Extending Upside, A near-atomic level model for fast protein folding, for predicting protein-protein interactions

Nabil F. Faruk[1], John Jumper[2], Benoit Roux[3], Tobin Sosnick[3]

[1]Biophysical Sciences, [2]Chemistry, [3]Biochemistry and Molecular Biology, University of Chicago

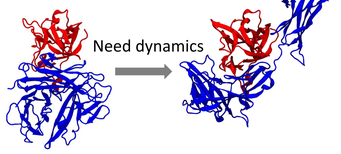
Biophysical Sciences
THE UNIVERSITY OF CHICAGO

Abstract

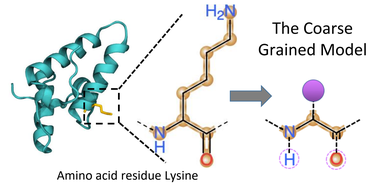
We are extending *Upside*, a near-atomic level model previously developed in our group for fast folding of proteins in implicit solvent via Langevin dynamics, for predicting protein-protein interactions and binding affinities. *Upside*'s speed arises from only explicitly accounting for the backbone N, Cα, and C atoms during the dynamics portion, while it infers the position of pendant groups representing side chains and propagates their forces onto the backbone. The reconstructed protein maintains detailed Rama maps, H-bonding, and side chain potentials, setting *Upside* apart from typical GG models. This provides a solid foundation to study protein-protein interactions. Here we utilize *Upside* for this application by retraining the potential with a benchmark set of protein complexes. We find mixed success.

Current Scope

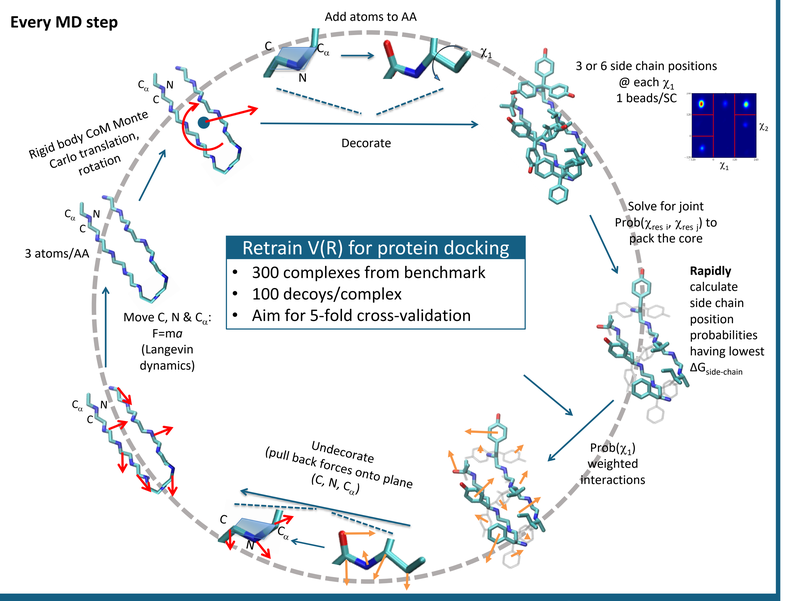
- Docking: predict the geometry of bound poses
- Want native pose (corresponding to crystal structure) to be highest ranked



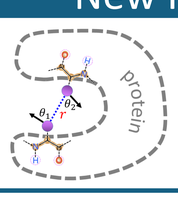
CG Model



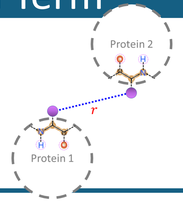
Upside Computational Loop



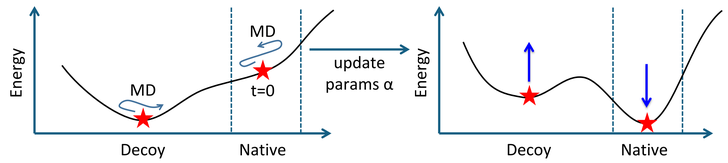
New Inter-protein Potential Term



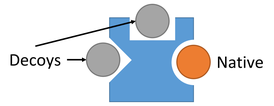
- $V_{SC} = V_{SC}(r, \theta_1, \theta_2) + V_{SC\_inter}(r)$
- Composed of cubic spline basis functions
- Advantages:
  - >4x fewer parameters to train
  - Modular



Training Objective



- Upside already trained on single protein chain data set for protein folding
- →info relevant to protein-protein interactions underrepresented
- We want to minimize native pose's potential energy w.r.t. params α and therefore maximize its (Boltzmann) probability:

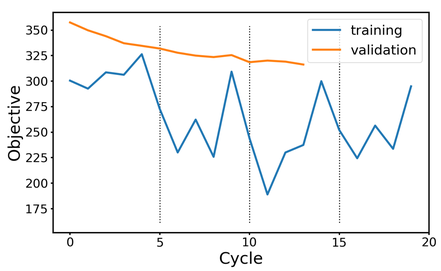


$$\max_{\alpha} \frac{1}{N}\sum_{i}^{N}\left(\frac{e^{-\beta\langle E\rangle_0^i}}{\sum_{k=0}^{m} e^{-\beta\langle E\rangle_k^i}}\right) \Rightarrow \min_{\alpha} \frac{1}{N}\sum_{i}^{N}\left[-\ln\left(\frac{e^{-\beta\langle E\rangle_0^i}}{\sum_{k=0}^{m} e^{-\beta\langle E\rangle_k^i}}\right)\right]$$

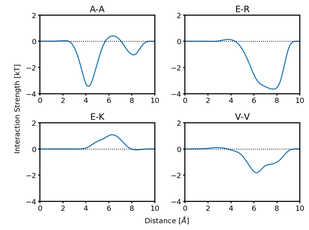
$\langle E\rangle_k^i$, $k = 0$ is native, $k \neq 0$ is decoy' for N complexes

Training Results

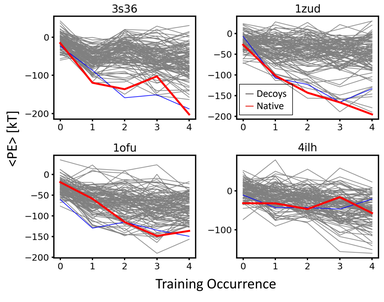
Minibatch (Size 48) Training



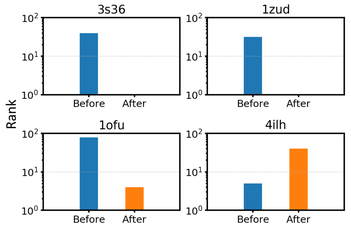
$V_{SC\_inter}(r)$ after training



Evolution of <PE> During Training for 4 Complexes



Change in Native Rank After Training



Rank: order of <PE> of native pose vs. 100 decoys. Rank 1 means it has the lowest <PE>

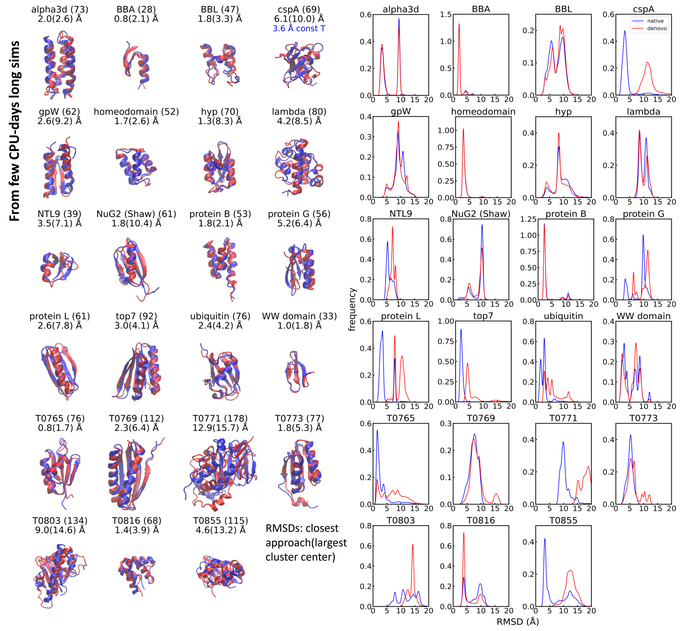
Summary Stats

	Rank Improvement	Rank Degradation	Top 5 Before	Top 5 After	Top 5 New
Training (240)	5573	641	67	122	73
Validation (60)	1015	238	19	26	13

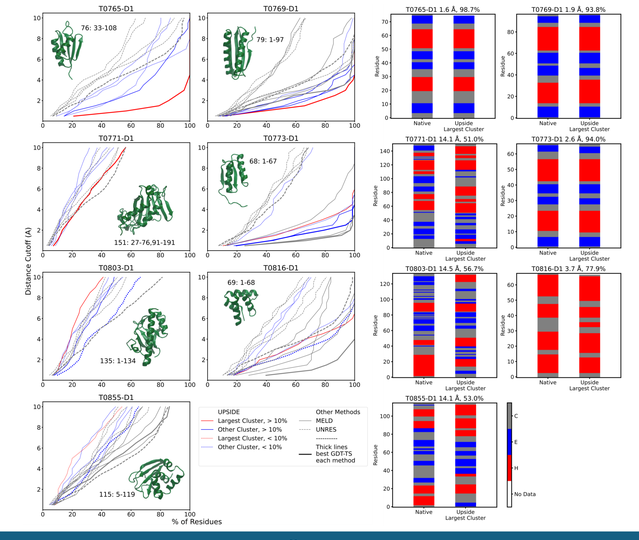
Ongoing and Future Work

- Try different forms of training objective
- More extensive validation and testing: higher degree of cross-validation, longer simulations, start from unbound conformations
- Expand training set

Success in De Novo Protein Folding



CASP Targets: Global Distance Test



References

- J. M. Jumper, K. F. Freed, and T. R. Sosnick, "New Methods using Rigorous Machine Learning for Coarse-Grained Protein Folding and Dynamics," Ph.D. Thesis, 2017.
- P. J. Kundrotas et al., "Dockground: A comprehensive data resource for modeling of protein complexes," *Protein Science*, vol. 27, no. 1, pp. 172–181, Epub Oct. 2017.
- D.. Kozakov *et al.*, "The ClusPro web server for protein–protein docking," *Nat. Protocols*, vol. 12, no. 2, pp. 255–278, Feb. 2017.
- M. Parisien, K. F. Freed, and T. R. Sosnick, "On Docking, Scoring and Assessing Protein-DNA Complexes in a Rigid-Body Framework," *PLOS ONE*, vol. 7, no. 2, p. e32647, Feb. 2012.

Acknowledgements

NSERC CRSNG
NIH
NSF