

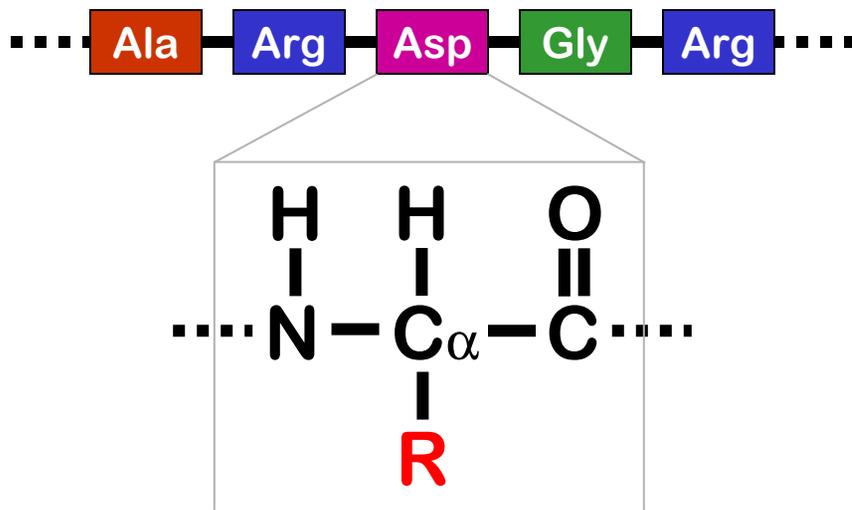
Homotopy Optimization Methods and Protein Structure Prediction

Daniel M. Dunlavy

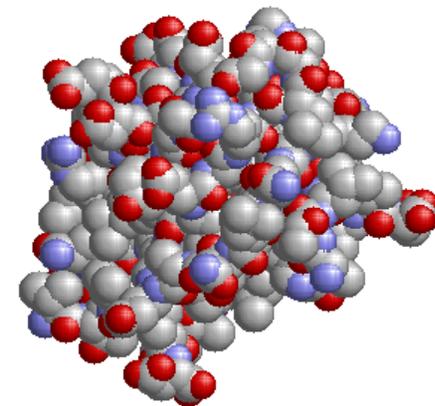
*Applied Mathematics and Scientific Computation
University of Maryland, College Park*

Protein Structure Prediction

Amino Acid Sequence



Protein Structure



Given the amino acid sequence of a protein (1D), is it possible to predict its native structure (3D)?

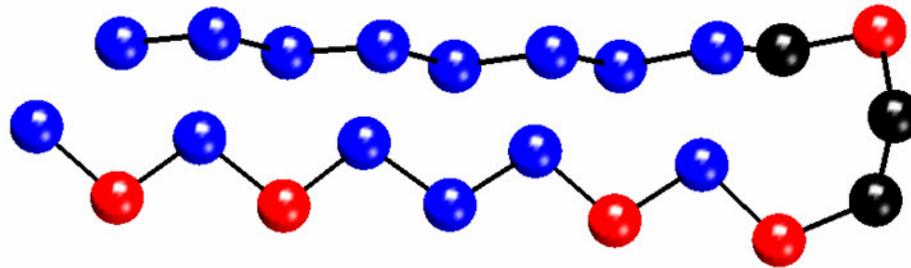
Protein Structure Prediction

- **Given:**
 - Protein model
 - Properties of constituent particles
 - Potential energy function (force field)
- **Goal:**
 - Predict native (lowest energy) conformation
 - Thermodynamic hypothesis [Anfinsen, 1973]
 - Develop hybrid method, combining:
 - Energy minimization [numerical optimization]
 - Comparative modeling [bioinformatics]
 - Use **template** (known structure) to predict **target** structure

Protein Model: Particle Properties

- Backbone model

- Single chain of particles with residue attributes
- Particles model C_{α} atoms in proteins



- Properties of particles

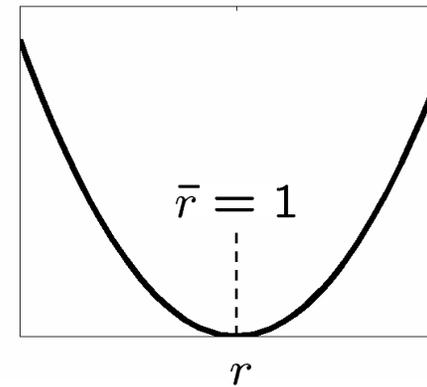
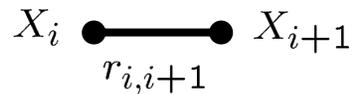
- Hydrophobic, Hydrophilic, Neutral
- Diverse hydrophobic-hydrophobic interactions

[Veitshans, Klimov, and Thirumalai. *Protein Folding Kinetics*, 1996.]

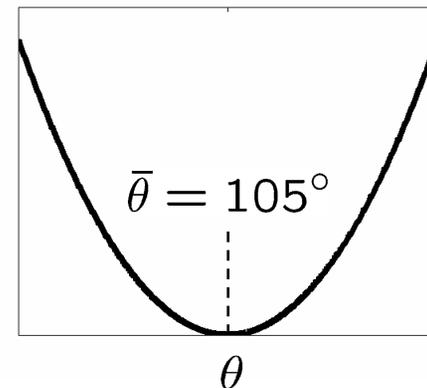
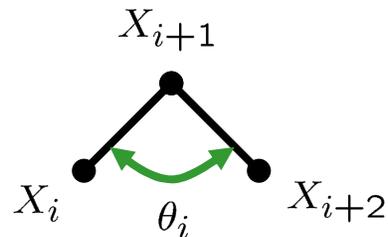
Potential Energy Function

$$E(X) = E_{bl}(X) + E_{ba}(X) + E_{dih}(X) + E_{non}(X)$$

$$E_{bl}(X) = \sum_{i=1}^{n-1} \frac{k_r}{2} (r_{i,i+1} - \bar{r})^2$$



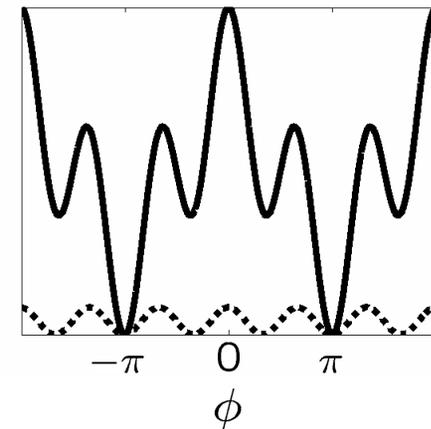
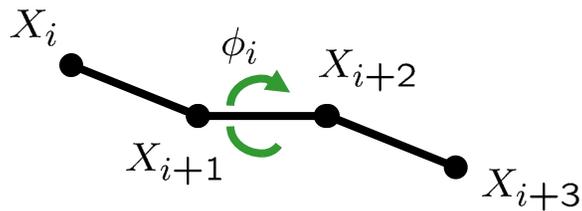
$$E_{ba}(X) = \sum_{i=1}^{n-2} \frac{k_\theta}{2} (\theta_i - \bar{\theta})^2$$



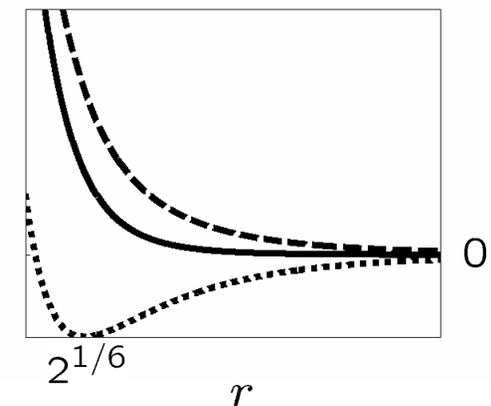
Potential Energy Function

$$E(X) = E_{bl}(X) + E_{ba}(X) + E_{dih}(X) + E_{non}(X)$$

$$E_{dih}(X) = \sum_{i=1}^{n-3} [A_i(1 + \cos \phi_i) + B_i(1 + \cos 3\phi_i)]$$



$$E_{non}(X) = \sum_{i=1}^{n-3} \sum_{j=i+3}^n \gamma_{ij} \left\{ \alpha_{ij} \left(\frac{\bar{r}}{r_{ij}} \right)^{12} - \beta_{ij} \left(\frac{\bar{r}}{r_{ij}} \right)^6 \right\}$$



Homotopy Optimization Method (HOM)

- Goal

- Minimize energy function of **target** protein:

$$E^1(X^*) = \min_{X \in \mathbb{R}^{3n}} E^1(X), \quad (E^1 : \mathbb{R}^{3n} \rightarrow \mathbb{R})$$

- Steps to solution

- Energy of **template** protein: $E^0(X^0) = \min_{X \in \mathbb{R}^{3n}} E^0(x)$

- Define a **homotopy** function:

- $H(X, \lambda) = \rho^0(\lambda)E^0(X) + \rho^1(\lambda)E^1(X)$

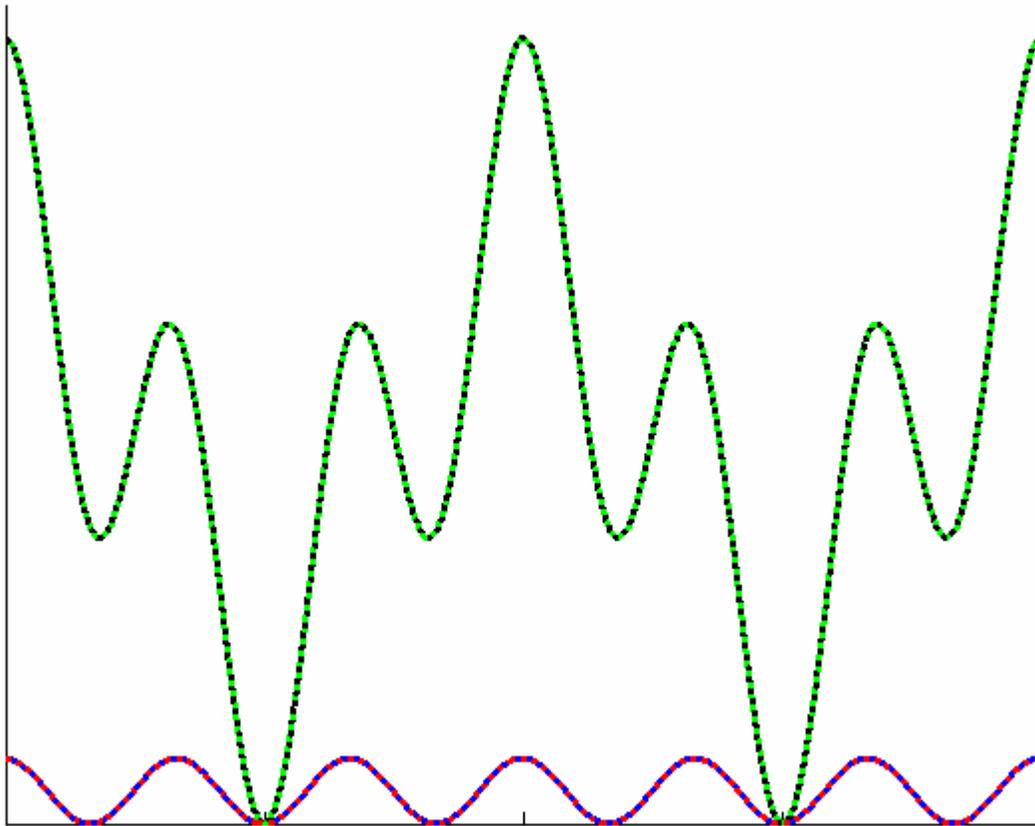
- Deforms template protein into target protein

- Produce sequence of minimizers of $H(X, \lambda)$ starting at $\lambda = 0$ and ending at $\lambda = 1$

Energy Landscape Deformation

Dihedral Terms

$$E_{dih}(X) = \sum_{i=1}^{n-3} [A_i(1 + \cos \phi_i) + B_i(1 + \cos 3\phi_i)]$$

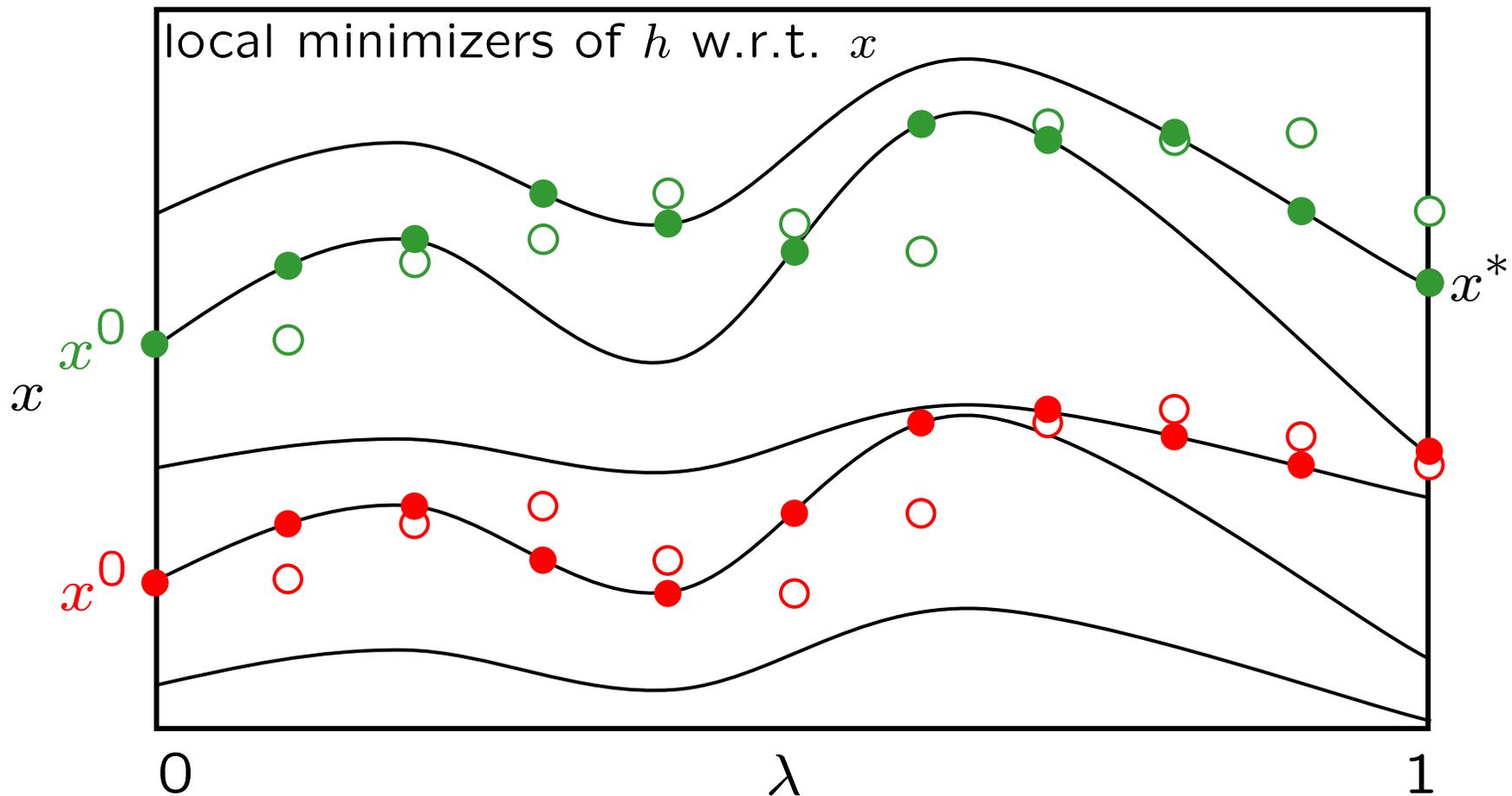


$$\lambda = 1.00$$

Neutral Particles		
Template		Target
≥ 2	—	≥ 2
< 2	—	< 2
≥ 2	⋯	< 2
< 2	⋯	≥ 2

Illustration of HOM

$$f^1(x^*) = \min_{x \in \mathbb{R}} f^1(x) \quad h(x, \lambda) = (1-\lambda)f^0(x) + \lambda f^1(x)$$



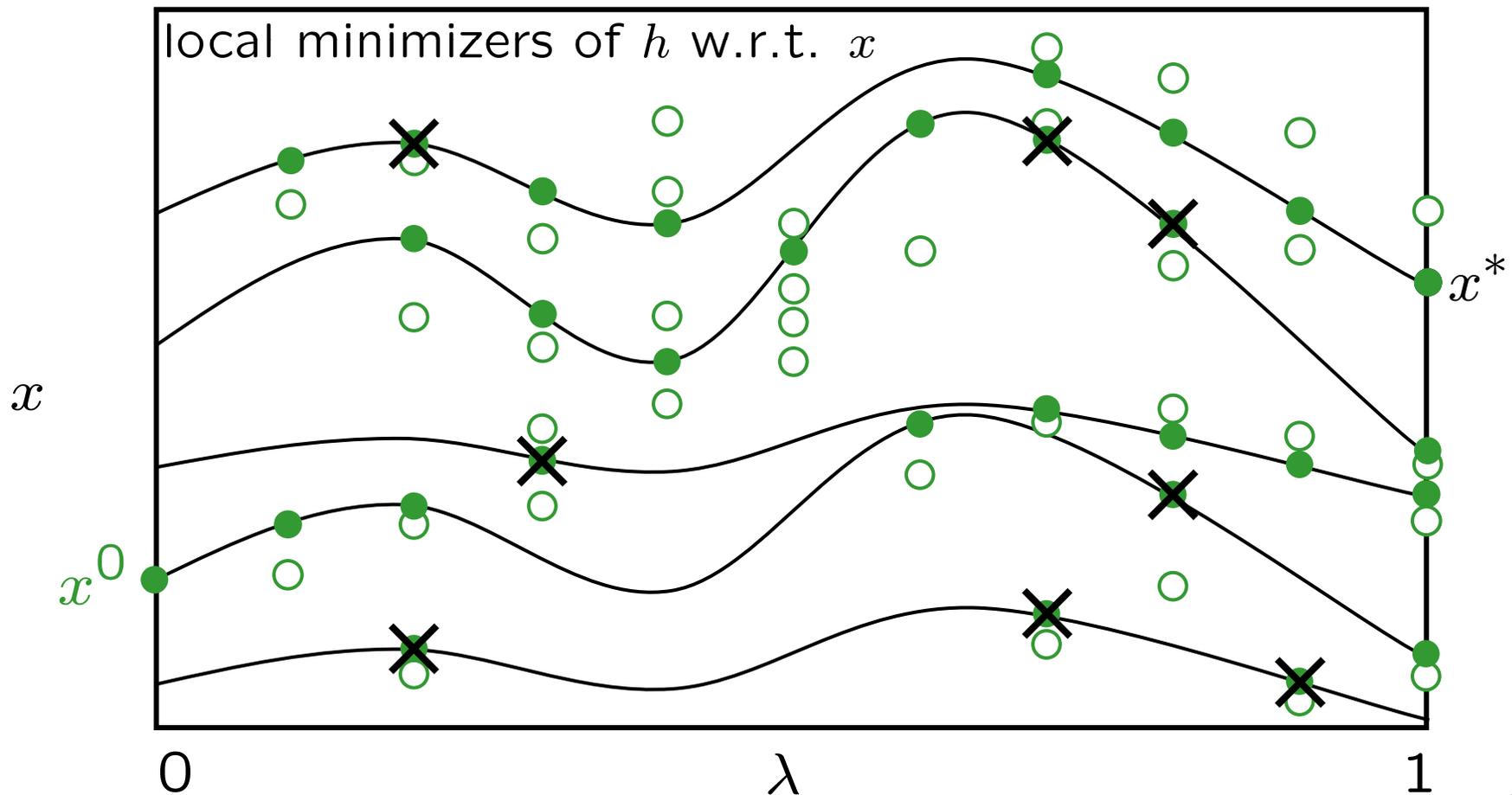
Homotopy Optimization using Perturbations & Ensembles (HOPE)

- Improvements over HOM
 - Produces ensemble of sequences of **local** minimizers of $h(x, \lambda)$ by perturbing intermediate results
 - Increases likelihood of predicting **global** minimizer
- Algorithmic considerations
 - Maximum ensemble size
 - Determining ensemble members

Illustration of HOPE

Maximum ensemble size = 2

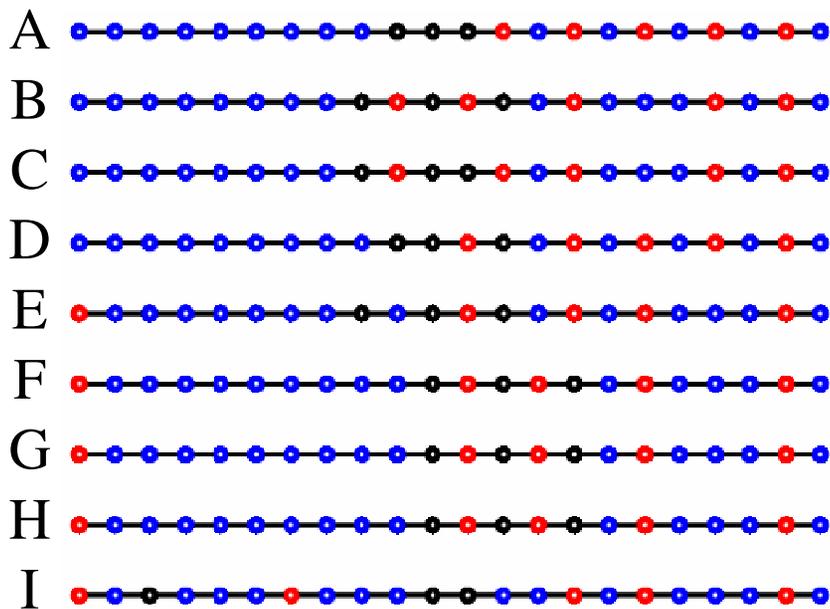
$$f^1(x^*) = \min_{x \in \mathbb{R}} f^1(x) \quad h(x, \lambda) = (1-\lambda)f^0(x) + \lambda f^1(x)$$



Numerical Experiments

9 chains (22 particles) with known structure

Loop Region

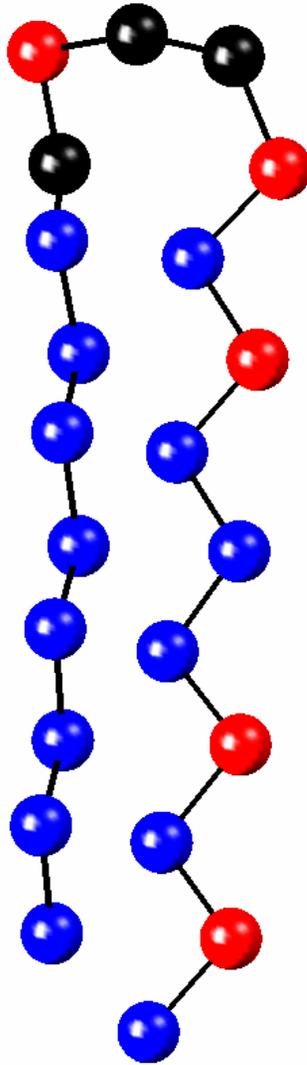


Sequence Homology (%)

	A	B	C	D	E	F	G	H	I
A	100								
B	77	100							
C	86	91	100						
D	91	86	77	100					
E	73	82	73	82	100				
F	68	68	59	77	86	100			
G	68	68	59	77	86	100	100		
H	68	68	59	77	86	100	100	100	
I	73	59	64	68	77	73	73	73	100

Hydrophobic Hydrophilic Neutral

Numerical Experiments



Numerical Experiments

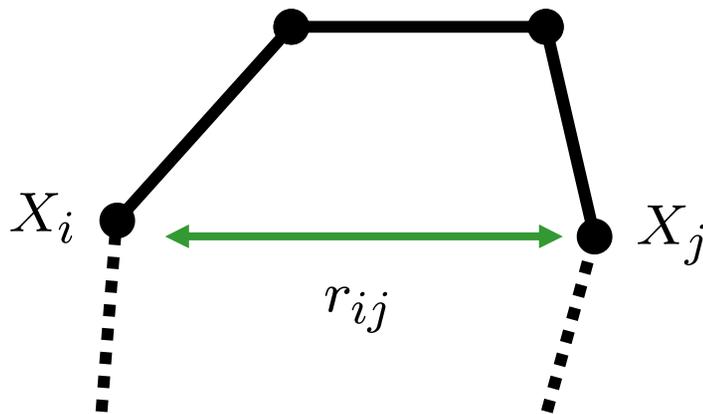
- 62 template-target pairs
 - 10 pairs had identical native structures
- Methods
 - HOM vs. Newton's method w/trust region (N-TR)
 - HOPE vs. simulated annealing (SA)
 - Different ensemble sizes (2,4,8,16)
 - Averaged over 10 runs
 - Perturbations where sequences differ
- Measuring success
 - Structural overlap function: $0 \leq \chi \leq 1$
 - Percentage of interparticle distances off by more than 20% of the average bond length (\bar{r})
 - Root mean-squared deviation (RMSD)

Ensemble SA
Basin hopping
 $T_0 = 10^5$
Cycles = 10
Berkeley schedule

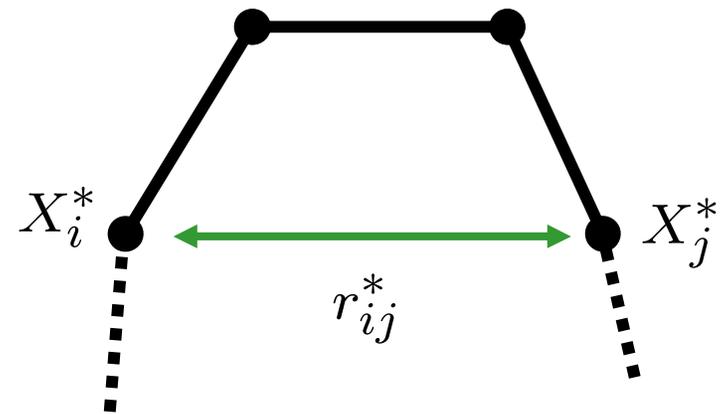
Structural Overlap Function

$$\chi = 1 - \frac{2}{n^2 - 5n + 6} \sum_{i=1}^{n-3} \sum_{j=i+3}^n \Theta \left(0.2\bar{r} - |r_{ij} - r_{ij}^*| \right)$$

$$\Theta(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$



Predicted



Native

RMSD

Measures the distance between corresponding particles in the predicted and lowest energy conformations when they are optimally superimposed.

$$RMSD(X) = \min_{S(X)} \sqrt{\frac{1}{n} \sum_{i=1}^n \|X_i - X_i^*\|^2}$$

where $S(X)$ is a rotation and translation of X

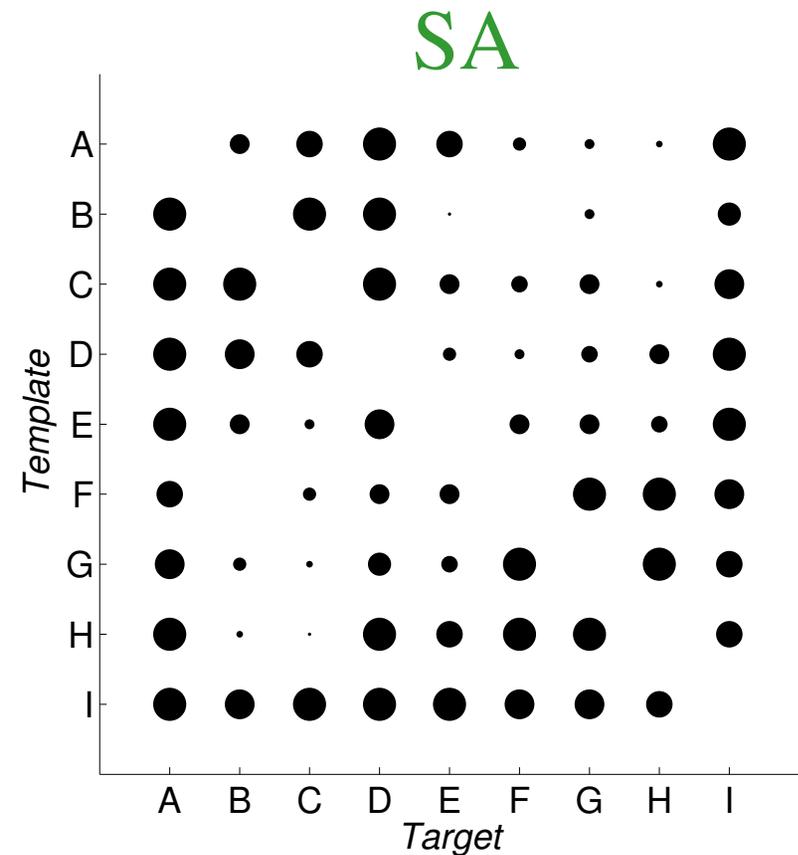
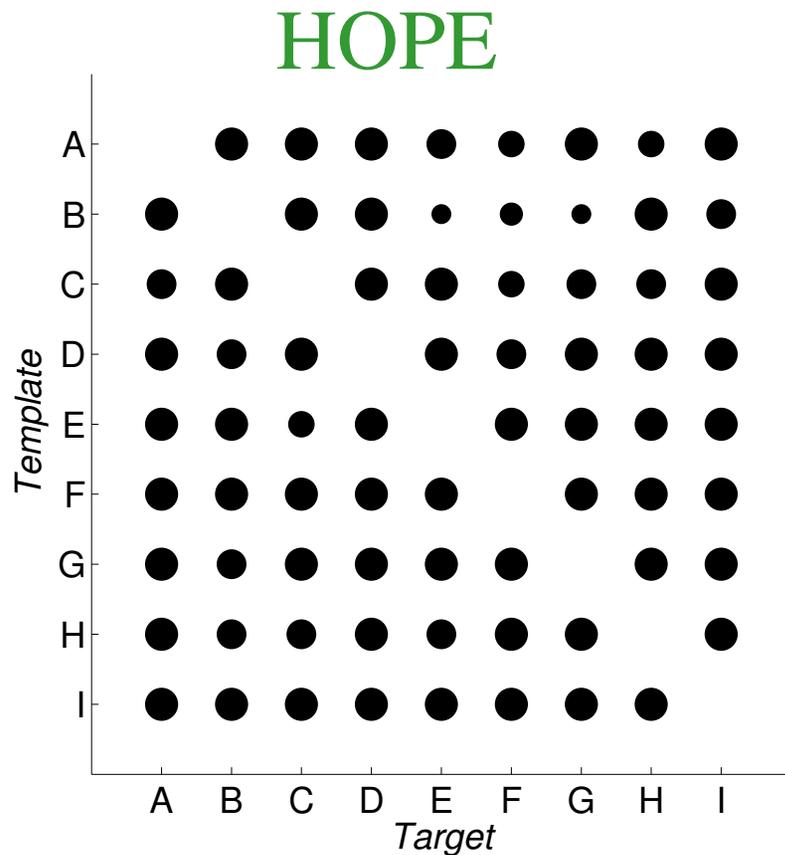
Results

Method		$\chi = 0$	Success	Mean χ	Mean RMSD	Time (sec)
<i>HOM</i>		15	0.24	0.36	0.38	10
<i>N-TR</i>		4	0.06	0.45	0.55	1

Method	Ensemble Size	$\chi = 0$	Success	Mean χ	Mean RMSD	Time (sec)
<i>HOPE</i>	2	33.40	0.54	0.14	0.17	35
	4	43.10	0.70	0.08	0.11	65
	8	54.60	0.88	0.03	0.04	115
	16	59.00	0.95	0.01	0.02	200
<i>SA</i>	2	13.10	0.21	0.27	0.36	52
	4	20.80	0.34	0.19	0.26	107
	8	28.50	0.46	0.13	0.19	229
	16	40.20	0.65	0.08	0.12	434

Results

Success of HOPE and SA with ensembles of size 16 for each template-target pair. The size of each circle represents the percentage of successful predictions over the 10 runs.



Conclusions

- Homotopy optimization methods
 - More successful than standard minimizers
- HOPE
 - For problems with f^0, x^0 (E^0, X^0) readily available
 - Solves protein structure prediction problem
 - Outperforms ensemble-based simulated annealing
- Future work
 - Protein Data Bank (templates), TINKER (energy)
 - Convergence analysis for HOPE

Acknowledgements

- Dianne O'Leary (UM)
 - Advisor
- Dev Thirumalai (UM), Dmitri Klimov (GMU)
 - Model, numerical experiments
- Ron Unger (Bar-Ilan)
 - Problem formulation
- National Library of Medicine (NLM)
 - Grant: F37-LM008162

Thank You

Daniel Dunlavy – HOPE

<http://www.math.umd.edu/~ddunlavy>

ddunlavy@math.umd.edu