Ubiquitin structures predicted from scheme2 latent state perturbations. The structure predicted from perturbed latent states (pink) are superposed with the unperturbed predictions (gray). All structures were rendered at the same scale.



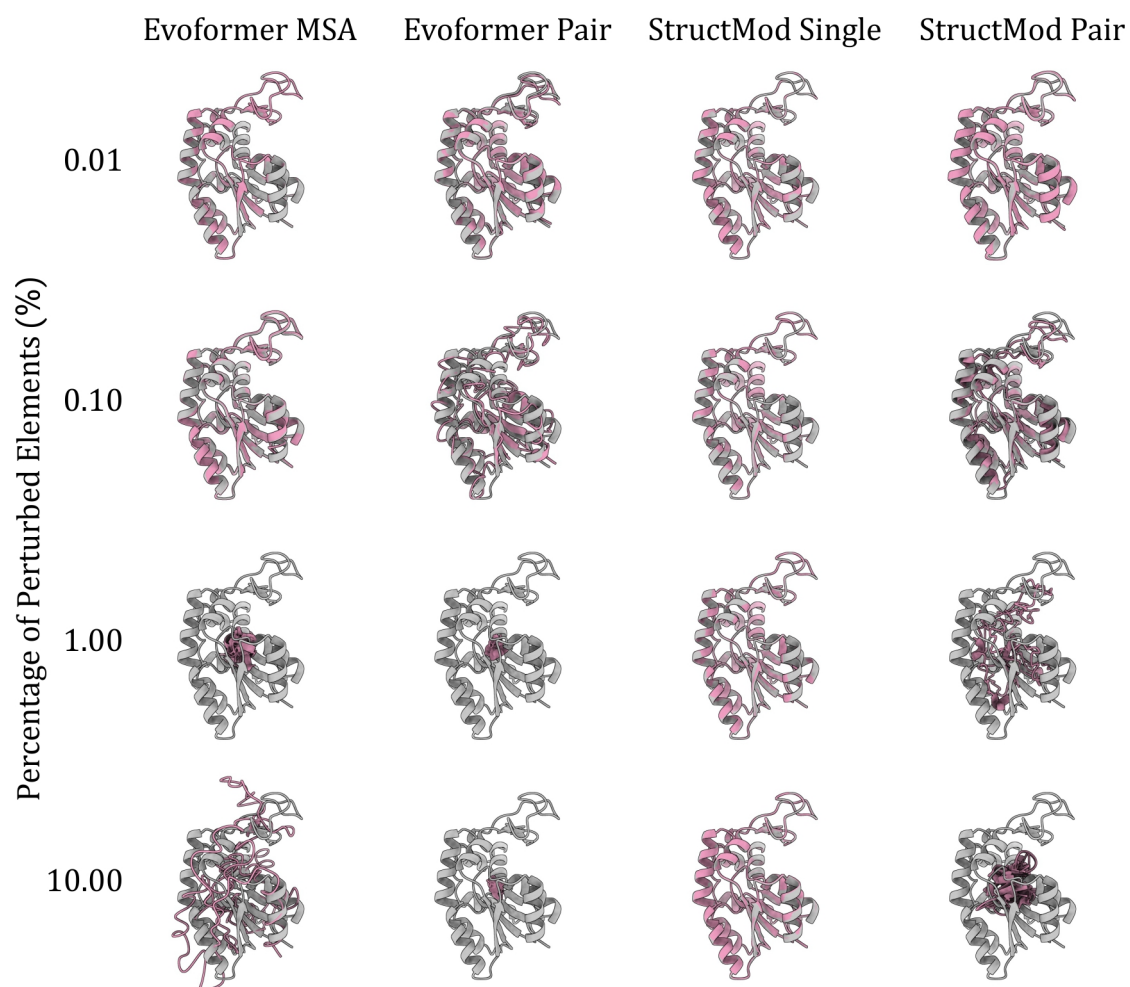
**Figure S4.** Latent features distribution and adenylate kinase (AdK) structural deviations from perturbed latent features. From left to right: Evoformer MSA, Evoformer pair, structural module (StructMod) single, and StructMod pair representations. **(a)** Permillage value distributions of the AF2 latent tensor elements used to predict AdK conformations. The 1st and 99th percentiles are indicated and  $n$  is the total number of elements in each tensor. **(b)** Backbone root mean squared deviations ( $\log_{10}$  RMSDs) of AdK conformations predicted from perturbed latent tensors relative to the native unperturbed prediction. Statistical tests: scheme0 ( $n = 5$ ) versus scheme1 ( $n = 5$ ) by independent samples  $t$ -test; scheme0 ( $n = 5$ ) versus scheme2 ( $n = 1$ ) by one-sample  $t$ -test. \*\*\*:  $p < 0.001$ ; \*\*:  $p < 0.01$ ; \*:  $p < 0.05$ ; ns: not significant.



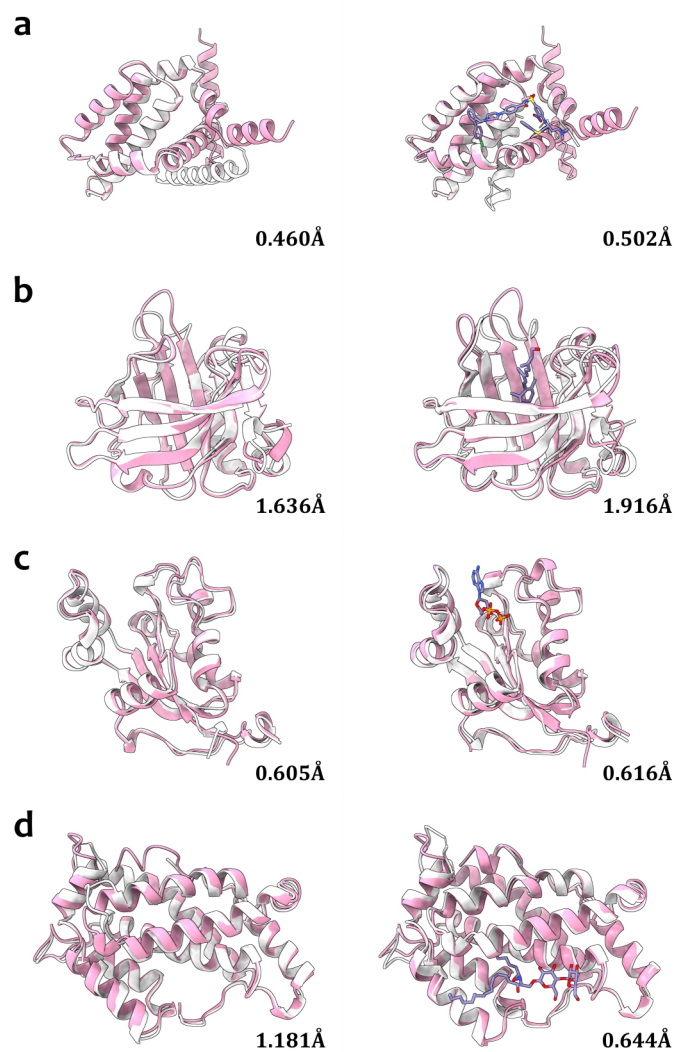
**Figure S5.** Adenylate kinase structures predicted from scheme0 latent state perturbations. The structure predicted from perturbed latent states (pink) are superposed with the unperturbed predictions (gray). All structures were rendered at the same scale.



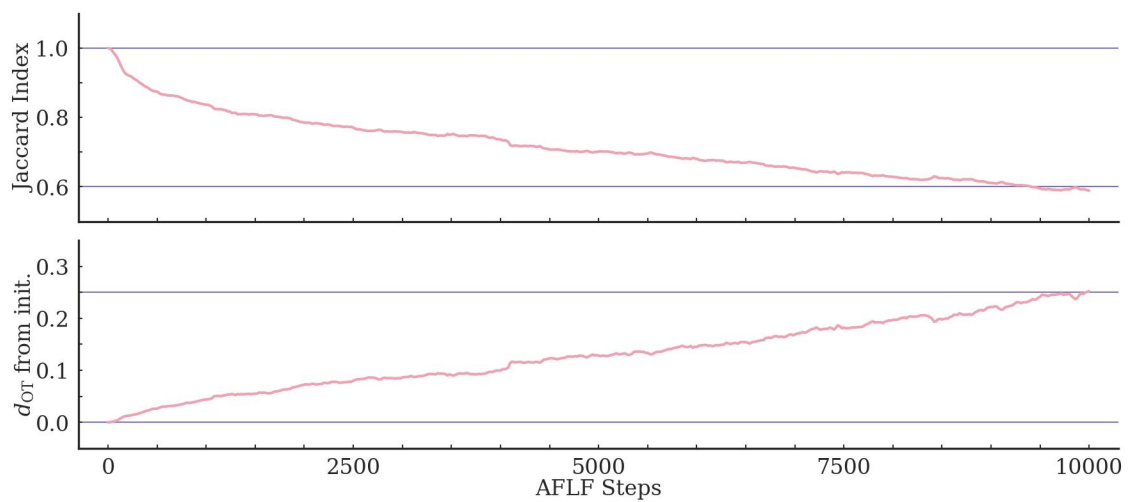
**Figure S6.** Adenylate kinase structures predicted from scheme1 latent state perturbations. The structure predicted from perturbed latent states (pink) are superposed with the unperturbed predictions (gray). All structures were rendered at the same scale.



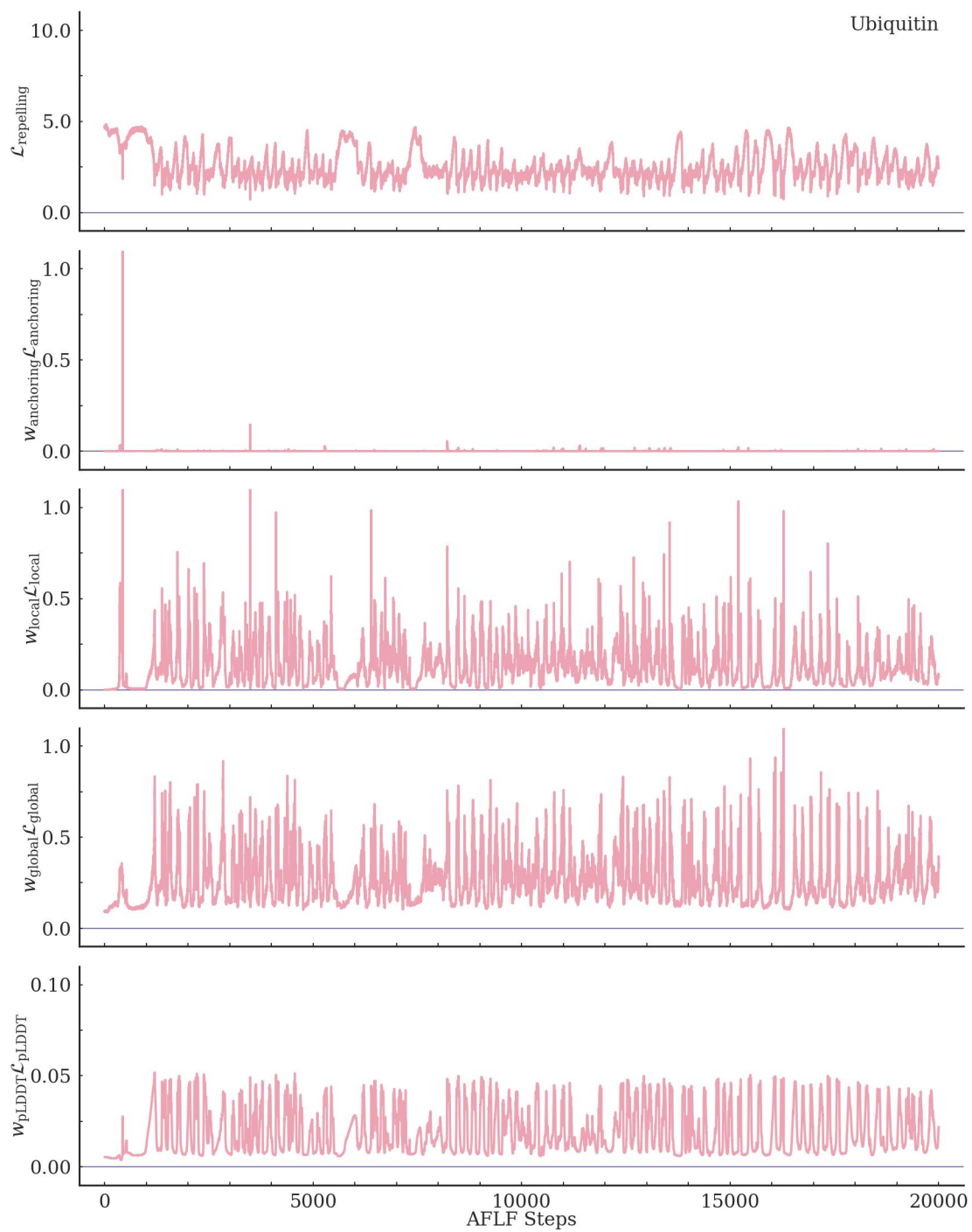
**Figure S7.** Adenylate kinase structures predicted from scheme2 latent state perturbations. The structure predicted from perturbed latent states (pink) are superposed with the unperturbed predictions (gray). All structures were rendered at the same scale.



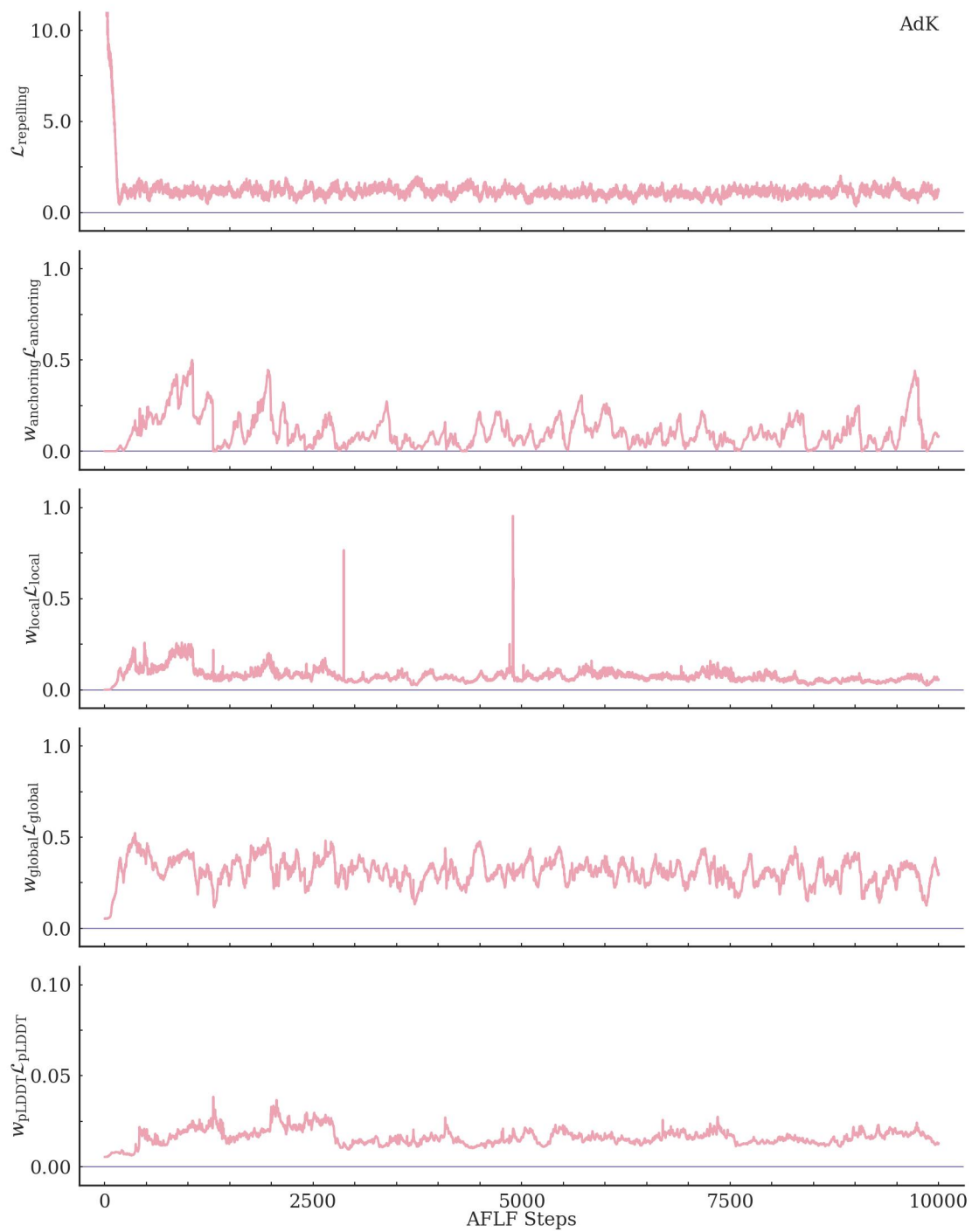
**Figure S8.** The representative AFLF generated structures for the cryptic binding sites. **(a)** The anti-apoptotic protein Bcl- $x_L$ ; **(b)** The human nucleoside diphosphate kinase A (NDPK-A); **(c)** The glycolipid transfer protein (GLTP). The minimum  $C\alpha$  RMSDs between the AFLF generated conformations (pink) and the experimental crystal structures (proteins in gray and ligands in purple) are listed.



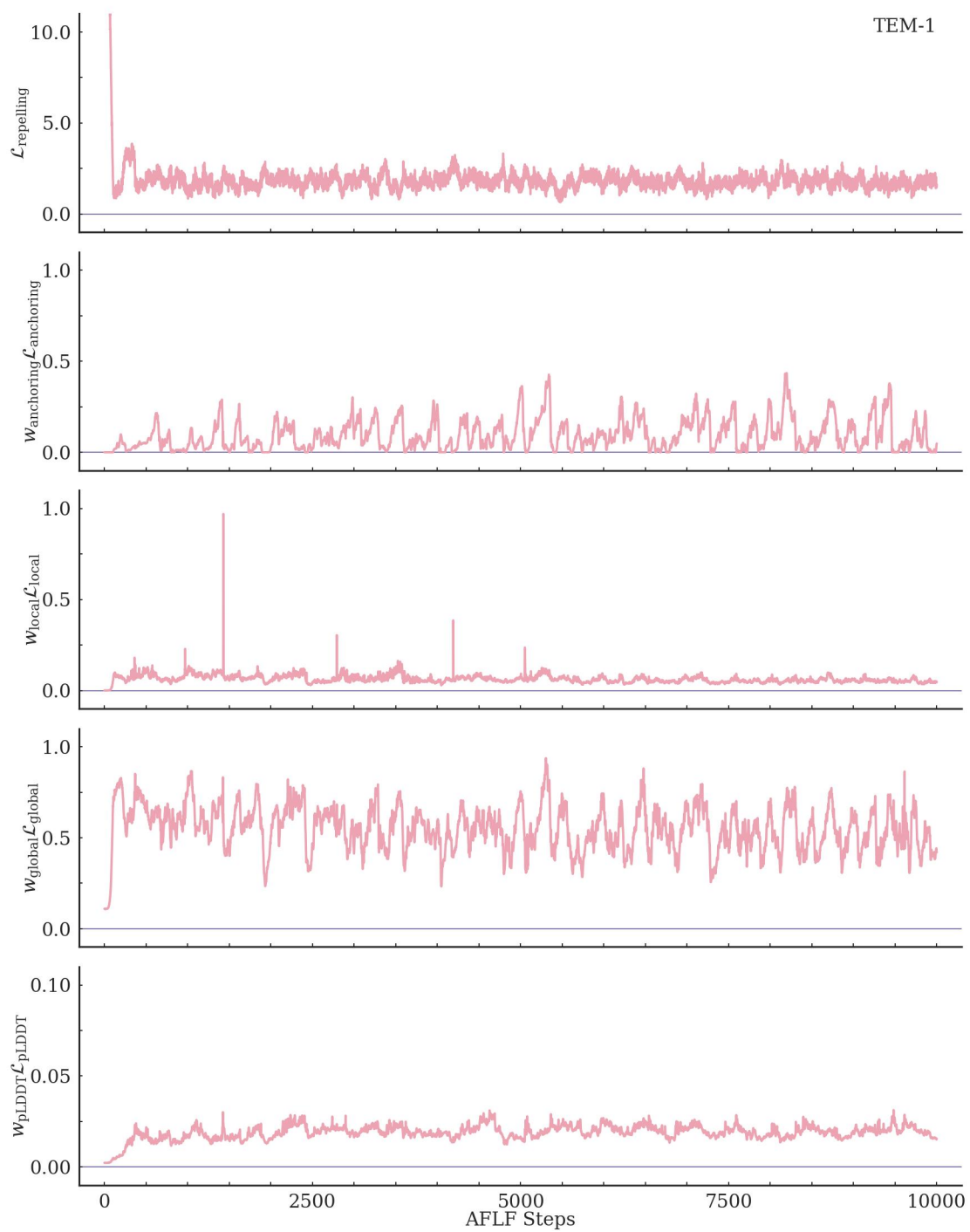
**Figure S9.** The evolution of the Jaccard indices (top) and the optimal transport distances (bottom) with regard to the initial state along the adenylate kinase (AdK) trajectory.



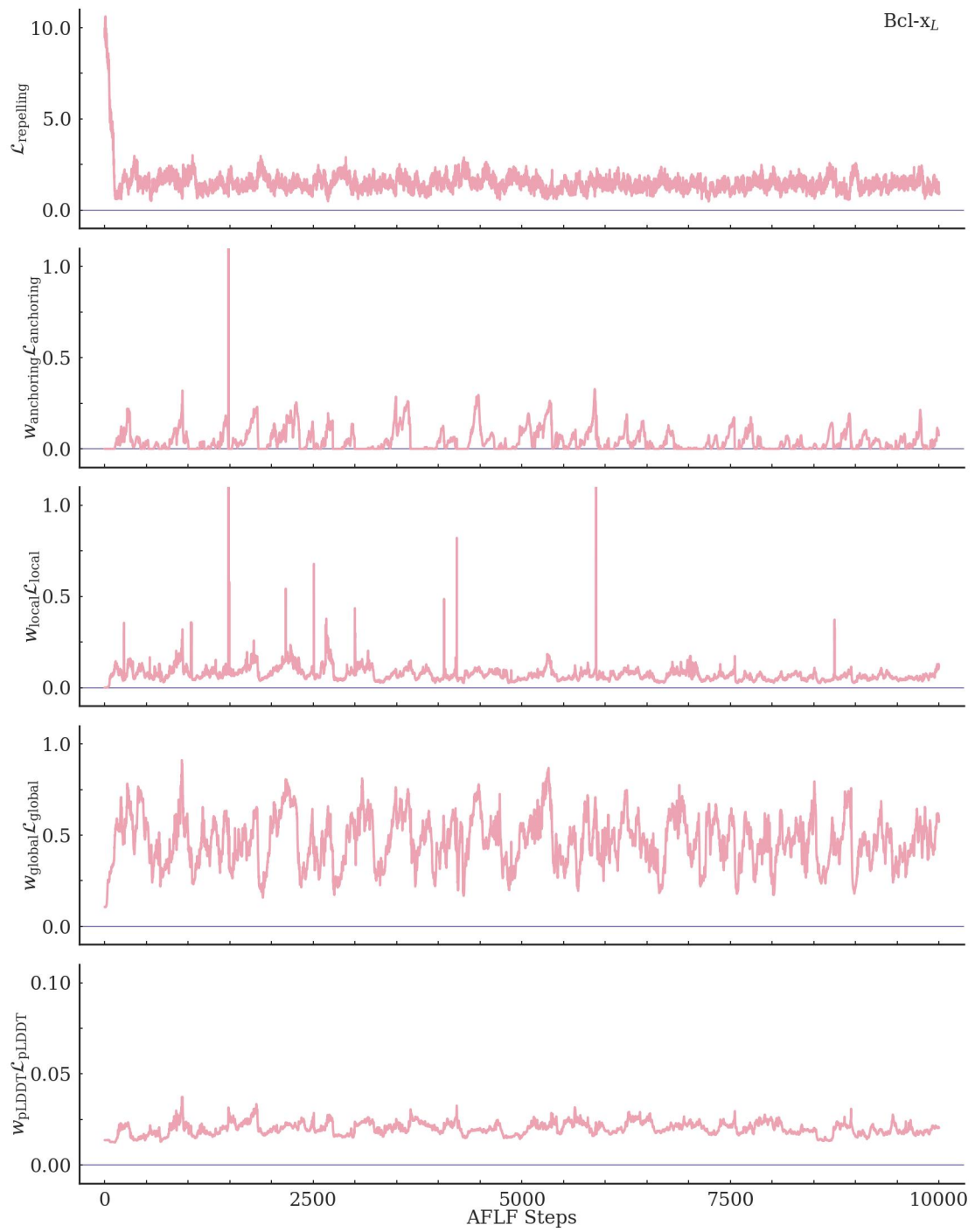
**Figure S10.** The evolution of the loss terms along the ubiquitin trajectory.



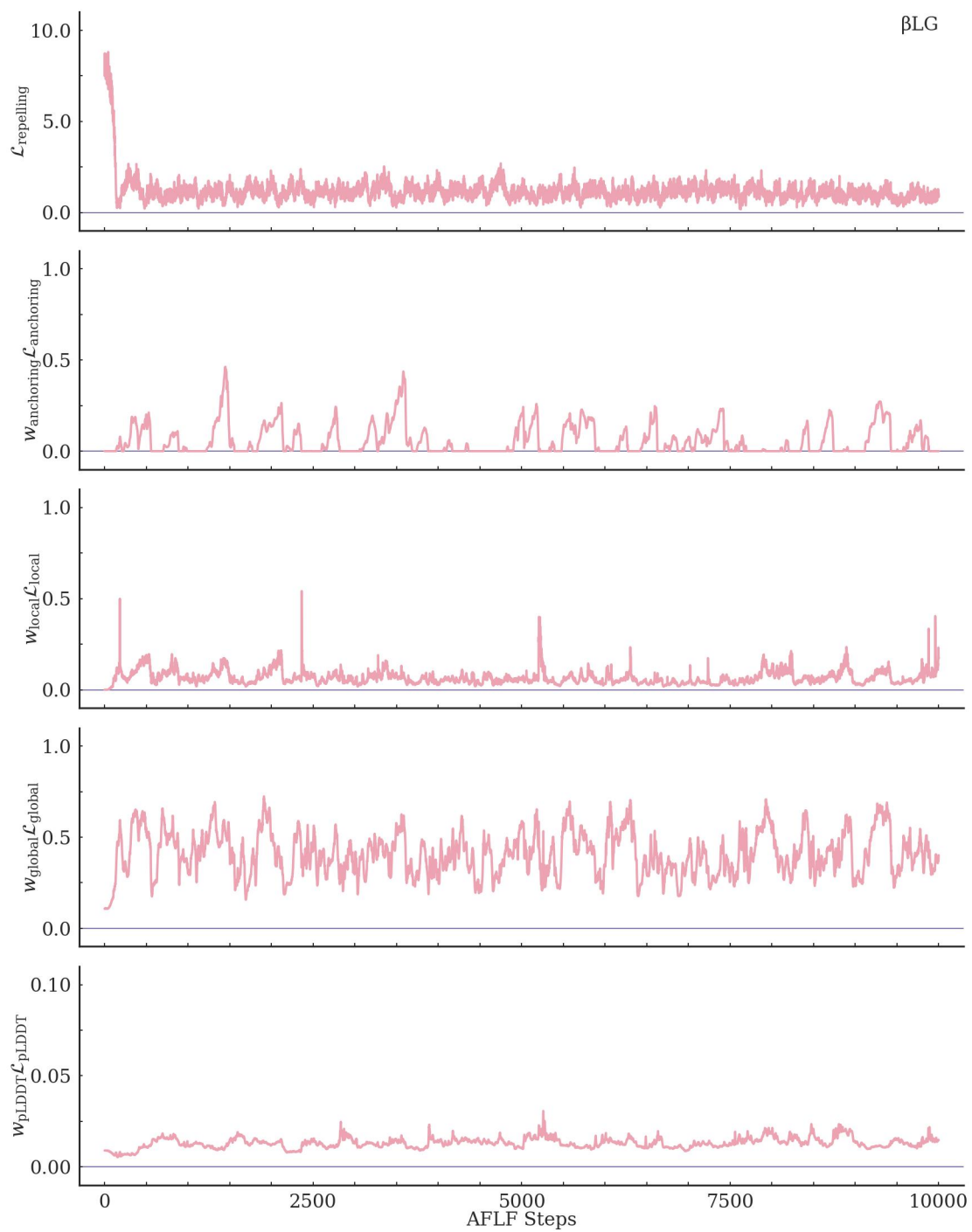
**Figure S11.** The evolution of the loss terms along the adenylyate kinase (AdK) trajectory.



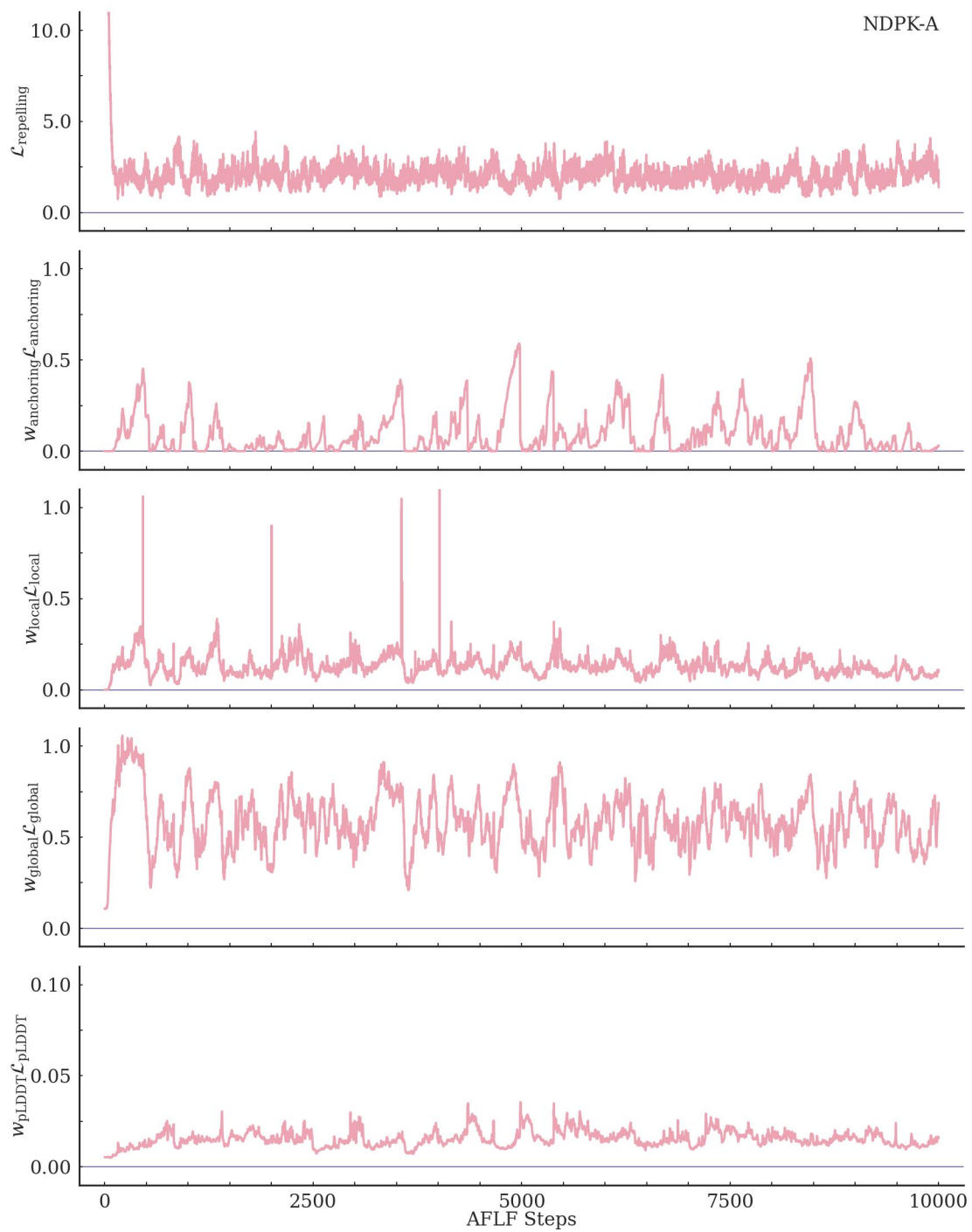
**Figure S12.** The evolution of the loss terms along the class A  $\beta$ -lactamase TEM-1 trajectory.



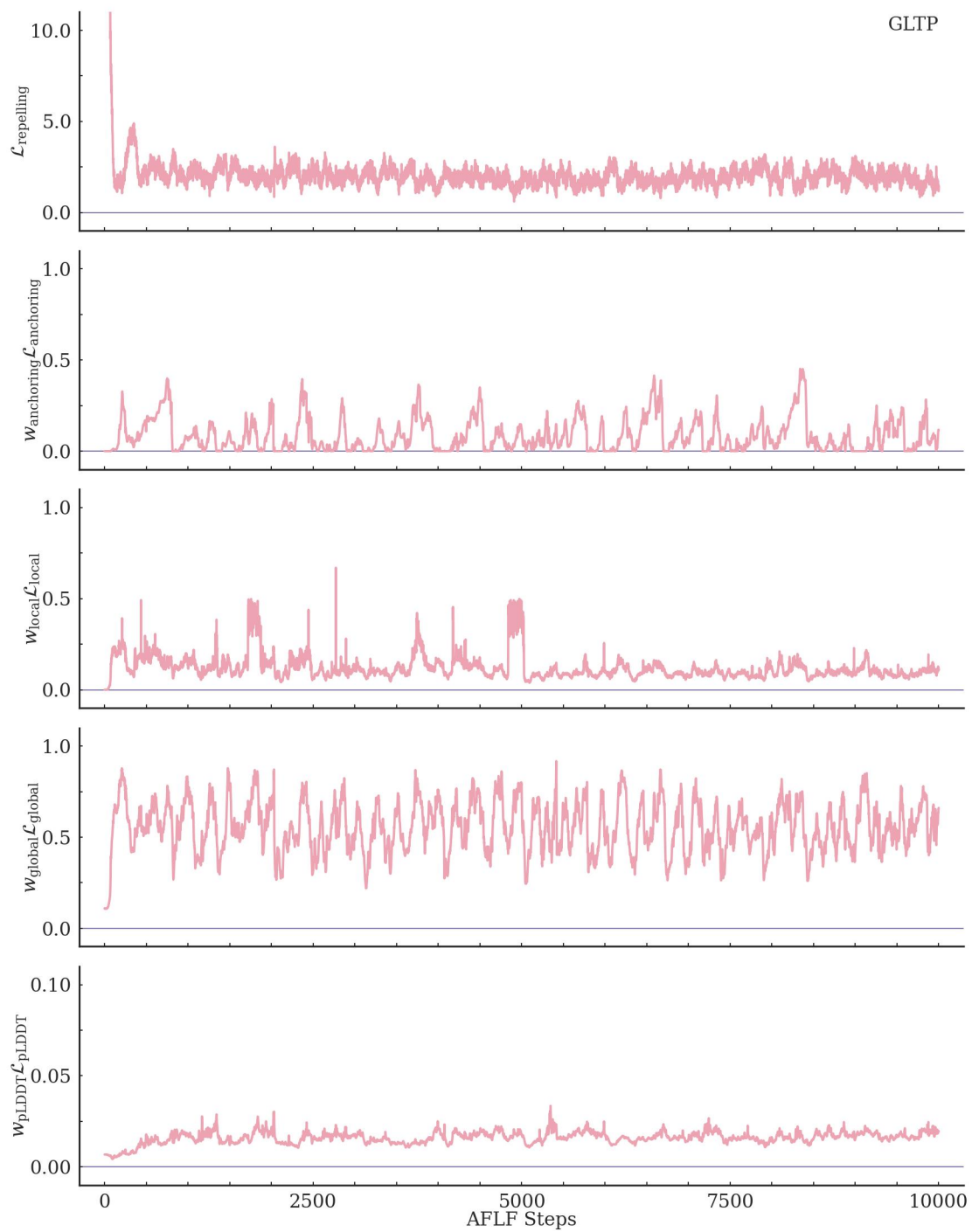
**Figure S13.** The evolution of the loss terms along the anti-apoptotic protein Bcl-x<sub>L</sub> trajectory.



**Figure S14.** The evolution of the loss terms along the bovine  $\beta$ -lactoglobulin ( $\beta$ LG) trajectory.



**Figure S15.** The evolution of the loss terms along the human nucleoside diphosphate kinase A (NDPK-A) trajectory.



**Figure S16.** The evolution of the loss terms along the human glycolipid transfer protein (GLTP) trajectory.

### 3 Supporting Tables

**Table S1.** Experimental and simulated residue-wise B-factors for ubiquitin (unit: Å<sup>2</sup>).

<b>Residue</b>	<b>Exp.</b>	<b>MD</b>	<b>AFLF</b>	<b>Residue</b>	<b>Exp.</b>	<b>MD</b>	<b>AFLF</b>
MET 1	10.38	15.32	17.17	ASP 39	14.96	26.34	9.83
GLN 2	9.07	11.58	4.85	GLN 40	10.76	21.56	7.19
ILE 3	5.07	6.66	3.92	GLN 41	3.87	14.06	4.10
PHE 4	4.68	5.00	4.35	ARG 42	6.97	14.19	5.06
VAL 5	3.87	5.69	3.29	LEU 43	3.51	8.21	2.54
LYS 6	6.12	8.47	4.18	ILE 44	5.55	9.14	4.20
THR 7	7.48	28.25	5.76	PHE 45	4.70	9.44	3.78
LEU 8	14.15	85.86	11.23	ALA 46	7.15	32.59	6.15
THR 9	19.24	84.06	14.41	GLY 47	11.68	34.02	11.86
GLY 10	18.74	52.76	12.40	LYS 48	8.82	27.60	14.38
LYS 11	11.91	23.08	12.07	GLN 49	7.18	18.31	9.68
THR 12	9.85	12.51	5.01	LEU 50	7.41	11.93	6.22
ILE 13	11.84	13.33	5.40	GLU 51	11.90	15.30	13.54
THR 14	9.63	11.06	4.55	ASP 52	16.56	17.80	11.96
LEU 15	9.03	10.59	3.63	GLY 53	11.77	19.62	11.47
GLU 16	11.50	12.44	3.87	ARG 54	9.05	12.87	11.13
VAL 17	8.85	8.60	3.95	THR 55	9.03	9.92	8.16
GLU 18	7.08	10.28	9.98	LEU 56	8.29	7.26	3.86
PRO 19	7.07	12.03	10.43	SER 57	9.00	13.63	7.11
SER 20	6.28	12.20	7.98	ASP 58	7.91	17.77	17.42
ASP 21	7.70	8.06	14.38	TYR 59	8.45	13.98	9.83
THR 22	6.01	8.86	7.88	ASN 60	13.94	14.59	11.50
ILE 23	9.92	7.75	5.91	ILE 61	11.78	9.39	5.69
GLU 24	11.81	13.49	7.39	GLN 62	15.52	14.12	10.66
ASN 25	10.96	9.66	7.53	LYS 63	11.97	11.59	13.90
VAL 26	5.53	6.92	4.69	GLU 64	10.94	9.77	12.72
LYS 27	4.14	9.86	5.84	SER 65	6.90	7.28	9.23
ALA 28	7.74	10.05	7.49	THR 66	3.80	6.22	4.39
LYS 29	7.90	9.63	7.53	LEU 67	3.85	5.54	2.78
ILE 30	5.58	10.96	4.71	HIS 68	4.17	7.37	2.79
GLN 31	8.67	13.75	6.24	LEU 69	3.97	9.25	4.06
ASP 32	14.01	20.86	8.09	VAL 70	6.26	13.96	3.61
LYS 33	14.00	24.20	7.61	LEU 71	16.06	24.05	4.34
GLU 34	10.07	24.75	8.71	ARG 72	25.83	46.43	10.43
GLY 35	6.29	22.87	14.10	LEU 73	30.76	249.00	24.45
ILE 36	6.07	22.02	8.99	ARG 74	35.33	267.10	39.81
PRO 37	9.18	26.91	10.99	GLY 75	36.07	574.13	80.29
PRO 38	9.08	21.09	6.39	GLY 76	36.19	1212.43	323.16

**Exp.:** Experimental B-factors (PDB id: 1UBQ);

**MD:** Molecular dynamics simulated B-factors;

**AFLF:** AlphaFold Latent Flooding simulated B-factors.

**Table S2.** PDB ids for the protein crystal structures with cryptic binding sites.

<b>Protein</b>	<b>Abbreviation</b>	<b>Apo state</b>	<b>Holo state</b>
Class A $\beta$ -lactamase TEM-1	TEM-1	1JWP	1PZO
Anti-apoptotic protein Bcl-x <sub>L</sub>	Bcl-x <sub>L</sub>	3FDL	2YXJ
Bovine $\beta$ -lactoglobulin	$\beta$ LG	1BSQ	1GX8
Human nucleoside diphosphate kinase A	NDPK-A	3L7U	2HVD
Human glycolipid transfer protein	GLTP	1SWX	2EUM

**Table S3.** AFLF parameters used for each system.

Parameters	UBQ	AdK	TEM-1	Bcl-x <sub>L</sub>	$\beta$ LG	NDPK-A	GLTP
$\mathcal{L}_{\text{repelling}}$							
$\beta_{ij}$ or $\beta_i$	$\beta_{ij}$	$\beta_i$	$\beta_i$	$\beta_i$	$\beta_i$	$\beta_i$	$\beta_i$
$b$	1.0	0.0	10.0	10.0	10.0	10.0	10.0
$\Delta t$	10	10	10	10	10	10	10
$K$	10	50	40	40	40	40	40
$\varsigma$	4.0	4.0	4.0	4.0	4.0	4.0	4.0
$\varepsilon_k$	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$w_k$	0.05	0.1	0.1	0.1	0.1	0.1	0.1
$\psi$ dropout rate	0.5	0.5	0.5	0.5	0.5	0.5	0.5
$\mathcal{L}_{\text{local}}$							
$w_{\text{local}}$	0.5	0.5	0.5	0.5	0.5	0.5	0.5
$\psi$ dropout rate	0.2	0.2	0.2	0.2	0.2	0.2	0.2
$\mathcal{L}_{\text{global}}$							
$\Delta d$	2.0	2.0	2.0	2.0	2.0	2.0	2.0
$L$	10	32	32	32	32	32	32
$\varsigma'$	4.0	4.0	4.0	4.0	4.0	4.0	4.0
$w_{\text{global}}$	3.0	0.5	0.5	0.5	0.5	0.5	0.5
$\psi$ dropout rate	0.2	0.2	0.2	0.2	0.2	0.2	0.2
$\mathcal{L}_{\text{anchoring}}$							
$\delta_{\text{Huber}}$	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$w_{\text{anchoring}}$	1.0	1.0	3.0	3.0	3.0	3.0	3.0
$\psi$ dropout rate	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\mathcal{L}_{\text{pLDDT}}$							
$w_{\text{pLDDT}}$	0.1	0.1	0.1	0.1	0.1	0.1	0.1
AdaM optimizer							
$\eta$ learning rate	0.002	0.002	0.003	0.003	0.003	0.003	0.003
$\beta_1$	0.9	0.9	0.9	0.9	0.9	0.9	0.9
$\beta_2$	0.999	0.999	0.999	0.999	0.999	0.999	0.999

**Table S4.** AFLF computational times.

<b>Protein</b>	<b>Number of residues</b>	<b>First-step wall time<sup>†</sup> (s)</b>	<b>Per-iteration wall time (s)</b>
Ubiquitin	76	155.48	5.57
AdK	214	179.22	16.87
TEM-1	263	186.36	23.26
Bcl-x <sub>L</sub>	158	174.41	11.57
$\beta$ LG	162	177.08	11.77
NDPK-A	152	161.27	10.49
GLTP	209	176.11	16.42

<sup>†</sup> The AFLF initial inference incurs by default a Just-In-Time (JIT) compilation overhead.