

## Protein Modeling Methods

BINF 731

# Protein Structure Analysis

Iosif Vaisman

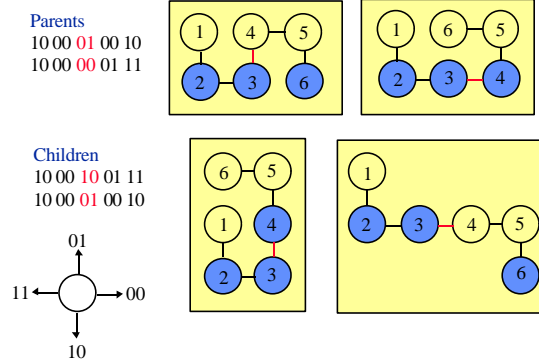
2002

- **Ab initio methods:**  
solution of a protein folding problem  
search in conformational space
- **Energy-based methods:**  
energy minimization  
molecular simulation
- **Knowledge-based methods:**  
homology modeling  
fold recognition

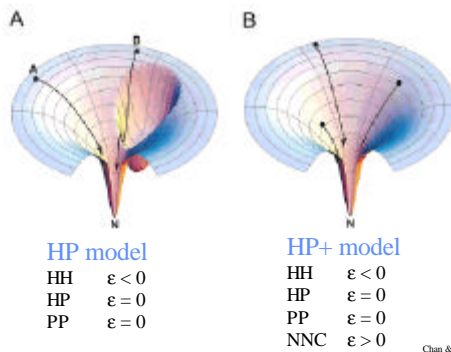
### Ab initio Methods

- Simplified models
  - simplified alphabet (HP)
  - simplified representation (lattice)
- Build-up techniques
- Deterministic methods
  - quantum mechanics
  - diffusion equations
  - DFT
- Stochastic searches
  - Monte Carlo
  - genetic algorithms

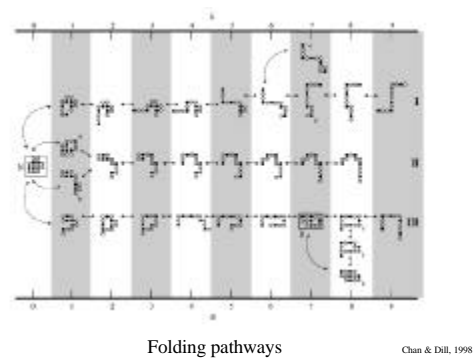
### Genetic Algorithms Applications



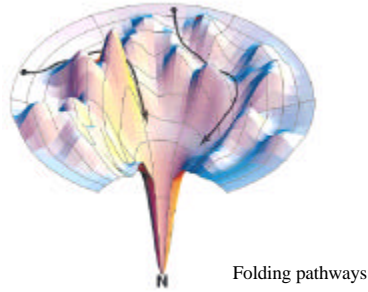
### HP Lattice Models



### HP Lattice Models

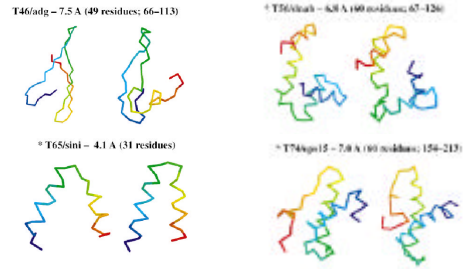


## HP Lattice Models



Chan & Dill, 1998

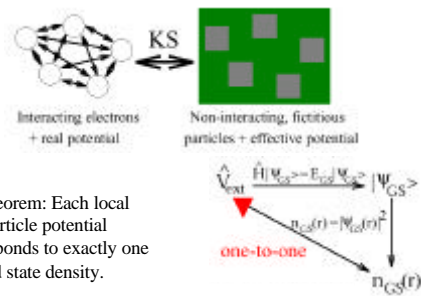
## Hierarchical *ab initio* prediction



Lattice models  
Knowledge-based scoring functions

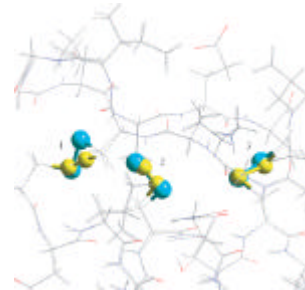
Samudrala et al., 1999

## Density Functional Theory



Adapted from Willfried Aulrich, OSU

## Density Functional Theory



DFT optimization of NMR structure (1PNH)

Andreoni et al., 1999

## Protein Modeling Methods

- **Ab initio methods:**  
solution of a protein folding problem  
search in conformational space
- **Energy-based methods:**  
energy minimization  
molecular simulation
- **Knowledge-based methods:**  
homology modeling  
fold recognition

## Knowledge

Knowledge is a pattern that exceeds certain threshold of interestingness.

- Factors that contribute to interestingness:
- coverage
  - confidence
  - statistical significance
  - simplicity
  - unexpectedness
  - actionability

## Knowledge-based methods

Finding patterns in known structures

Deriving rules (usually in the form of PMF)

Applying the rules

## Fold Recognition

Pattern searching

sequence patterns

structure patterns

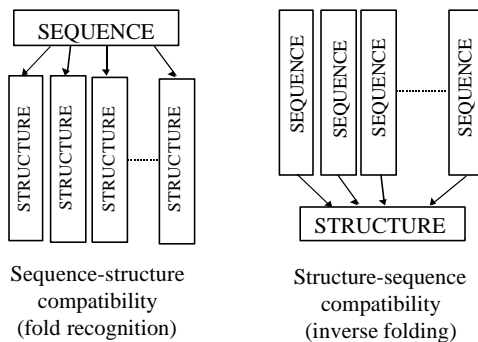
residue composition patterns

Threading

sequence-structure compatibility

structure-sequence compatibility

## Threading



## Threading

- Only the local environment is taken into account
- Non-local contacts are assumed with generic peptide
- No gaps are allowed in the alignment

## Threading Programs

THREADER

(<http://insulin.brunel.ac.uk/threader/threader.html>)

NCBI Threading

(<http://www.ncbi.nlm.nih.gov/Structure/RESEARCH/threading.shtml>)

## Homology Modeling

- Identification of structurally conserved regions (using multiple alignment)
- Backbone construction (based on SCR)
- Loop construction (KB or conformational search)
- Side-chain restoration (KB, rotamer, or MM)
- Structure verification and evaluation
- Structure refinement (energy minimization)

## Homology Modeling Programs

Modeller

(<http://guitar.rockefeller.edu/modeller>)

Swiss-Model

(<http://www.expasy.ch/swissmod>)

Whatif

(<http://www.cmbi.kun.nl/whatif>)

## Swiss-Model

- **Method:**  
Knowledge-based approach.
- **Requirements:**  
At least one known 3D-structure of a related protein.  
Good quality sequence alignments.
- **Procedures:**  
Superposition of related 3D-structures.  
Generation of a multiple alignment.  
Generation of a framework for the new sequence.  
Rebuild lacking loops.  
Complete and correct backbone.  
Correct and rebuild side chains.  
Verify model structure quality and check packing.  
Refine structure by energy minimisation and molecular dynamics.

## Swiss-Model Request Types

- **First Approach mode.**
- **Optimise mode.**
- **Combine mode.**
- **GPCR mode.**

## Methods and Programs used by Swiss-Model

- **Sequence Alignment**  
BLAST (Altschul S.F., *J. Mol. Biol.* **215**:403, 1990)  
SIM (Huang, X., Miller, M. *Adv. Appl. Math.* **12**:337, 1991)  
ProModII (Peitsch, M.C. *Unpublished*, Server-specific tool)
- **Knowledge Based Protein Modelling**  
ProMod (Peitsch M.C. *Biochem Soc Trans* **24**:274, 1996)
- **Energy Minimisation**  
Gromos96 (van Gunsteren W.F. <http://igc.ethz.ch/gromos/>)
- **Model evaluation**  
Swiss-PdbViewer  
(<http://www.expasy.ch/spdbv/mainpage.html>)

## Model Confidence Factors

The Model B-factors are determined as follows:

- The number of template structures used for model building.
- The deviation of the model from the template structures.
- The Distance trap value used for framework building.

The Model B-factor is computed as:

$85.0 * (1 / \# \text{ selected template str.}) * (\text{Distance trap} / 2.5)$

and

99.9 for all atoms added during loop and side-chain building