Ab initio protein structure prediction: Progress and Prospects – Bonneau, Baker


**Abstract:** Homology-based comparative modeling, Fold recognition and Ab initio protein structure prediction are some of the popular methods for protein structure prediction. This paper reviews features of current ab initio methods like reduced complexity models, scoring functions and relationship to true protein folding and discusses potential applications. It also briefly discusses Rosetta method which is one of the successful ab initio structure prediction methods.

**Discussion:** The goal of ab initio structure prediction is: given a protein’s amino acid sequence predict the structure of its native state. This problem can be decomposed into 2 sub problems: a. developing accurate potential and b. developing an efficient protocol for searching the resultant energy landscape. This paper focuses on methods for predicting tertiary structure in absence of homology to a known structure and discusses local prediction methods only in context of tertiary structure prediction.

To overcome computational intractability, reduced complexity models like lattice and off-lattice methods are used. **Lattice models** have a restricted ability to model subtle geometric considerations. But, the computational advantages of these models outweigh the problems associated with systematic biases. **Off Lattice Models** containing 4 to 32 $\varphi/\psi$ states can represent various strand, helix and loop conformations. Optimized models can reproduce native contacts, preserve secondary structure and fit overall coordinates of the native state. We can use **local structure prediction** to narrow our search. Local structures excised from proteins can fold independent of the full protein. Majority of the methods proving successful at CASP (a conference for evaluating protein folding methods) use secondary structure prediction.

**The energy (or scoring) function** must adequately represent forces responsible for protein structure: solvation, strand, hydrogen bonding etc. Any energy function designed to work with low complexity models must be robust to systematic errors and must be computationally efficient. **Solvation-based scoring** schemes favors placement of hydrophobic amino acids at buried positions and of hydrophilic acids at exposed positions. These simple and efficient scoring schemes are useful in predicting small hydrophobic cores. Most low-resolution potential-scoring functions utilize an empirically derived **pair potential**. The most common of these potentials are functions of position of a single center per residue ($C_{n} C_{g}$ or centroid/united atom center). These functions are computationally efficient. Many features of proteins, such as association of beta strands into sheets, can be described by **sequence independent scoring functions**. Homologous sequences available for protein families can be used by doing **multiple sequence alignments** on these sequences followed by making contact (between amino acid side chains) prediction based on covariance patterns in these alignments. **High resolution potential methods** must be improved in order for ab initio methods to improve. Some progress has been made in modeling difficult terms like solvation and electrostatics.

Low resolution methods have been successful in accounting for both protein folding rates and distribution of structure in transition state from knowledge of final folded state. Annotation of Open Reading Frames lacking sequence homology to proteins of known functions is one of the potential uses for ab initio prediction. Ab initio models are well suited for adding missing regions to homology models, thereby producing much more complete sets of models. Accurate structures can be produced by using an ab initio method in conjunction with NMR chemical shift data and sparse NMR constraints.

In **Rosetta method**, tertiary structures are generated by using a Monte Carlo search of possible combinations of likely local structures, minimizing a scoring function that accounts for nonlocal interaction. This method produces structures with hydrophobic residue, paired beta strands and other protein like features that are consistent with local sequence-structure biases.

Ab initio methods are making progress from being an area of academic interest to that of practical relevance.