Protein Structure Refinement using Replica Exchange Monte Carlo Sampling

Mohsen Chitsaz
Molecular Biophysics and Biochemistry and Computer Science

April 30, 2012

What are proteins?

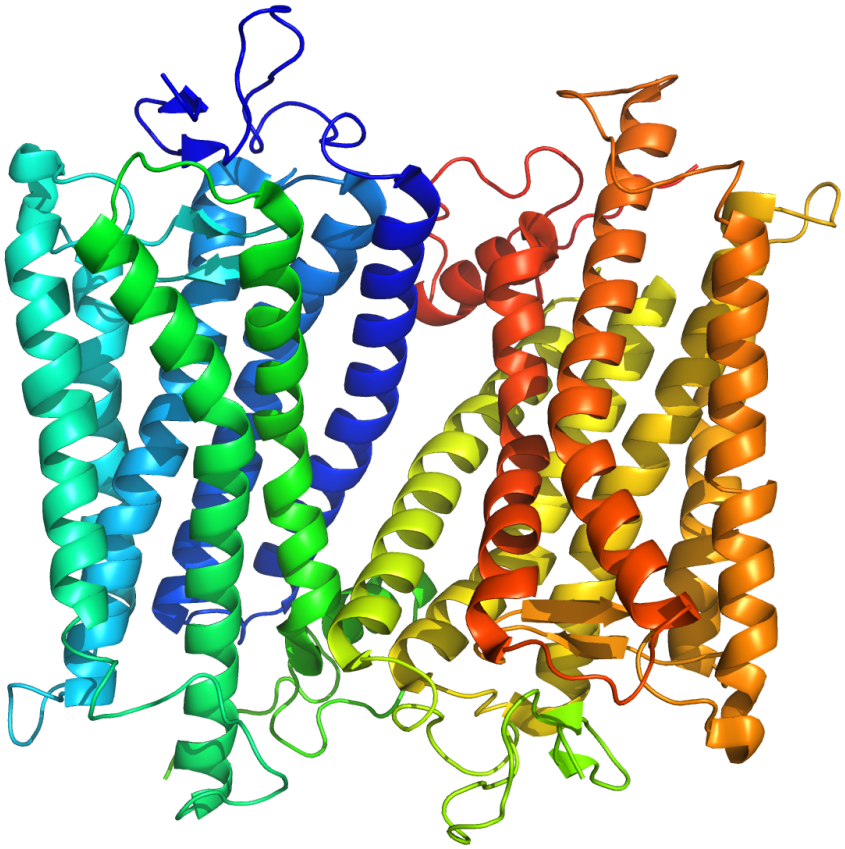
- Big molecules with about 10,000 atoms
- Composed of smaller chemical building blocks called amino acids
- Only 20 different amino acids in nature
- Responsible for almost all tasks in living organisms

Where can proteins be found?

- Every cell of living organisms
- Human cells are operated by cooperation of about 25,000 different proteins
- Viruses are mainly composed of proteins

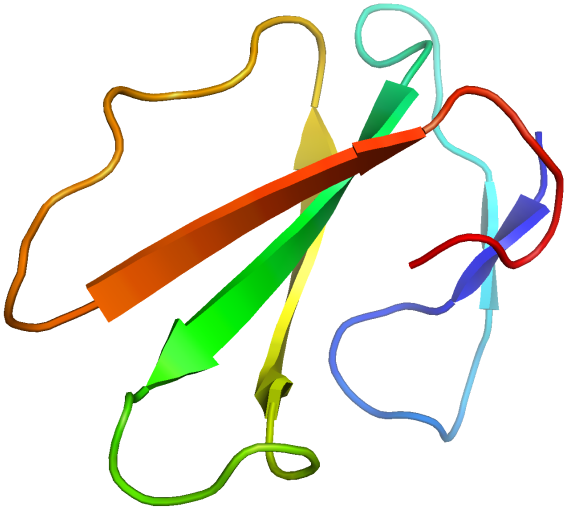
Proteins are crucial!

- Rhodopsin, a light sensor that results in seeing sense

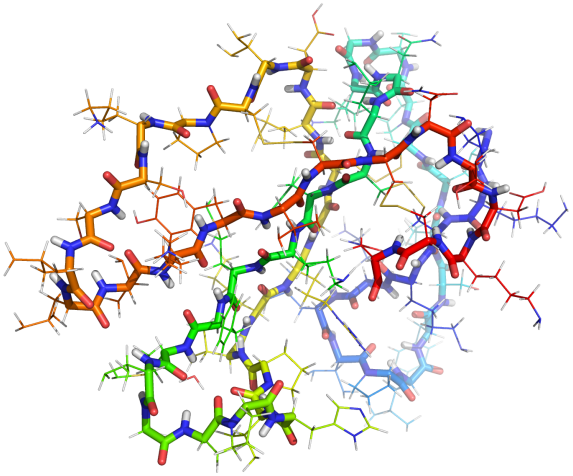


Proteins can be dangerous!

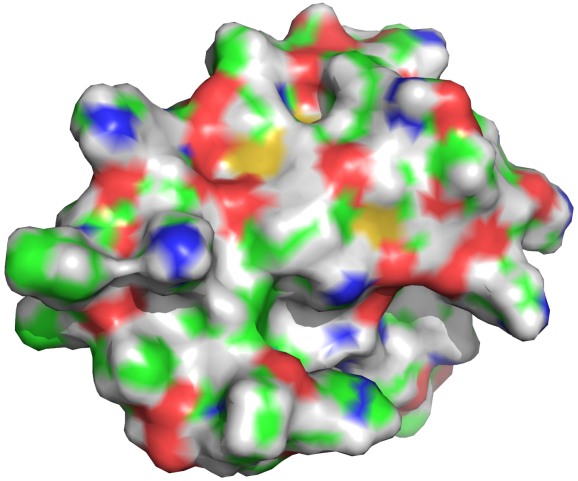
- King Cobra venom contains a protein that makes the heart stop beating



Cartoon representation

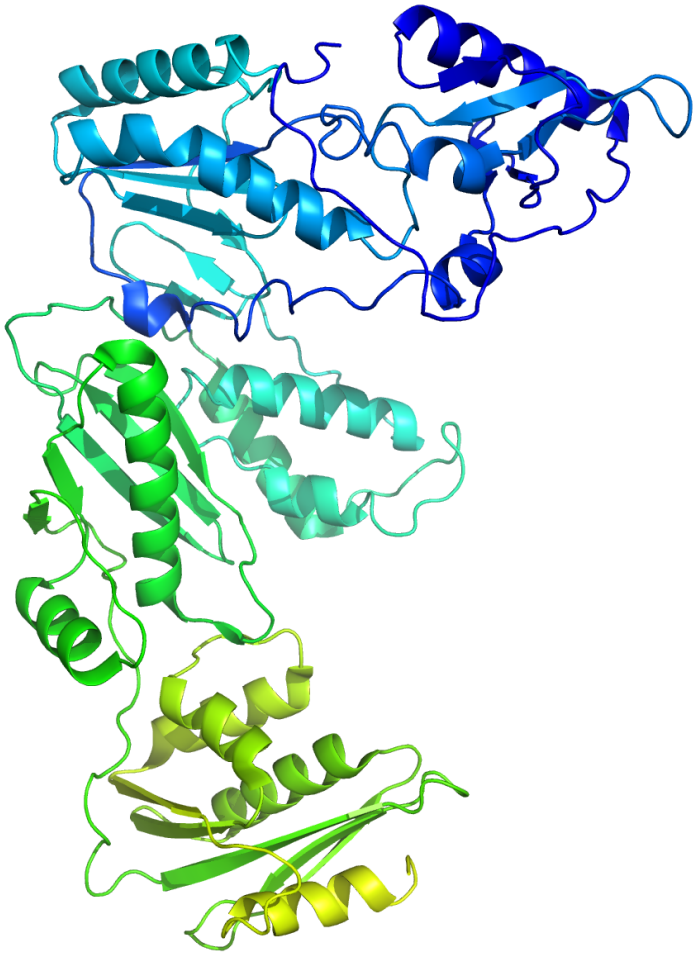


Stick representation



Surface representation

HIV Replication Machinery

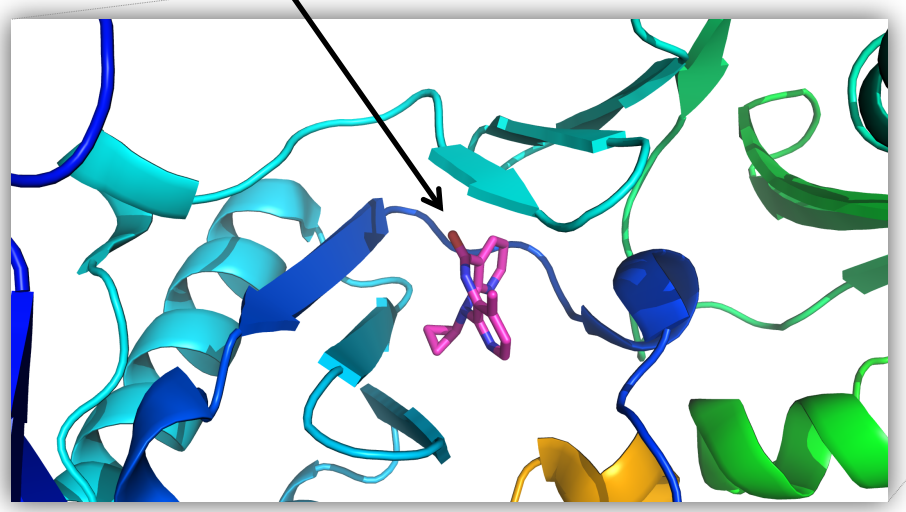


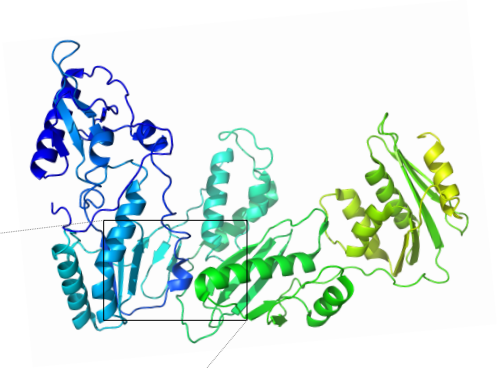
Reverse Transcriptase
(copies virus genetic information)

Integrase
(integrates virus DNA into the human's DNA)

Nevirapine (HIV drug)

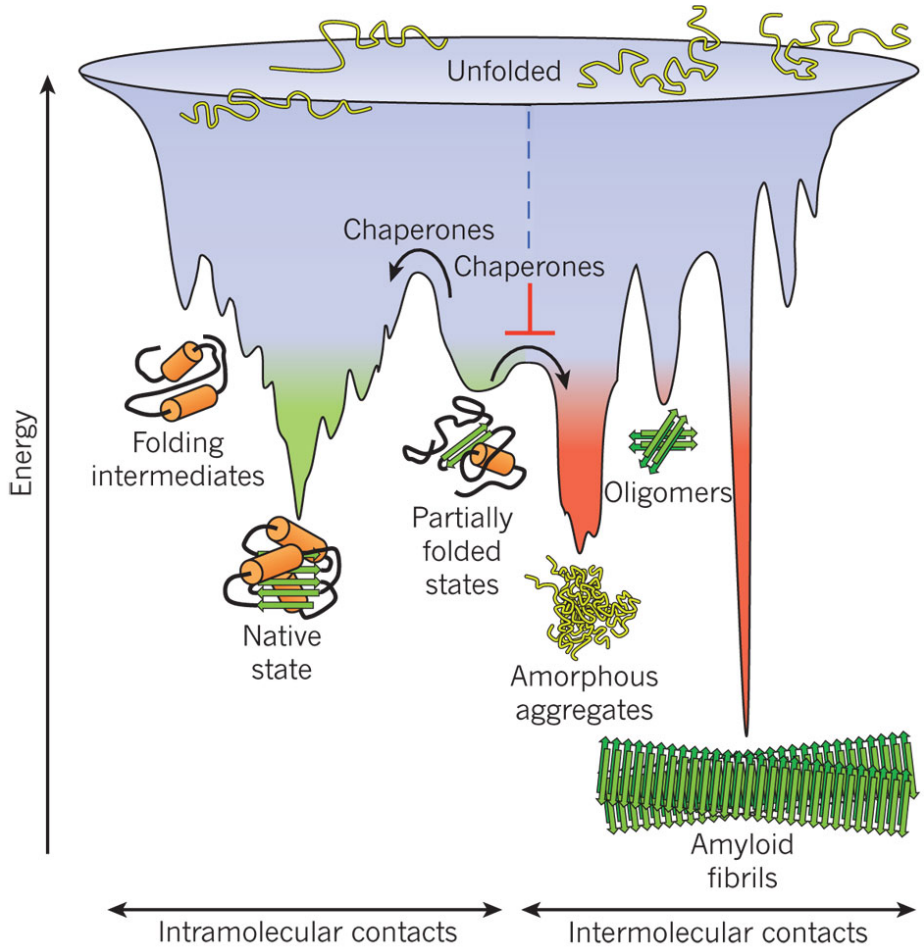
HIV Reverse Transcriptase Inhibitor





Reverse Transcriptase
(responsible for replication of the virus)

Protein Energy Landscape



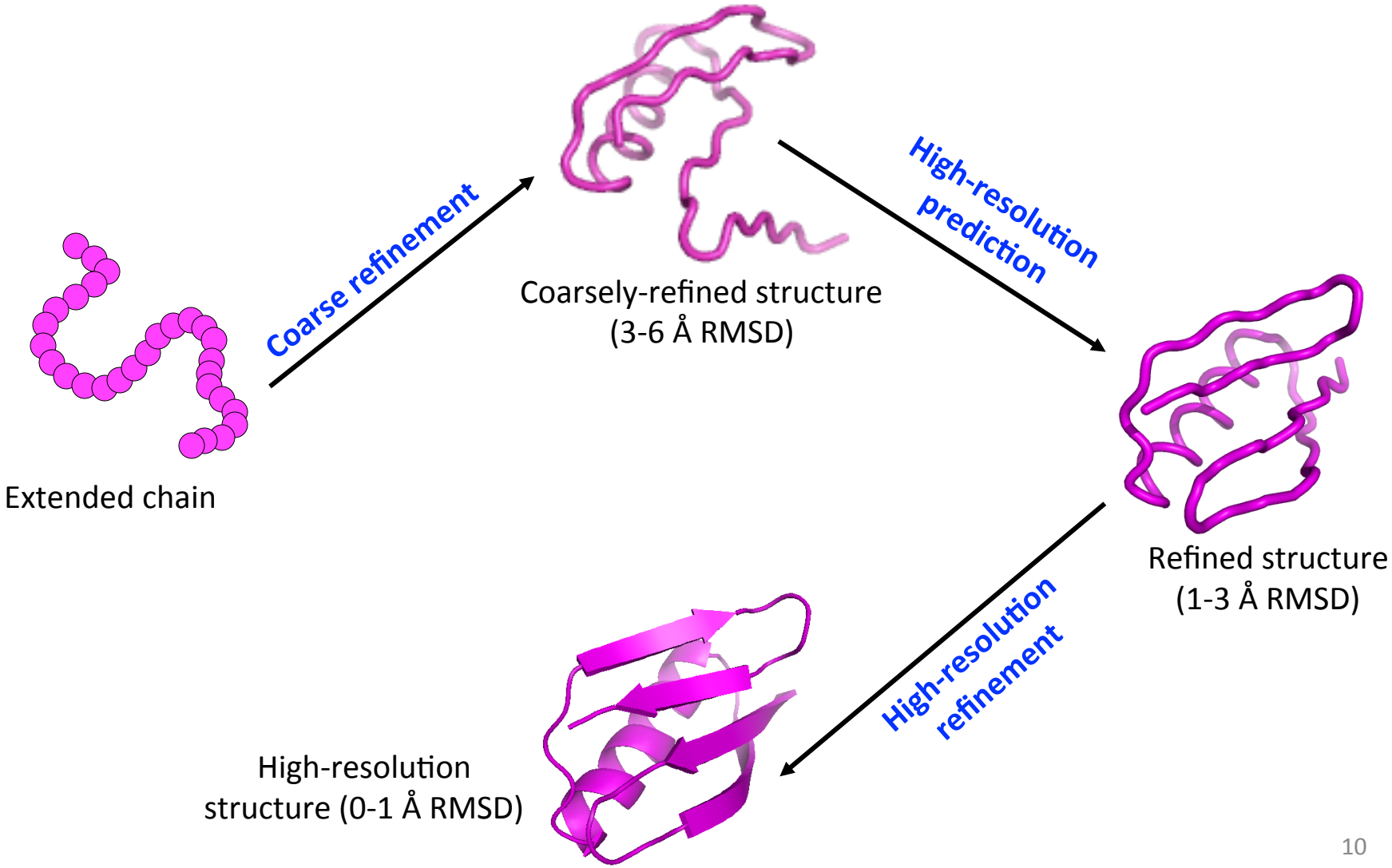
F. Ulrich Hart, Andreas Bracher & Manajit Hayer-Hartl, Nature, 475, 324-332, 2011

Modeling Proteins

- Energy
  - Quantum Mechanics (Schrodinger, DFT, HF, …)
  - Molecular Mechanics (Coulomb's law, Van der waals, Hydrogen bond, …)
  - Continuum Mechanics (Poisson Boltzmann, …)

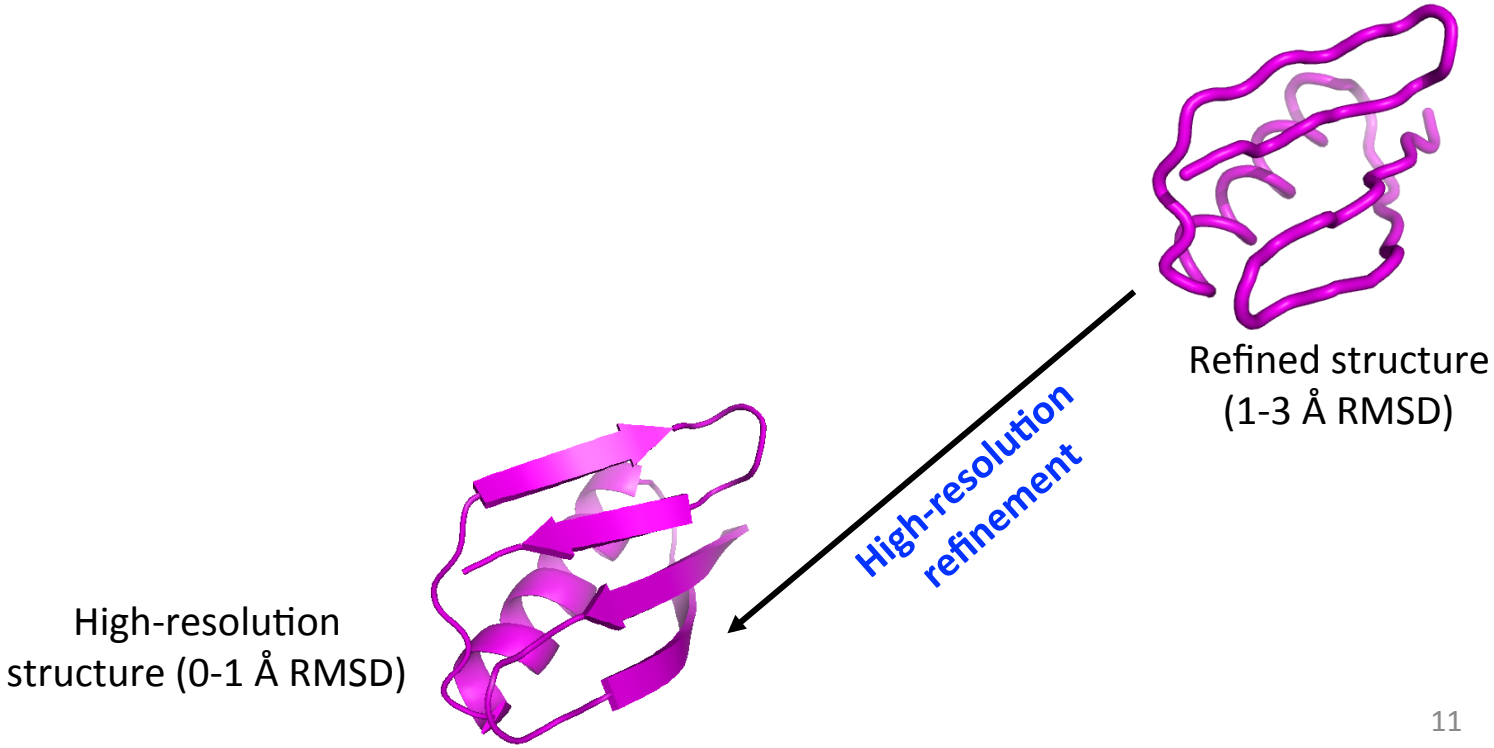
- Conformation
  - Dihedral space
  - Planar angels
  - Side chain conformations
  - three dimensional configuration of atoms

Protein Structure Prediction



High-resolution Protein Structure Refinement

- Accurate protein structure is needed for drug design
- Bottle-neck in protein structure prediction

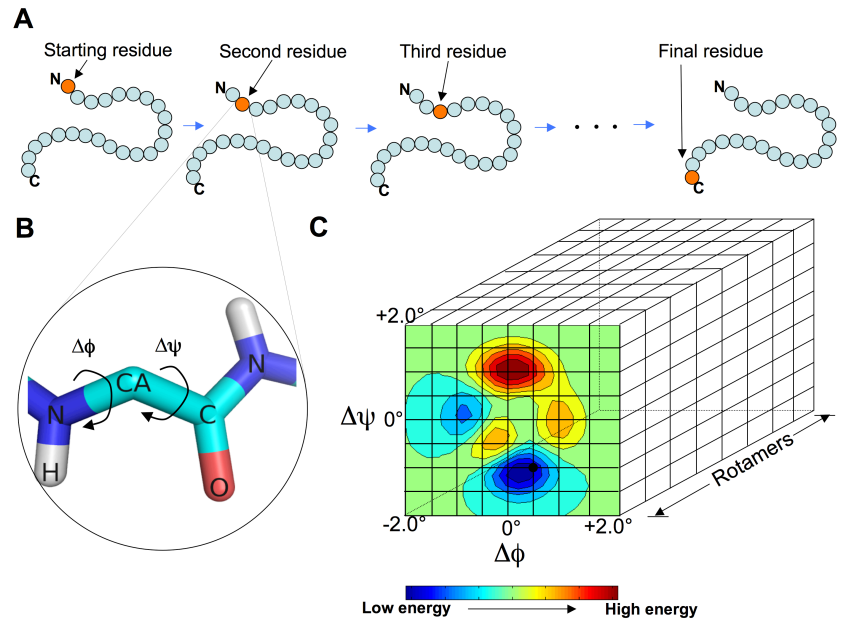


Replica Exchange Monte Carlo GRID

1. Randomly choose amino acid

2. Randomly pick a point on the GRID

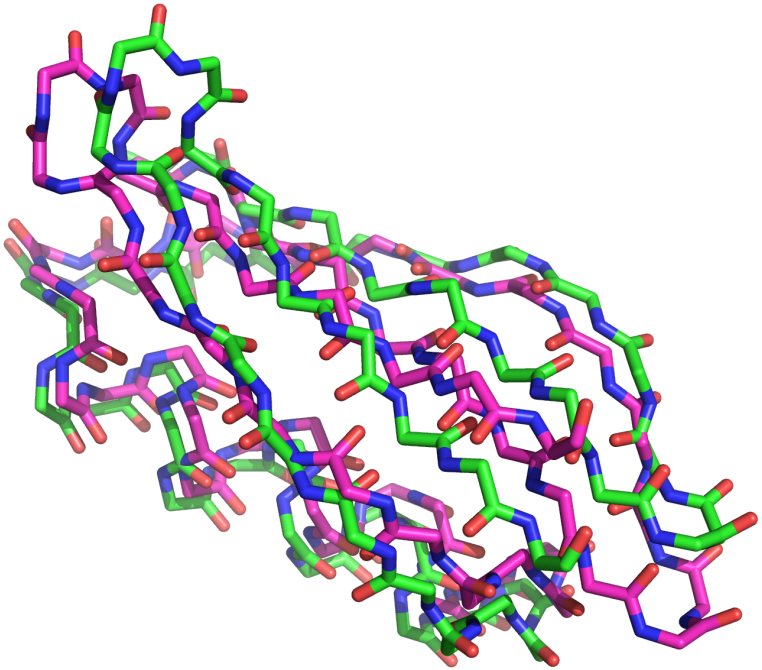
3. Find the lowest energy side-chain

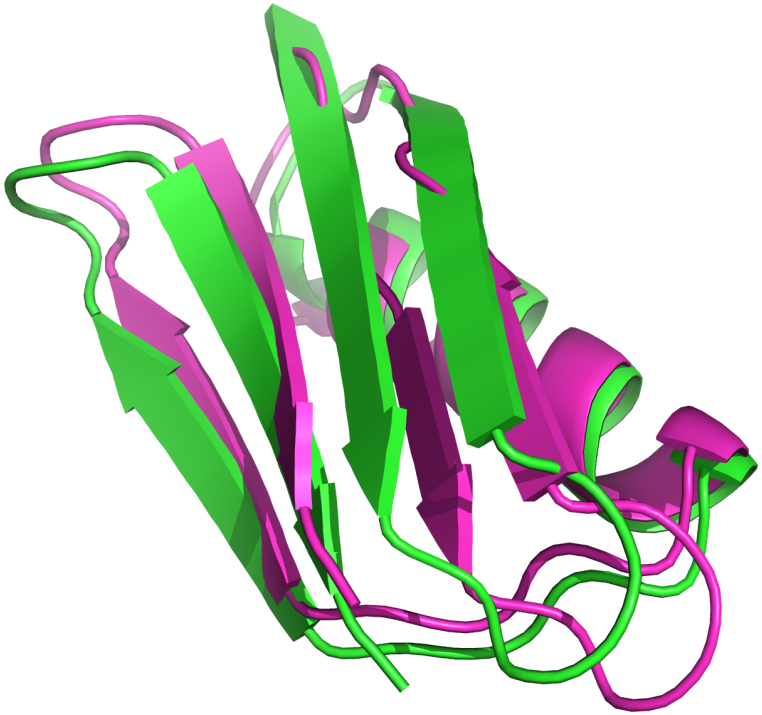


4. Based on metropolis acceptance criterion accept or reject

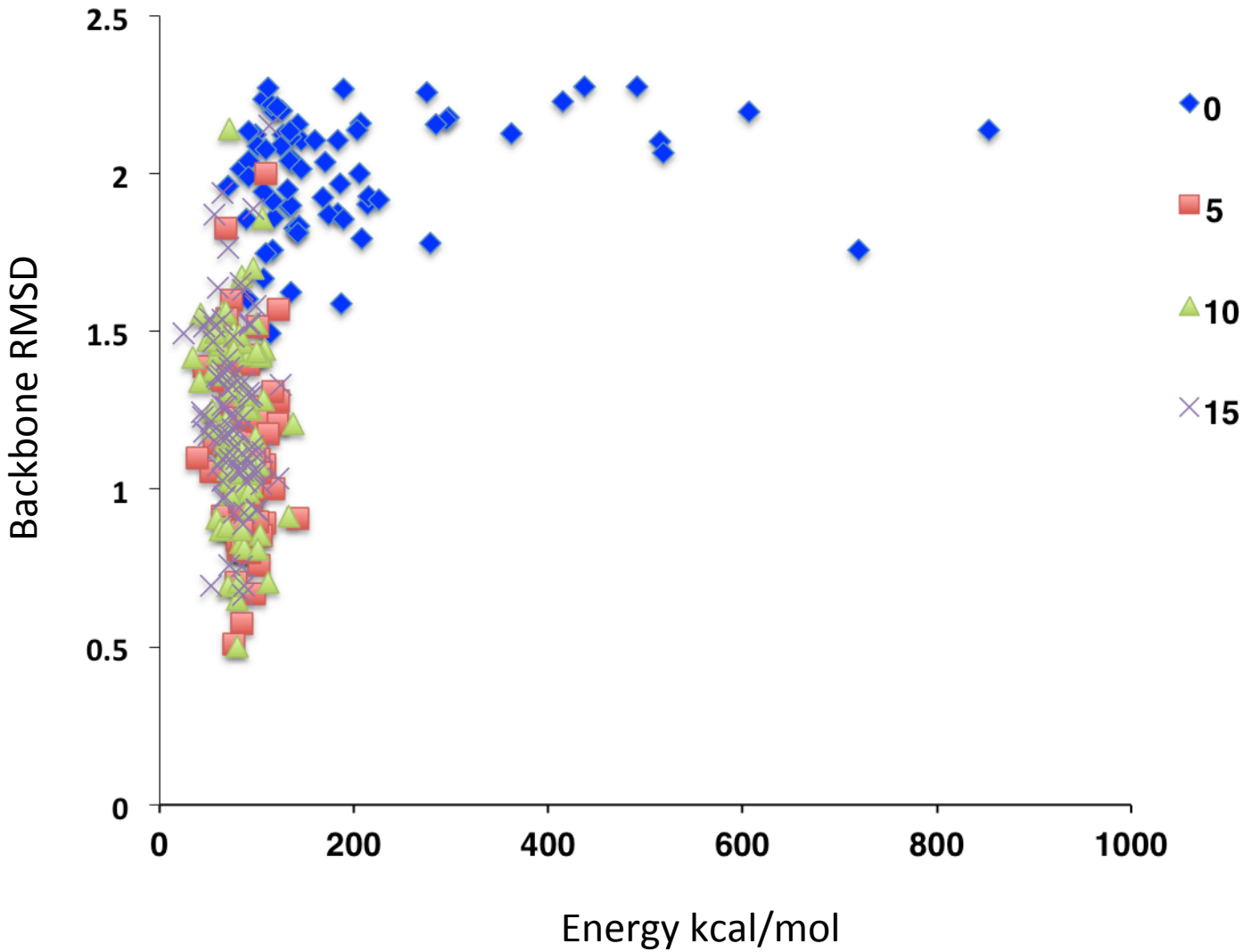
5. Every several steps swap configurations based on their temperatures

Disrupted Protein Structure

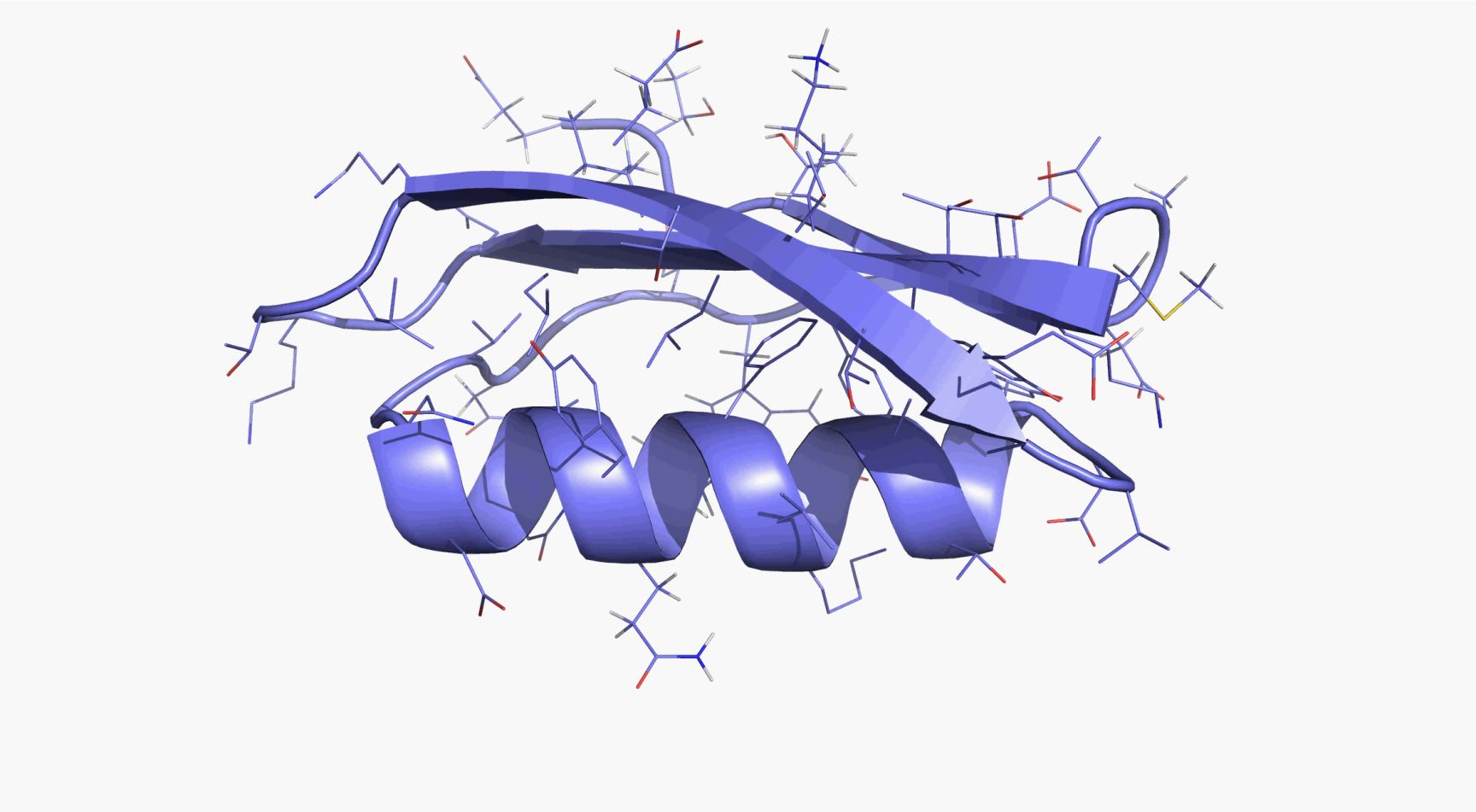




Refined Structure



Replica Exchange Monte Carlo GRID



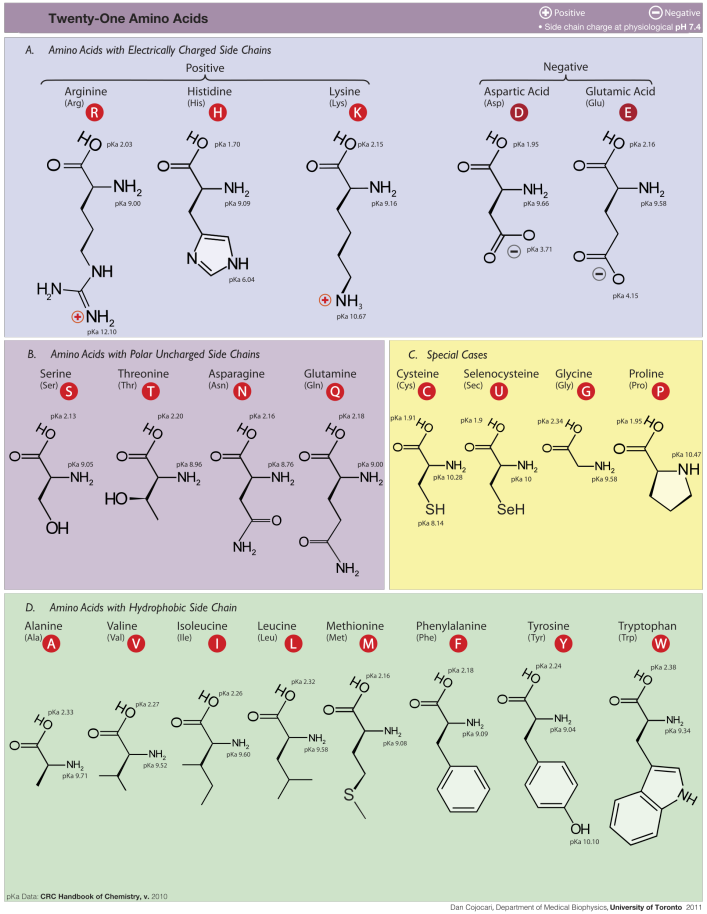
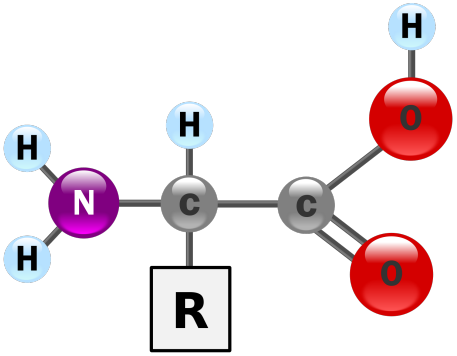
Parallelism Scheme

- Each individual trajectory runs on a single core
- Processes communicate every 200 steps
- Efficiency of parallelism is about 98%

Future works

- Benchmark the algorithm and show more the efficacy of the algorithm
- Turn the algorithm into computational protein design algorithm
- Use the algorithm with coarser resolution energy force-fields and use it for full protein structure prediction

Thanks for listening!

Questions

Protein Building Blocks (Amino Acids)