# Solving Protein Structure by Gamers Playing Foldit

PhD. Student: **Haokui ZHOU**
Supervisor: **Prof. Guoping ZHAO**

Clinical Microbial Genomics Laboratory

Department of Microbiology

The Chinese University of Hong Kong

What is a protein?
sequence -> structure -> function

- A chain (sequence) of amino acids
  - made up of 20 amino acids
  - main chain & side chain
- Folds into 3D structure (shape)
  - folding of the backbone (main chain)
  - 3D arrangement of the side chains
- Biological functions
  - Binding to substrates (receptor, ligand, …)
  - Catalyzing chemical reactions

Structure determines function

Substrate
Active site
Binding
Products

Enzyme@wiki
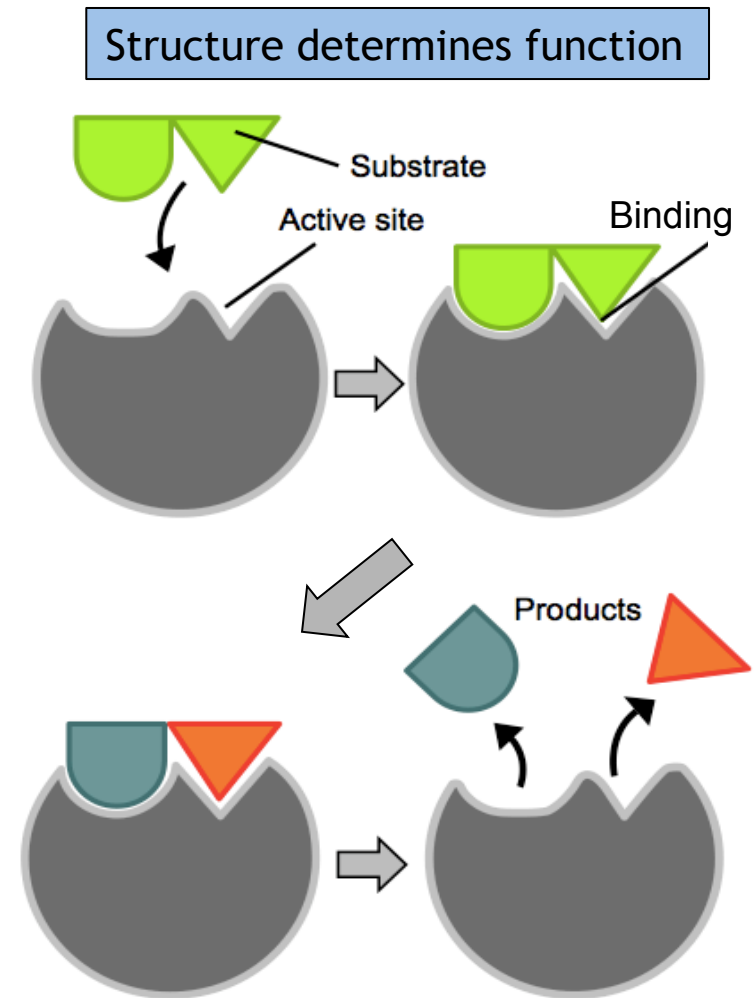
Hierarchy of protein structure

Primary structure
(amino acid sequence)

↓

Secondary structure
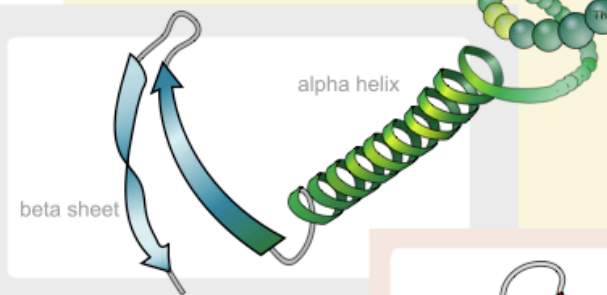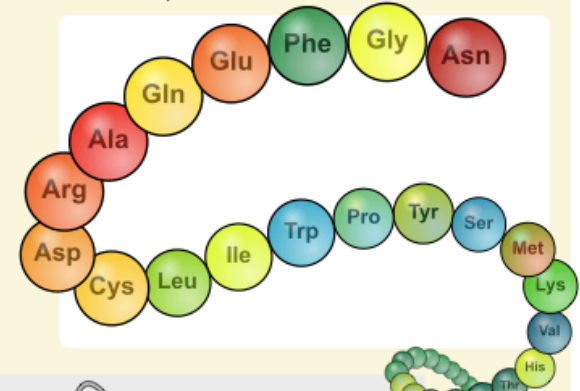(α-helix, ß-sheet, loop, ...)

↓

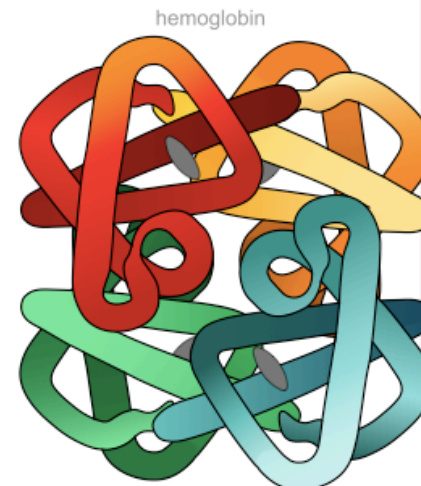**Tertiary structure**
(3D assembly of secondary structures)

↓

Quaternary structure
(Complex structures of multiple subunits)

Primary structure
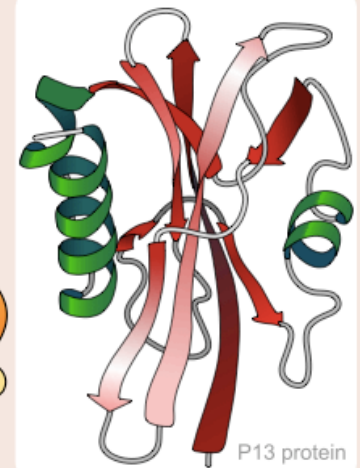amino acid sequence

Phe Gly Asn Glu Gln Ala Arg Asp Cys Leu Ile Trp Pro Tyr Ser Met Lys Val His Thr

alpha helix

beta sheet

Secondary structure
regular sub-structures

hemoglobin

Tertiary structure
three-dimensional structure

P13 protein

Quaternary structure
complex of protein molecules

Protein_structure@wiki

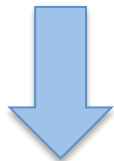# Protein 3D structure determination (physical methods)

- X-ray crystallography
  - x-rays pass through a crystal of protein
  - produce a diffraction pattern
  - calculate the electron density map
  - decipher the molecular bonds
- Nuclear Magnetic Resonance (NMR)
  - a solution of protein is placed in a magnetic field
  - determine frequencies on the resonance of atoms in the protein
- Cryo-electron microscopy (frozen)
  - a form of transmission electron microscopy (EM)
  - sample is studied at cryogenic temperatures (liquid nitrogen temperatures)

Protein 3D structure prediction (computational methods)
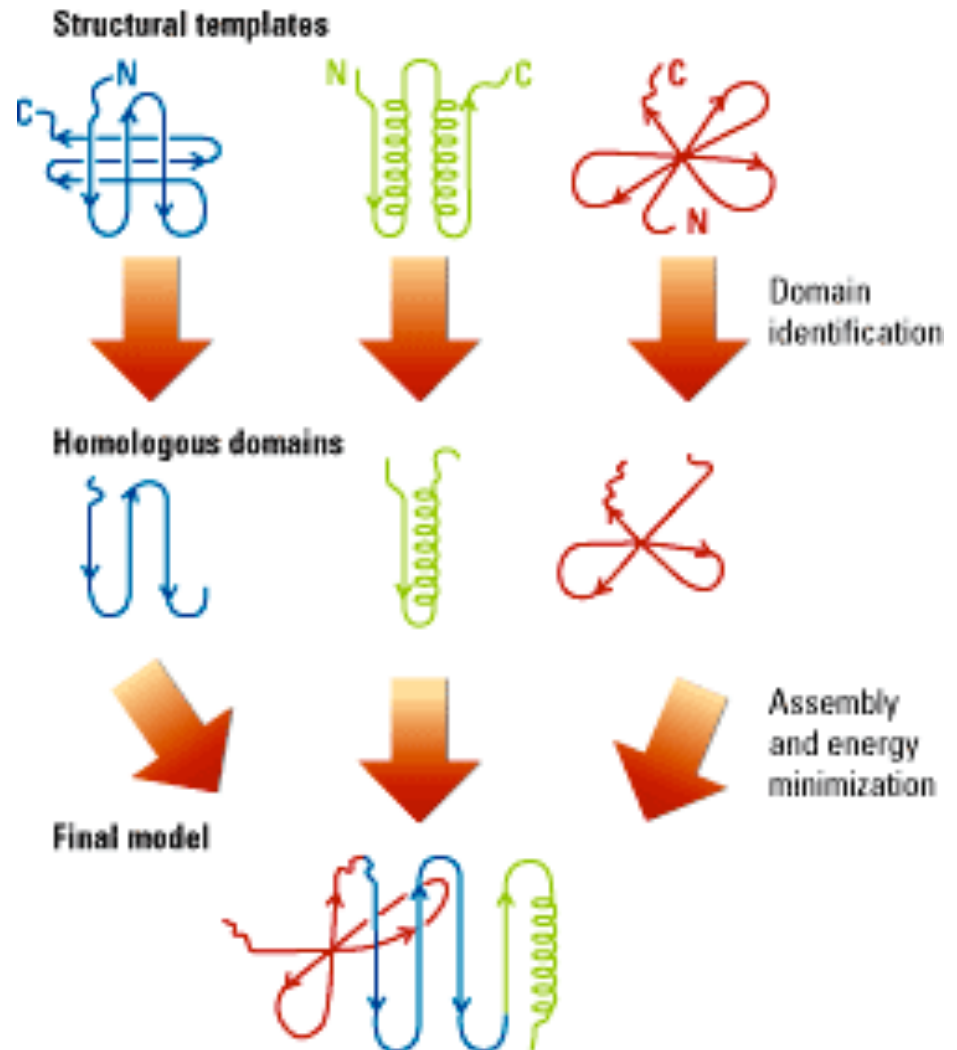
- *Ab initio* modeling
  - start from primary protein sequence
  - find protein conformation with minimized free energy
  - a big challenge of computational biology
- Comparative modeling
  - start from homologues protein with known structure
  - map similar sequence regions to identify similar structures and refinements
- Fold recognition
  - thread sequence onto template structure with folds

The state-of-the-art Rosetta methods

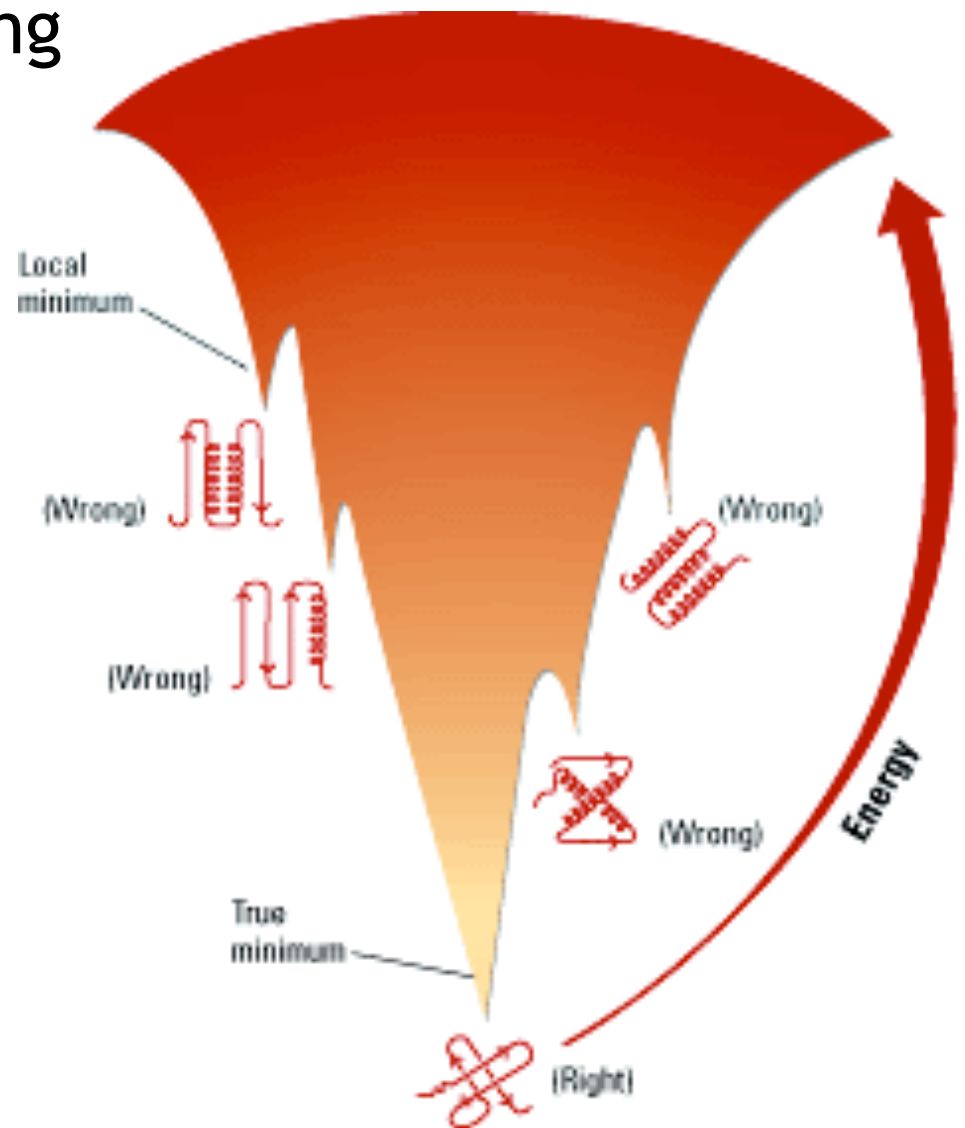**Predicts local folds**
using threading or multiple alignments

Links domains
using energy functions (hydrophobic burial, hydrogen bonding, and electrostatics)

Structural templates

Domain identification

Homologous domains

Assembly and energy minimization

Final model

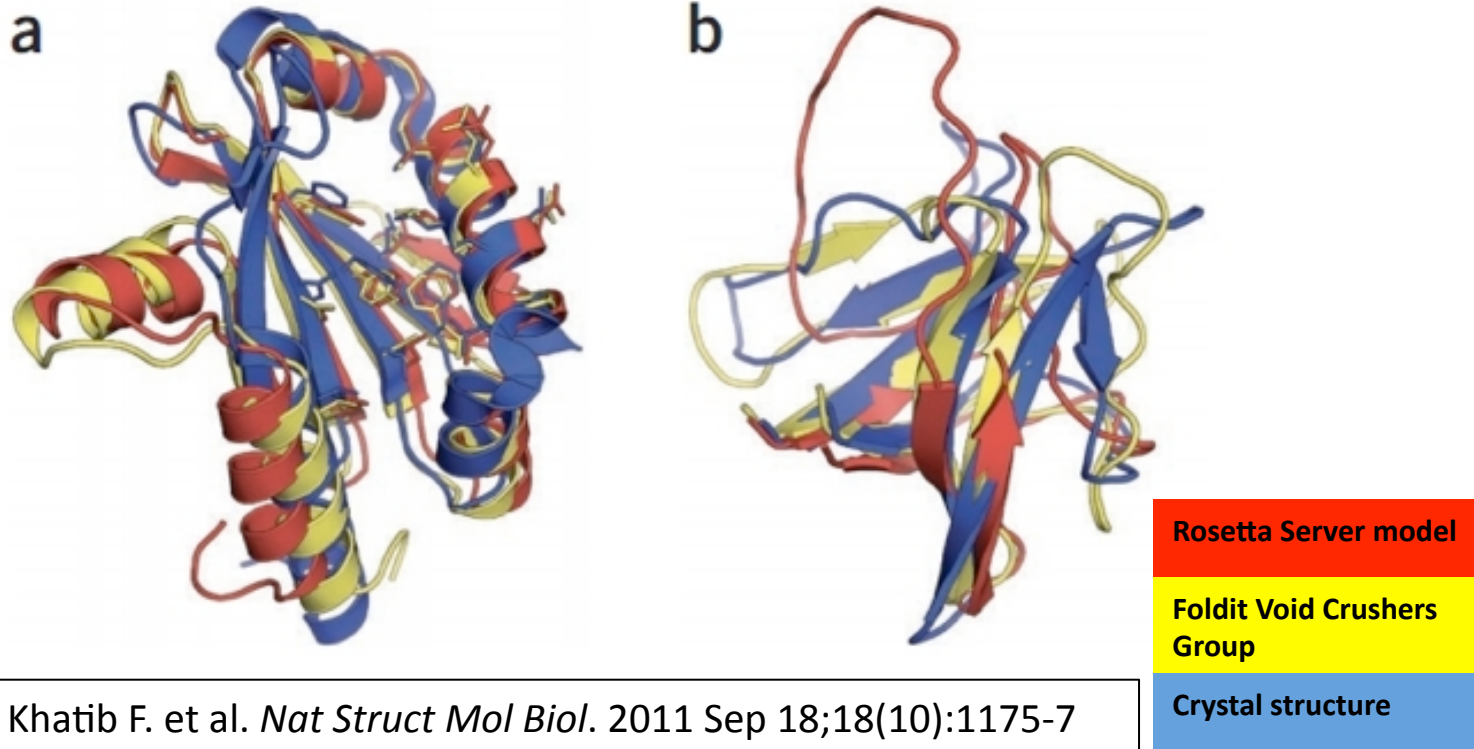Simons, K. T., et al. Proteins Suppl. 1999, 3, 171–176

The Rosetta methods
minima energy searching and limitations

- Protein takes a native structure with the lowest free energy
- Computational searching algorithms may sample local minima foldings (local traps)
- The searching space is immense

Anfinsen, C. B. Science 1973, 181, 223–230

Foldit: a story of social success

a b

Rosetta Server model

Foldit Void Crushers Group

Crystal structure

Khatib F. et al. *Nat Struct Mol Biol*. 2011 Sep 18;18(10):1175-7

- Outperform many prediction algorithms in the CASP9 contest.
- Key in solving the crystal structure of Mason-Pfizer Monkey Virus (M-PMV) retroviral protease.

What's the game?
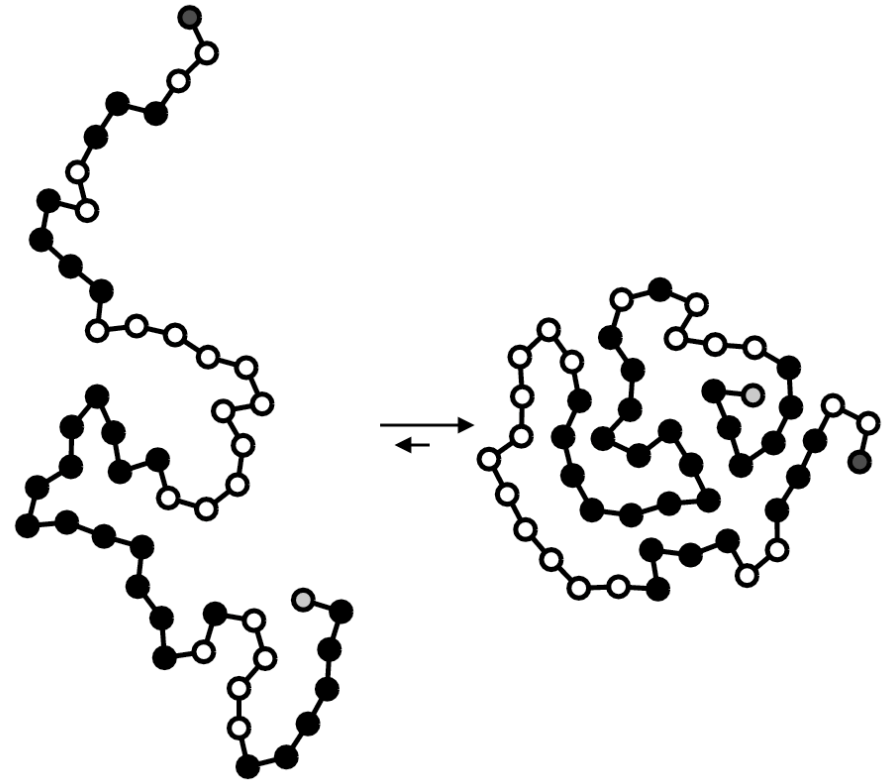## Rosetta algorithms + Motivated social gamers

Copper S. et al. Nature. 2010 Aug 5;466(7307):756-60.

To win the game
tips for high scoring

- Compact blob
  - core
  - surface
- Polar/charged core + nonpolar surface
- Nonpolar core + polar/charged surface
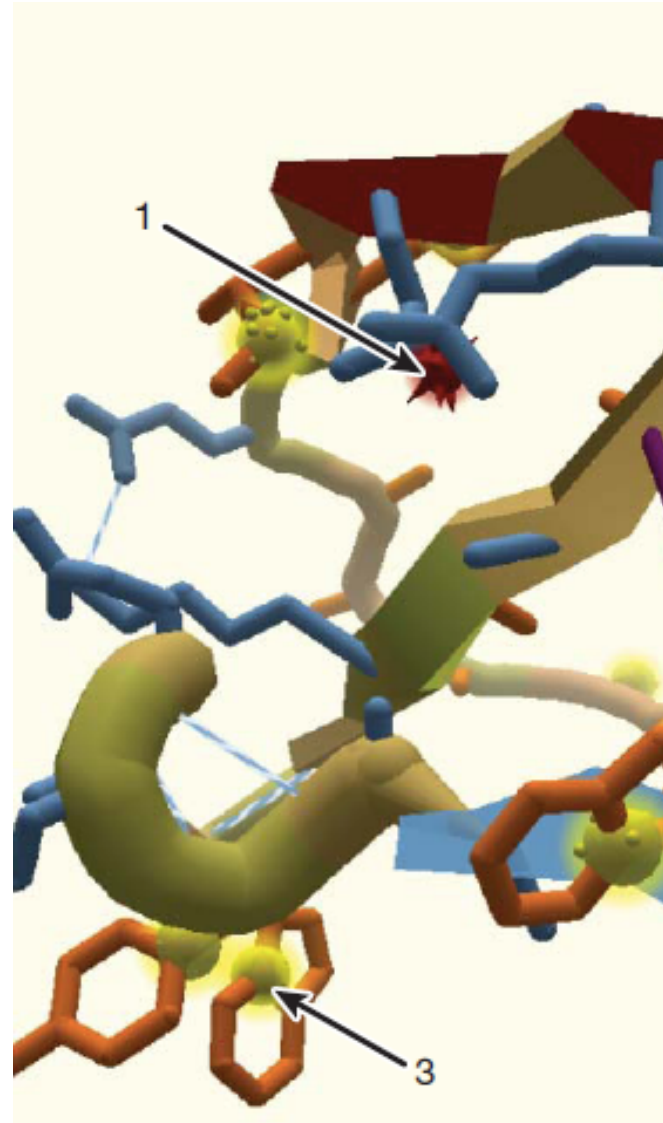- Hydrogen bonds
- Local motifs

hydrophobic amino acids (black spheres)
hydrophilic amino acids (white spheres)

Protein_folding@wiki
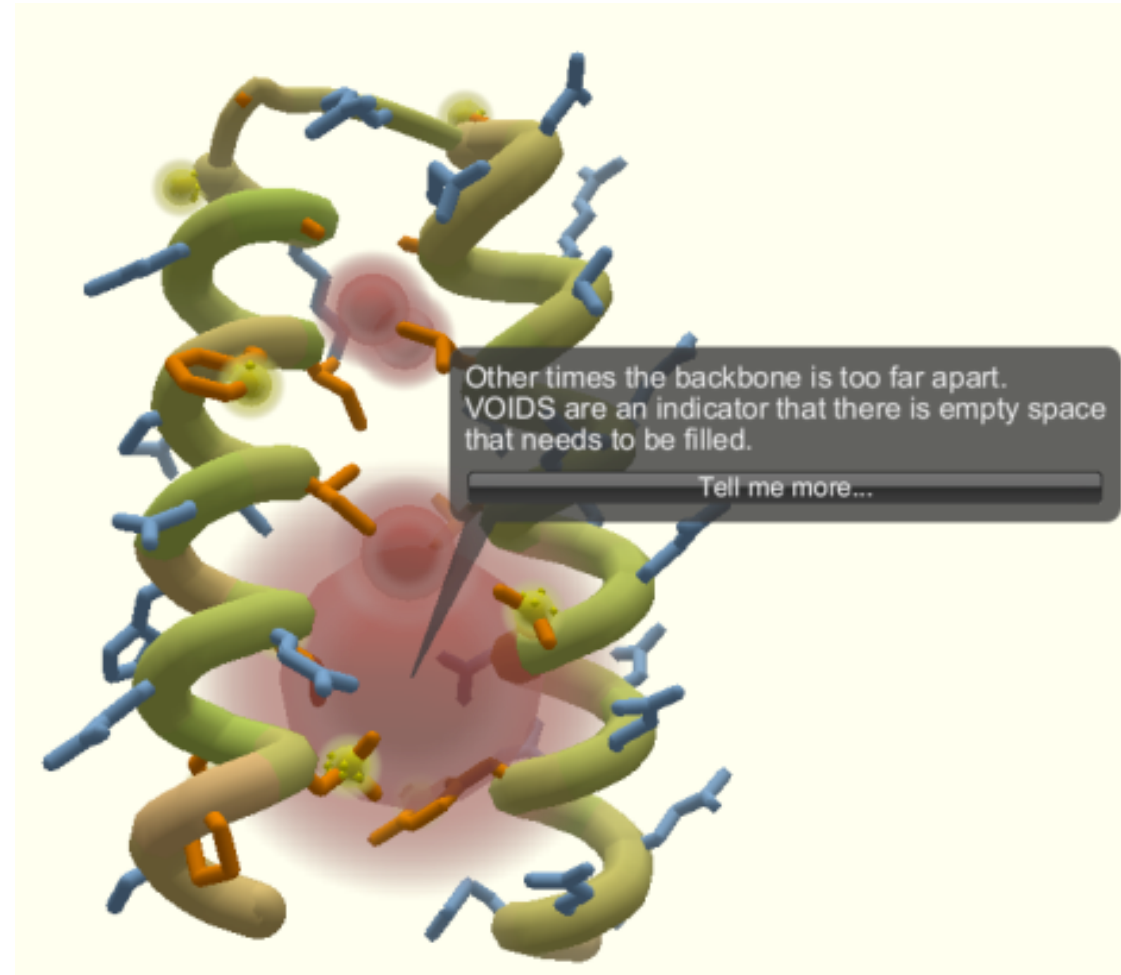
How to play (1)
shake the side-chains

- Side chains too close
  - Side chain clash
  - Indicated by red stars
- use **SHAKE** to fix clashes
-
- hydrophobic side chain
  - exposed (yellow blob)
- hydrophilic side chain

How to play (2) backbone packing

- Backbone
  - too close (collide)
  - too apart (gaps)
- Pull the backbone manually
- Or use the **WIGGLE** to improve the backbone automatically

Other times the backbone is too far apart. VOIDS are an indicator that there is empty space that needs to be filled.

Tell me more...
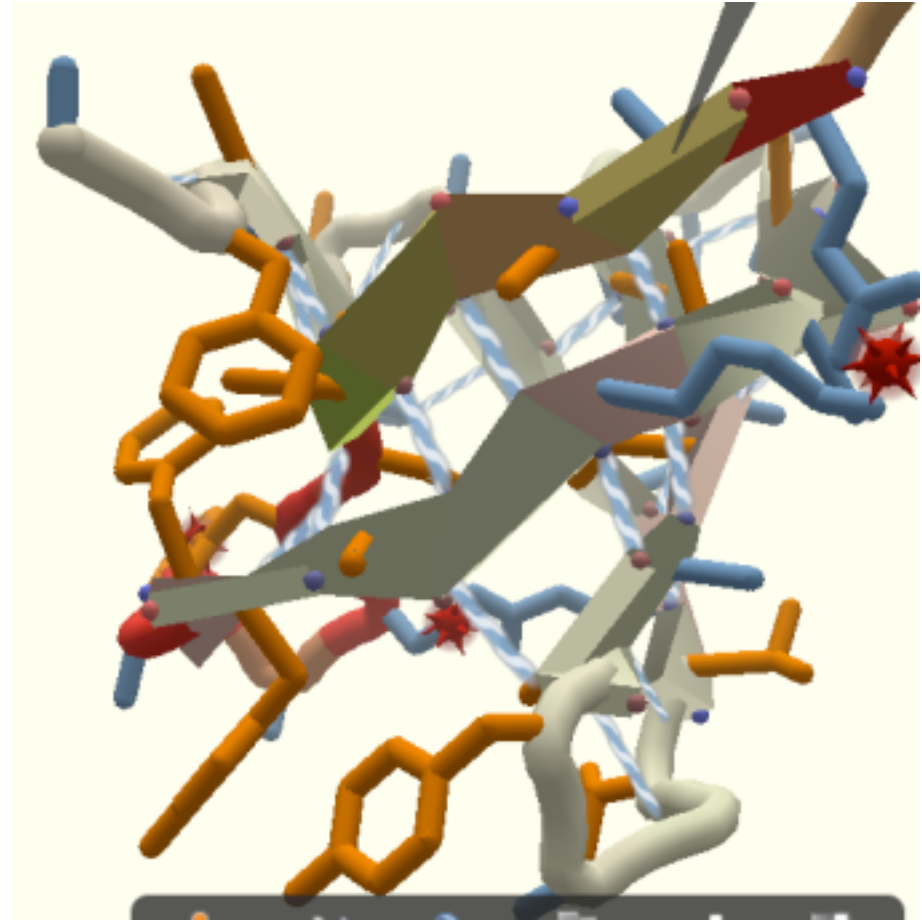
How to play (3) hydrogen bonding

- zig-zag sheets line up together

- connect these sheets with **RUBBER BANDS**

- deal with problematic segments by using the **REBUILD** tool
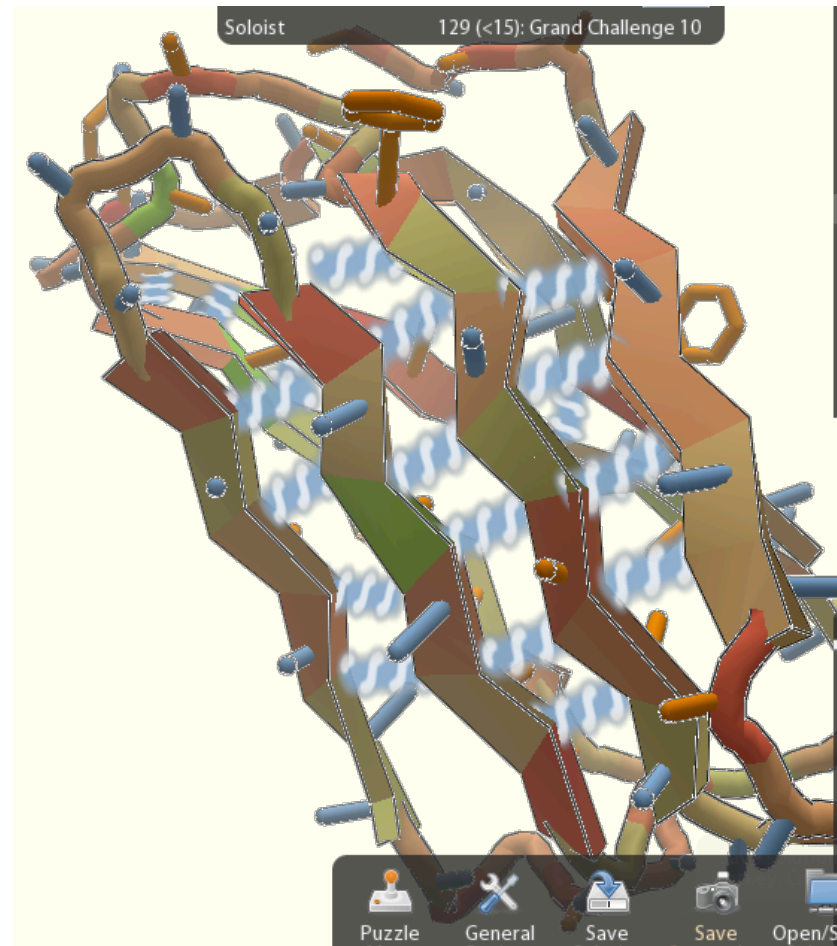
How to play (4) Hydrophobics and hydrophilics

- Orange sidechains are **HYDROPHOBIC**

- **EXPOSEDS** may appear on unburied hydrophobics

- **Blue** sidechains need plenty of space (outward)

- rotate with the **TWEAK** tool

Play and contributing ...

- Software availability
  - http://fold.it/portal/info/science

- Community and wikis
  - http://foldit.wikia.com/wiki/Foldit_Wiki

- Puzzles releases
  - http://fold.it/portal/puzzles

Thanks for your attention :)