Statistical and Computational Issues in Ab Initio Protein Structure Prediction

Ingo Ruczinski

Department of Biostatistics
Johns Hopkins University

Email: ingo@jhu.edu

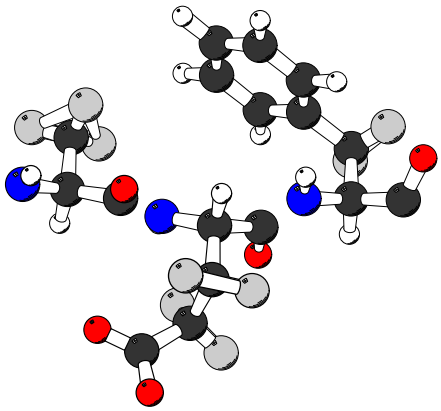
http://biostat.jhsph.edu/∼iruczins

Collaborators

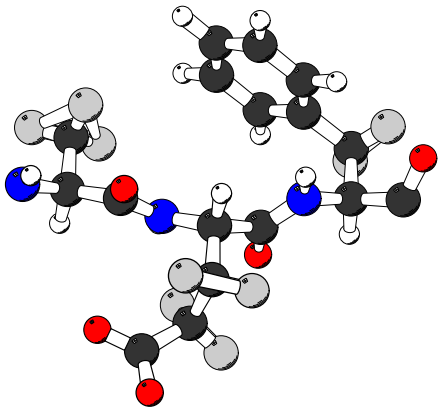
David Baker	University of Washington
Richard Bonneau	Institute for Systems Biology
Chris Bystroff	Rensselaer Polytechnic Institute
Charles Kooperberg	Fred Hutchinson Cancer Research Ctr
Carol Rohl	University of Washington
Kim Simons	Harvard University
Charlie Strauss	Los Alamos National Laboratory
Jerry Tsai	Texas A&M

What are Proteins?

Without peptide bonds

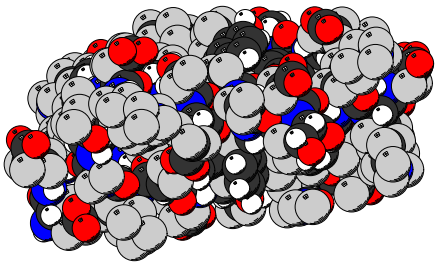


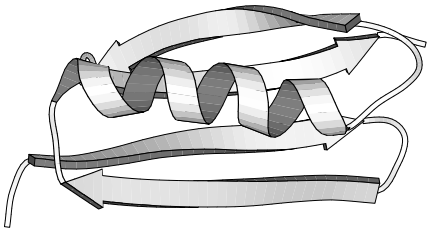
With peptide bonds



The building blocks of proteins are amino acids.

2D and 3D Protein Structure

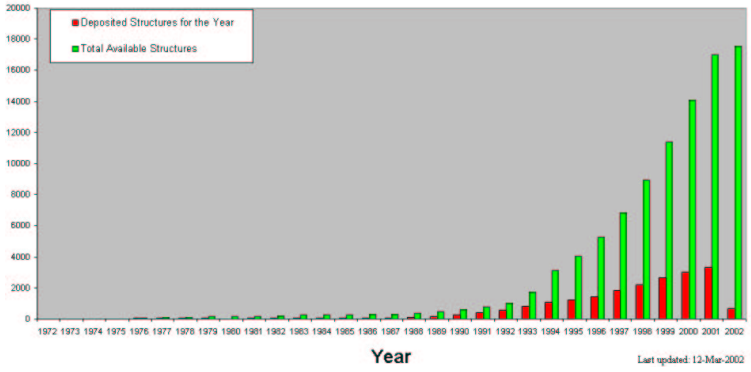




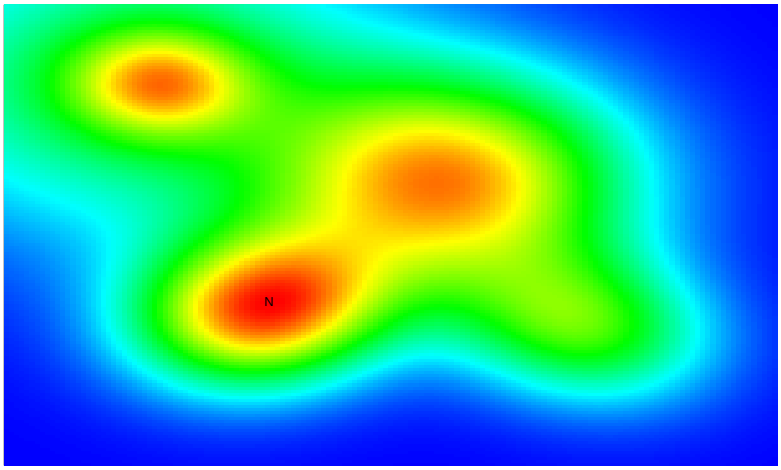
Both figures show the same protein, highlighting the tertiary and secondary structure.

Motivation

- What are proteins? Why do we care about them?
- Why do we care about protein structure?
- Why do we need to predict protein structures?
- How does the computational approach work?



Energy Landscape



The free energy of a structure changes with its geometry.

A Scoring Function for Ab Initio Protein Folding

$$P(\text{structure}|\text{sequence}) \propto P(\text{sequence}|\text{structure}) \times P(\text{structure})$$

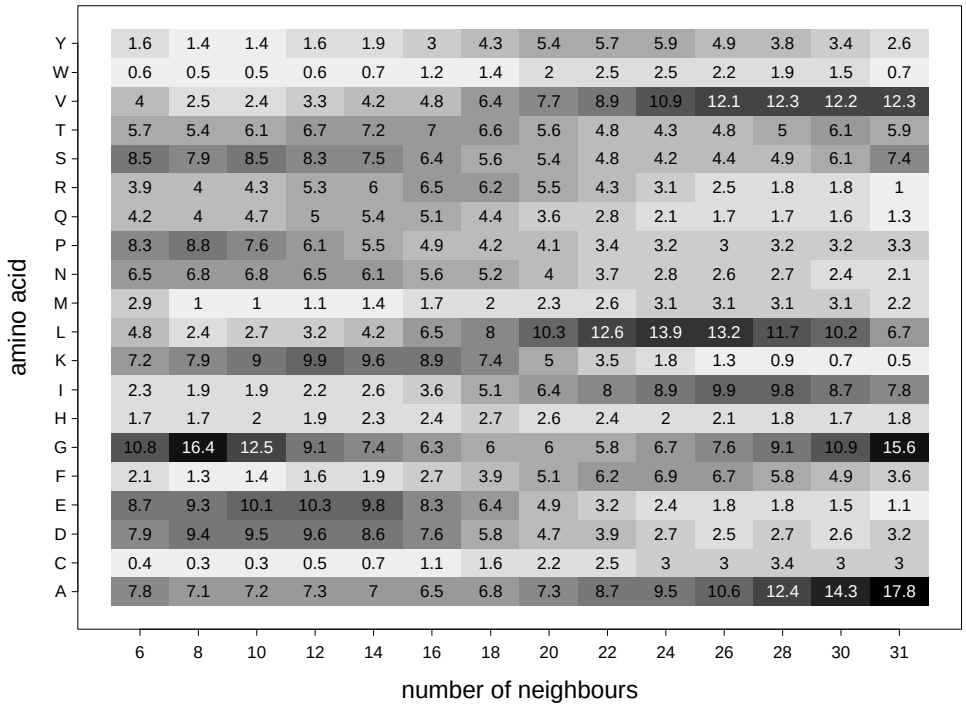
Sequence dependent:

- hydrophobic burial
- residue pair interaction

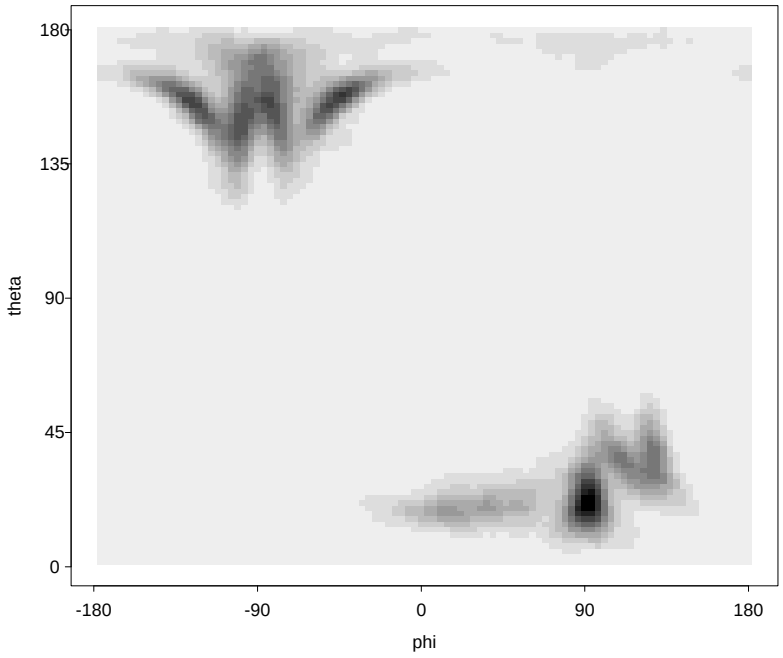
Sequence independent:

- helix-strand packing
- strand-strand packing
- sheet configurations
- vdW interactions

Hydrophobic Burial

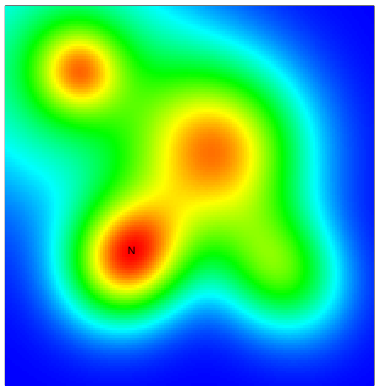


Strand-strand Interaction

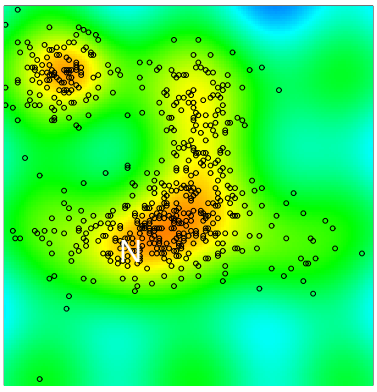


Energy Landscape (2)

True landscape

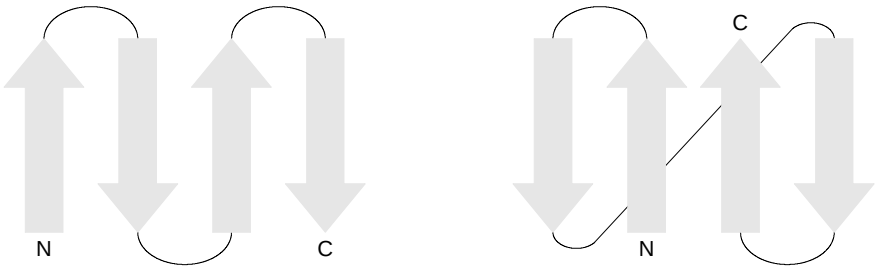


Our scoring function



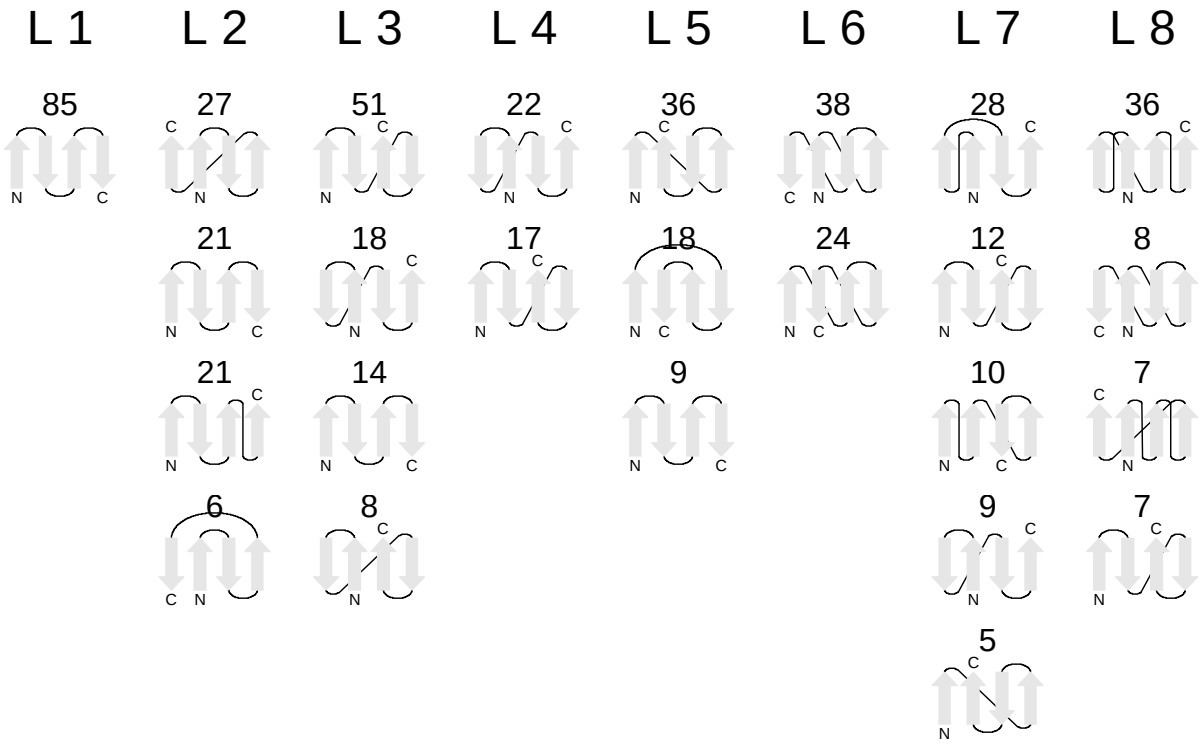
Decoys tend to cluster near low energy states

Beta-Sheet Motifs

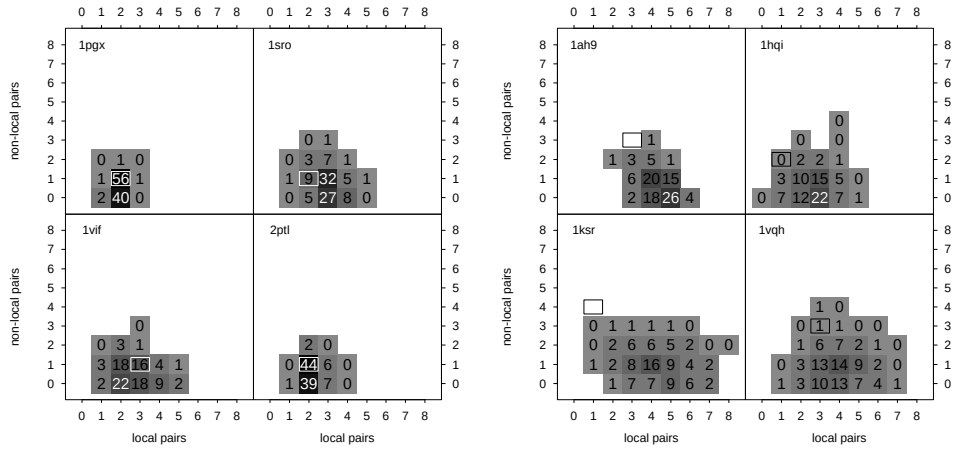


Two possible motifs for 4-stranded sheets.

Likely Sheet Topologies

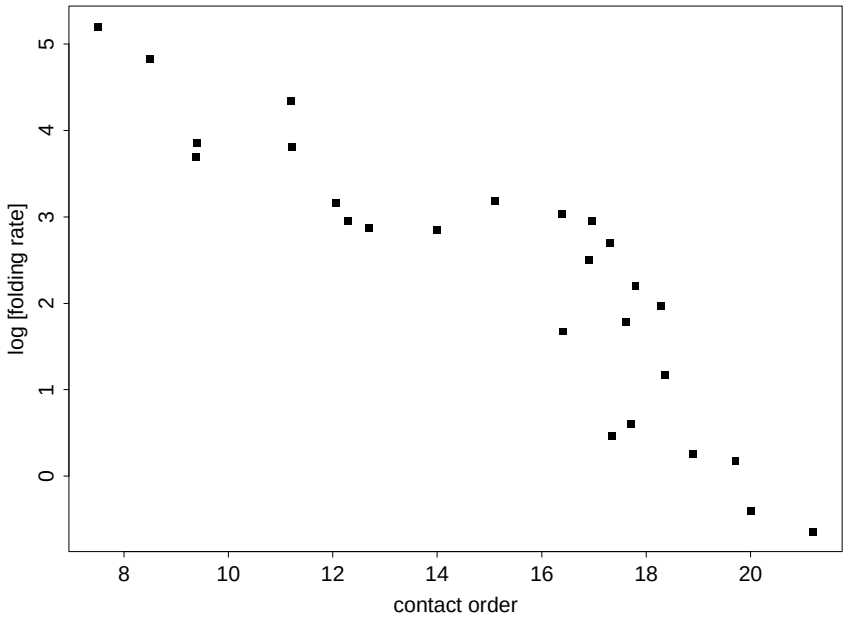


Bias towards Local Conformations

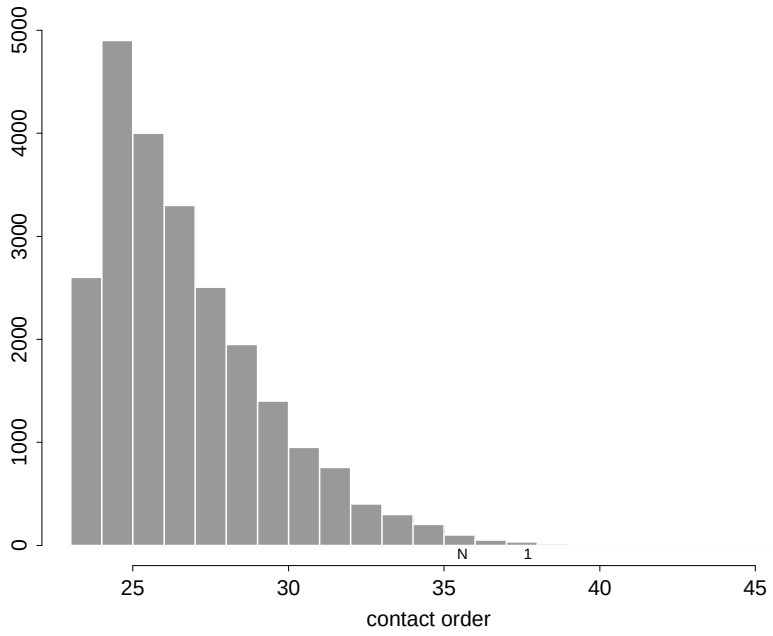


Local structures are easier to generate than non-local structures.

Contact Order and Folding Rate



Contact Order Distribution



Using All Atom Information

The Lennard-Jones potential for atoms i and j is given by the equation

$$E = \left[\left(\frac{R_i + R_j}{d_{ij}} \right)^{12} - 2 \left(\frac{R_i + R_j}{d_{ij}} \right)^{6} \right] \sqrt{\epsilon_i \epsilon_j}.$$

R_i and R_j are the van der Waals radii, ϵ_i and ϵ_j are the well depths for atoms i and j, and d_{ij} is the distance between atoms i and j.

Then

$$\prec f(aa|\mathbf{E}) \succ := \int_{\mathbf{E}_{-0}} P(aa, \mathbf{E}) f(aa|\mathbf{E}) d(\mathbf{E}_{-0})$$

might be helpful in obtaining additional information to the environment term.