Boltzmann-Aligned Inverse Folding Model as a Predictor of Mutational Effects on Protein-Protein Interactions

Xiaoran Jiao, Weian Mao, Wengong Jin, Peiyuan Yang, Hao Chen, Chunhua Shen

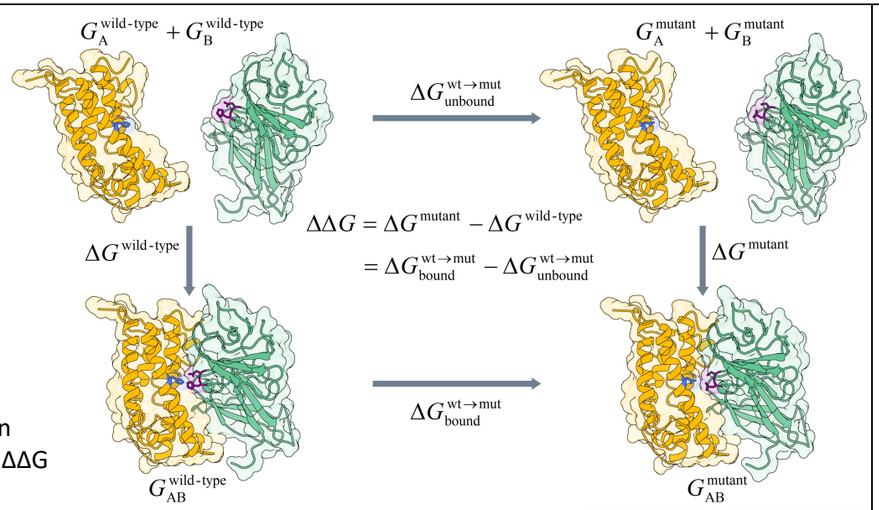
ICLR

Importance

- Predicting ΔΔG is critical for modulating protein-protein interactions essential for developing treatments, but limited by data scarcity.
- The availability of untrained protein sequence and structure data is becoming limited, emphasizing the need for new data modalities to train advanced models.

Contribution

- Establish a bidirectional connection between log-likelihood in inverse folding models and ΔΔG values, enabling mutual enhancement.

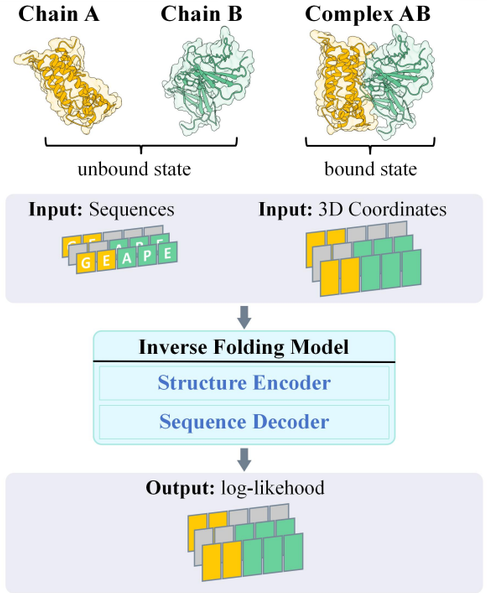


Boltzmann Alignment: Bridging Probability & Energy

Thermodynamic Cycle + Boltzmann Distribution

$$\begin{aligned}\Delta G &= G_{\text{bnd}} - G_{\text{unbnd}} = -k_{\text{B}}T \cdot (\log p_{\text{bnd}} - \log p_{\text{unbnd}}) \\ &= -k_{\text{B}}T \cdot (\log p(\mathcal{X}_{\text{bnd}} \mid \mathcal{S}_{\text{AB}}) - \log p(\mathcal{X}_{\text{unbnd}} \mid \mathcal{S}_{\text{AB}})) \\ &= -k_{\text{B}}T \cdot \left(\log \frac{p(\mathcal{S}_{\text{AB}} \mid \mathcal{X}_{\text{bnd}}) \cdot p(\mathcal{X}_{\text{bnd}})}{p(\mathcal{S}_{\text{AB}})} - \log \frac{p(\mathcal{S}_{\text{AB}} \mid \mathcal{X}_{\text{unbnd}}) \cdot p(\mathcal{X}_{\text{unbnd}})}{p(\mathcal{S}_{\text{AB}})}\right) \\ &= -k_{\text{B}}T \cdot \log \frac{p(\mathcal{S}_{\text{AB}} \mid \mathcal{X}_{\text{bnd}}) \cdot p(\mathcal{X}_{\text{bnd}})}{p(\mathcal{S}_{\text{AB}} \mid \mathcal{X}_{\text{unbnd}}) \cdot p(\mathcal{X}_{\text{unbnd}})}\end{aligned}$$

$$\begin{aligned}\Delta\Delta G &= \Delta G^{\text{mut}} - \Delta G^{\text{wt}} \\ &= -k_{\text{B}}T \cdot \left(\log \frac{p(\mathcal{S}_{\text{AB}}^{\text{mut}} \mid \mathcal{X}_{\text{bnd}}^{\text{mut}}) \cdot p(\mathcal{X}_{\text{bnd}}^{\text{mut}})}{p(\mathcal{S}_{\text{AB}}^{\text{mut}} \mid \mathcal{X}_{\text{unbnd}}^{\text{mut}}) \cdot p(\mathcal{X}_{\text{unbnd}}^{\text{mut}})} - \log \frac{p(\mathcal{S}_{\text{AB}}^{\text{wt}} \mid \mathcal{X}_{\text{bnd}}^{\text{wt}}) \cdot p(\mathcal{X}_{\text{bnd}}^{\text{wt}})}{p(\mathcal{S}_{\text{AB}}^{\text{wt}} \mid \mathcal{X}_{\text{unbnd}}^{\text{wt}}) \cdot p(\mathcal{X}_{\text{unbnd}}^{\text{wt}})}\right) \\ &= -k_{\text{B}}T \cdot \left(\log \frac{p(\mathcal{S}_{\text{AB}}^{\text{mut}} \mid \mathcal{X}_{\text{bnd}}^{\text{mut}})}{p(\mathcal{S}_{\text{AB}}^{\text{mut}} \mid \mathcal{X}_{\text{unbnd}}^{\text{mut}})} - \log \frac{p(\mathcal{S}_{\text{AB}}^{\text{wt}} \mid \mathcal{X}_{\text{bnd}}^{\text{wt}})}{p(\mathcal{S}_{\text{AB}}^{\text{wt}} \mid \mathcal{X}_{\text{unbnd}}^{\text{wt}})}\right)\end{aligned}$$



Benchmark Results and Downstream Applications

- State-of-the-Art ΔΔG Prediction on SKEMPIv2
- Broader Applicability: Validated in binding energy prediction, rigid protein-protein docking, and therapeutic antibody optimization.

Supervision	Method	Per-Structure		Overall				
		Pearson ↑	Spear. ↑	Pearson ↑	Spear. ↑	RMSE ↓	MAE ↓	AUROC ↑
✗	Rosetta	0.3284	0.2988	0.3113	0.3468	1.6173	1.1311	0.6562
	FoldX	0.3789	0.3693	0.3120	0.4071	1.9080	1.3089	0.6582
	ESM-1v	0.0073	-0.0118	0.1921	0.1572	1.9609	1.3683	0.5414
	PSSM	0.0826	0.0822	0.0159	0.0666	1.9978	1.3895	0.5260
	MSA Transformer	0.1031	0.0868	0.1173	0.1313	1.9835	1.3816	0.5768
	Tranception	0.1348	0.1236	0.1141	0.1402	2.0382	1.3883	0.5885
	B-factor	0.2042	0.1686	0.2390	0.2625	2.0411	1.4402	0.6044
	ESM-IF	0.2241	0.2019	0.3194	0.2806	1.8860	1.2857	0.5899
	MIF-Δlogit	0.1585	0.1166	0.2918	0.2192	1.9092	1.3301	0.5749
	RDE-Linear	0.2903	0.2632	0.4185	0.3514	1.7832	1.2159	0.6059
	BA-Cycle	0.3722	0.3201	0.4552	0.4097	1.8402	1.3026	0.6657
✓	DDGPred	0.3750	0.3407	0.6580	0.4687	1.4998	1.0821	0.6992
	End-to-End	0.3873	0.3587	0.6373	0.4882	1.6198	1.1761	0.7172
	MIF-Network	0.3965	0.3509	0.6523	0.5134	1.5932	1.1469	0.7329
	RDE-Network	0.4448	0.4010	0.6447	0.5584	1.5799	1.1123	0.7454
	DiffAffinity	0.4220	0.3970	0.6609	0.5560	1.5350	1.0930	0.7440
	Prompt-DDG	0.4712	0.4257	0.6772	0.5910	1.5207	1.0770	0.7568
	ProMIM	0.4640	0.4310	0.6720	0.5730	1.5160	1.0890	0.7600
	Surface-VQMAE	0.4694	0.4324	0.6482	0.5611	1.5876	1.1271	0.7469
	BA-DDG	**0.5453**	**0.5134**	**0.7118**	**0.6346**	**1.4516**	**1.0151**	**0.7726**