Structural biology: What for?

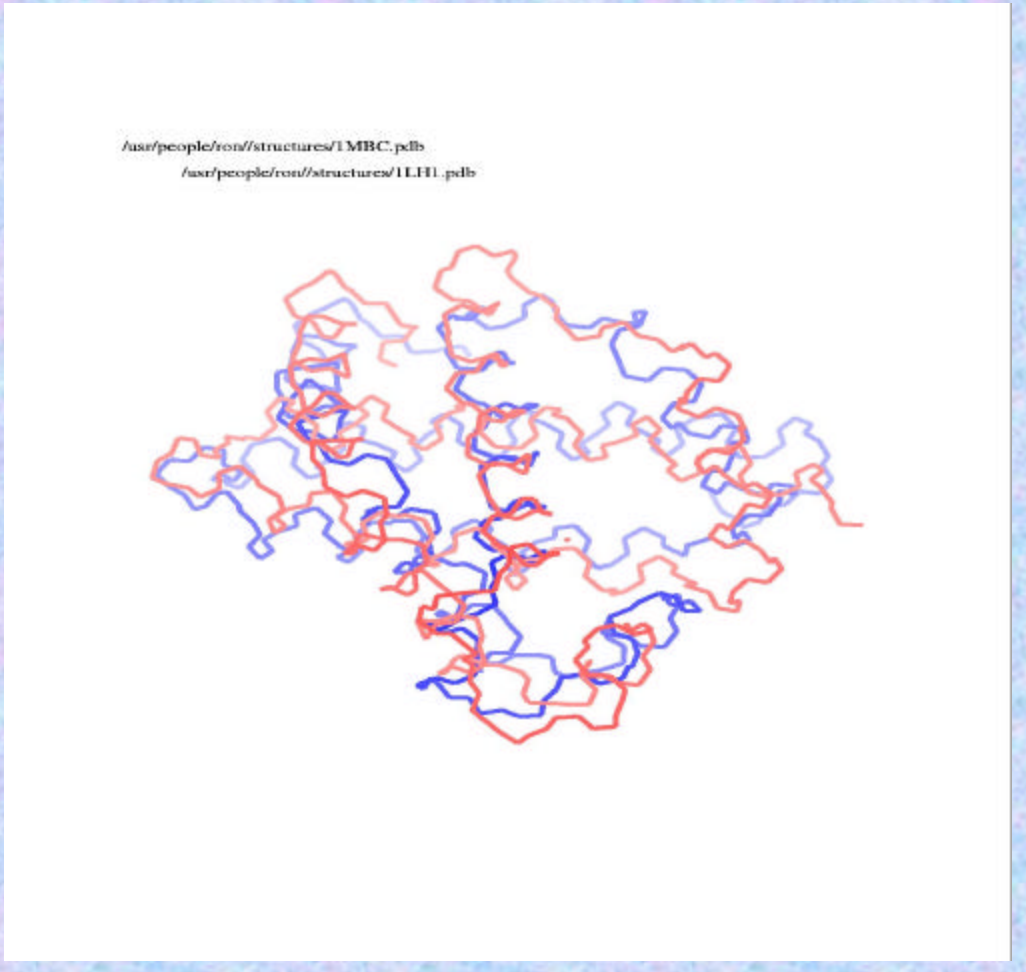
- Detecting remote evolutionary relationships by matching sequences to structures.
- Understanding evolutionary mechanisms (e.g., domain swaps)
- Identifying active sites
- Understanding protein protein interactions
- Rational design of protein structures functions and drugs

Studies of evolution

- Structure are conserved a lot better than sequences (proteins with less than 20 percent sequence identity may still share remarkably similar structures, e.g. myoglobin and hemoglobin)
  - Better detection
  - More meaningful evolutionary trees

Comparing Protein Structures (myoglobin and leghemoglobin)

Plant Versus mammalian protein



Mechanism of evolution

- Rather than single point mutation domain swaps
  - Identification of domains, candidates for domain swaps, better done on structures
  - Identify sites more accessible for mutations (loops)
  - Identify interesting sites: Active site, stabilizing contacts

Protein-protein interactions

- Identify candidates
- Mechanism of interactions:
  - What are the critical residues?
  - Strength of interactions
  - Competition with other proteins
  - prions

Design of protein structure and drugs

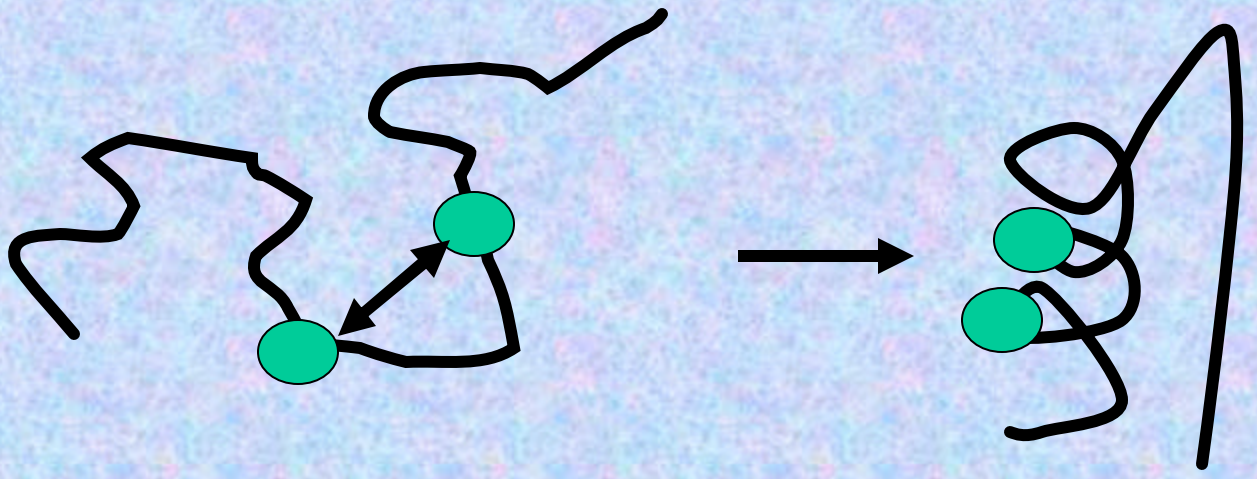
- Learn principles of protein structures
- Understand basic chemical physics principles of protein structure and protein interactions
- A starting point to learn about protein function and dynamic. Protein activation and conformational transitions.

Approaches to determine protein structures

- Experiments: X-ray crystallography, NMR, Clearly most accurate, can take months/years and sometimes fail.
- Homology (based on an assignment to a family) most accurate computationally, based on the availability of an evolutionary related protein, fast.
- Ab-initio folding, solve the protein folding problem. Based on sequence and chemical physics principles determine the three dimensional shape. General. Slow and less accurate.

Ab-initio folding in time

- Start from a random conformation of the chain and mimic the chemical physics of the process. Kinetic of folding (Blue Gene)



Ab-initio folding as a global optimization problem

- Anfinsen hypothesis: Proteins have a unique fold which is the thermodynamically most stable.
- Find the global free energy minimum: the most stable state of a thermodynamic system

Challenges in ab-initio folding

- The interaction potential (atomic or residue level) is not sufficiently accurate
- The cost of computing the correct fold is extremely high
- Overall, nature finds the correct fold in about 10^7 torsion transitions for a protein of length 100. Remarkably small considering the number of possible states 10^{100}. Can we mimic nature?

Potential advantages of ab-initio folding

- Predicting structures of novel folds.
- Chemical physics level understanding of intra-protein interactions. Useful for protein design. Useful for better understanding of interactions between protein and with non-peptide molecules (drugs)

Homology modeling

- Two independent steps
  - Identify a relative (sequence alignment, HMM, threading) and highly approximate alignment
  - Build a model based on the alignment

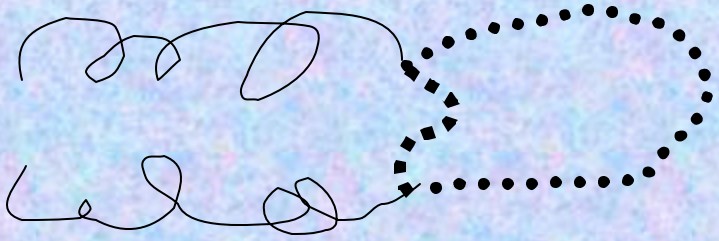
Assignment

- Must include a library as large as possible for protein with known structures.
  - The Protein Data Bank (PDB) includes about 18K structures compared to known 600K sequences... Lucky for us many sequences fold to a common structure. Only 500 structure families.
  - The annotation can and should use whatever information we have: sequence, activity, side chains, fitness to structure (threading)

Homology

- Once am alignment is provided need to build side chains (relatively simple essentially solved problem). Limited number of rotamers (conformation of a side chain attached to a site) that are determined locally

- Determine pieces of structure with no alignments (gaps). Most interesting missing parts are in loops. Still an open problem



Advantages

- Similar structures (within RMS of about 4 angstrom) can be modeled accurately
- Quick. Genome annotation can be done quickly and accurately. About 1/3 is annotated

Disadvantages

- Requires an extensive library of structures (we clearly miss some). So far we have much much less structures compared to sequences
- Accuracy of predictions is uncertain

Summary: Computational Studies of Protein Structures

- Ab-initio: Use physical chemistry principle to predict protein structures (very general and fundamental, expensive)
- Homology: Use existing database of proteins to create a knowledge base procedure to predict proteins shapes (reasonably accurate, fast, depends on the quality of the library, and scoring function)

- (there are also mixed approaches)