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 $C_\alpha$ atoms in proteins

  ![Protein Model Diagram]

• Properties of particles
  – Hydrophobic, Hydrophilic, Neutral
  – Diverse hydrophobic-hydrophobic interactions

Potential Energy Function

\[ E(X) = E_{bl}(X) + E_{ba}(X) + E_{dih}(X) + E_{non}(X) \]

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

\[ E_{ba}(X) = \sum_{i=1}^{n-2} \frac{k_\phi}{2} (\theta_i - \bar{\theta})^2 \]
Potential Energy Function

\[ E(X) = E_{bl}(X) + E_{ba}(X) + E_{dih}(X) + E_{non}(X) \]

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

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

• Goal
  – Minimize energy function of target protein:
    \[ E^1(X^*) = \min_{X \in \mathbb{R}^{3n}} E^1(X), \quad (E^1 : \mathbb{R}^{3n} \to \mathbb{R}) \]

• 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

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

\[ \lambda = 1.00 \]

<table>
<thead>
<tr>
<th>Neutral Particles</th>
<th>Template</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 2</td>
<td>≥ 2</td>
<td>≥ 2</td>
</tr>
<tr>
<td>&lt; 2</td>
<td>&lt; 2</td>
<td>&lt; 2</td>
</tr>
<tr>
<td>≥ 2</td>
<td>..........</td>
<td>&lt; 2</td>
</tr>
<tr>
<td>&lt; 2</td>
<td>..........</td>
<td>≥ 2</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>77</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>86</td>
<td>91</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>91</td>
<td>86</td>
<td>77</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>73</td>
<td>82</td>
<td>73</td>
<td>82</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>68</td>
<td>68</td>
<td>59</td>
<td>77</td>
<td>86</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>68</td>
<td>68</td>
<td>59</td>
<td>77</td>
<td>86</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>68</td>
<td>68</td>
<td>59</td>
<td>77</td>
<td>86</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>73</td>
<td>59</td>
<td>64</td>
<td>68</td>
<td>77</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>100</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Ensemble SA</th>
<th>Basin hopping</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_0 = 10^5$</td>
<td>Cycles = 10</td>
</tr>
<tr>
<td>Berkeley schedule</td>
<td></td>
</tr>
</tbody>
</table>
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 \hspace{2cm} Native
**RMSD**

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

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

where \( S(X) \) is a rotation and translation of \( X \)
# Results

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