Protein Refinement in
Discovery Studio 1.7

Presented by: Francisco Hernandez-Guzman, PhD

February 2007
Agenda

Part I: Protein Refinement in Discovery Studio® 1.7 (20mins)
  • The problem
  • Novel Algorithms at Accelrys

Part II: DEMO (25mins)

Conclusion
Importance and Difficulties

- Importance of side-chain and loop refinement
  - X-ray
    - Missing data
    - Data is constrained → crystal contacts
    - Proteins are dynamics entities
  - Homology modeling
    - Inherent errors and bias from template(s)
    - Missing data
  - *In-Silico* Structure Based Drug Design
    - Evidence indicates the need to consider flexibility to open active sites, and allow proper docking

- Difficulties in side-chain and loop refinement
  - Intrinsic flexibility → difficult combinatorial problem
    - 1 peptide → 2-3 rot. bonds.
Protein Refinement in Discovery Studio 1.7

- Protein refinement tools can be used to:
  - Correct errors in exp. structs. due to missing exp. data.
  - Provide alternate starting conformations for docking

- Discovery Studio 1.7 provides two novel algorithms for:
  - Side-chain refinement (rebuilding) → ChiRotor\(^1\)
  - Loop refinement (rebuilding) → Looper\(^2\)

Side-Chain Refinement using ChiRotor

ChiRotor

Optimizes:
- side-chains conf. → systematic CHARMM E search and minimization.

Selects:
- Best conf. based on CHARMM E.

Fast *ab initio* approach

↑ independent on the starting structure

↑ Rebuilds the side-chains
ChiRotor Methodology

- Protein 3D Structure
- Select a Set of \( n \) Residues For Refinement
- Remove all side chain atoms of selected residues

Start loop for \( i \) from 1 to \( n \)

Choose Residue \( i \)

Sample side chain conformations of residue \( i \) varying \( \chi_1 \)

Energy Minimize side chain atoms of residue \( i \) in CHARMM

Save 2 Best Conformations for residue \( i \)

End loop for \( i \)

Output: 2\( n \) partial structures

Construct complete structure using lowest energy conformer of each residue.

Energy minimize all selected side chains

Start loop for \( i \) from 1 to \( n \)

Replace side chain conformation \( i \) with the 2\(^{nd} \) conformer and energy minimize

Accept the structure if energy is lower.

End loop for \( i \)

Output: 1 Lowest Energy structure
ChiRotor Results

• Compared ChiRotor to SCWRL and SCAP
  - Benchmark includes
    • 24 high resolution X-ray crystal structures selected using PISCES
    • Resolution < 1.0 Å
    • Sequence identity < 20%
  - ChiRotor
    • More accurate in the core region
    • Relatively fast
      • Computation time is similar to SCAP (fast mode) and much faster (10 to 20 times) than SCAP (slow mode)
    • Not based on a rotamer library
    • Predicted conformations are already minimized by CHARMM thus no further refinement is necessary

<table>
<thead>
<tr>
<th>Method</th>
<th>Core RMSD</th>
<th>All RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChiRotor</td>
<td>0.8</td>
<td>1.7</td>
</tr>
<tr>
<td>SCWRL</td>
<td>1.1</td>
<td>1.7</td>
</tr>
<tr>
<td>SCAP (Fast mode)</td>
<td>0.9</td>
<td>1.6</td>
</tr>
<tr>
<td>SCAP (Slow mode)</td>
<td>0.7</td>
<td>1.5</td>
</tr>
</tbody>
</table>

ChiRotor Results

- The CHARMM polar force field is more accurate and faster
  - CPU time scales close to $O(N^2)$ where $N$ is the number of residues
Loop Refinement using **Looper**

- The Looper loop refinement algorithm
  - Systematically searches for backbone conformation
  - Uses CHARMm minimization
  - Ranks the conformation using CHARMm energy, including solvation energy term
  - Fast, *ab initio* approach independent of the starting structure
    → Rebuilds the loop *de novo*
Looper Methodology

- Looper first constructs and optimizes the loop backbone
  - Systematic search of loop conformation by sampling a minimum set of backbone dihedral angles $\phi$ and $\psi$
  - The loop is divided into two halves and each half is constructed independently from the end of the loop and then combined without the side-chain atoms
  - The loops are minimized by CHARMM and ranked by CHARMM energy

Continued
Looper Methodology

- Next, Looper constructs the loop side chains and optimizes the loops
  - Construct the side-chains using the ChiRotor and the loop conformations are ranked by CHARMM energies

- Finally, Looper re-ranks the conformations
  - CHARMM energy minimizations in the first two steps are done without including the solvation energy term
  - In this step, each top ranking loop conformation is re-scored by adding the solvation energy term calculated using Generalized Born approximation (in CHARMM)
Looper Results

- Compared Looper to Fiser\textsuperscript{1} and Jacobson\textsuperscript{2} publications

  - Benchmark
    - Test Data Sets were taken from *Fiser’s\textsuperscript{1} work*
    - For each loop length, 40 loops from 40 different proteins are modeled

  - RMSD were calculated between the model and experimental structure for backbone heavy atoms using the global RMSD definition by Fiser

  - Jacobson’s calculation are done in crystal environment which may result in better accuracy, but not practical for model building

Looper Results (accuracy)

- Looper is more accurate than Fiser’s
- Jacobson results are better for longer loops
- MODELER 9.0 loop refinement method (includes the DOPE function) is better than Fiser’s

MODELER 6.0 (Fiser)  Jacobson
Looper Results (speed)

- Looper results are similar to Jacobson’s stage 1, but much faster (>100 times faster) than stages 2 and 3
- Jacobson’s stage 3 results are better for longer loops, but with a high computational cost (>100 times slower than looper)
- Fiser’s timing is similar to Jacobson’s methods
DEMO
Conclusions

One of the main obstacles to any method for protein structure prediction: **Combinatorial problem**

- In the case of loop and side chain optimizations: intelligent assumptions → the difficulty is reduced, but still significant

- We demonstrated that it is possible to create efficient algorithms for loop (Looper) and side-chain (ChiRotor) optimizations by reducing the search to the minimal number of initial conformations of each amino-acid residue in combination with energy minimization

*Continued*
Conclusions

Continued

- The CHARMM scripts for side-chain and loop predictions are developed entirely on physical principles and do not use any bases of structural data or rotamer libraries.

- The script realization of the algorithms makes it easy to include them as part of any CHARMM modeling protocol.

- The accuracy of the methods is comparable to the accuracy of other known algorithms.
  - In the case of ChiRotor, we have achieved an extremely fast code without losing accuracy.
Acknowledgements

R&D
- Velin Spassov
- Lisa Yan
- Paul Flook
- Marc Fasnacht
- Jürgen Koska

Marketing
- Dana Haley-Vicente

Accelrys