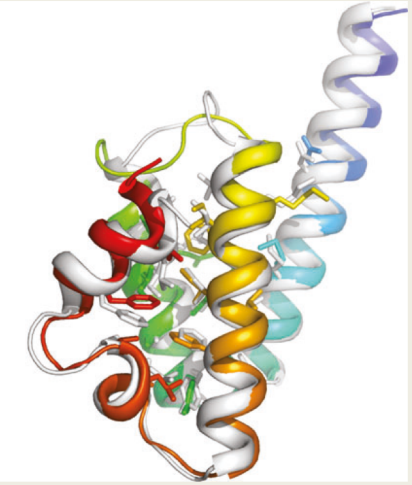
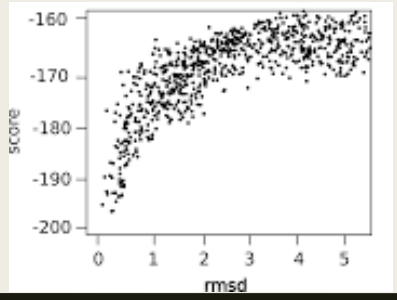


HOMOLOGY MODELLING

MCDB187
Marcus Gallagher-Jones
marcusgj@chem.ucla.edu

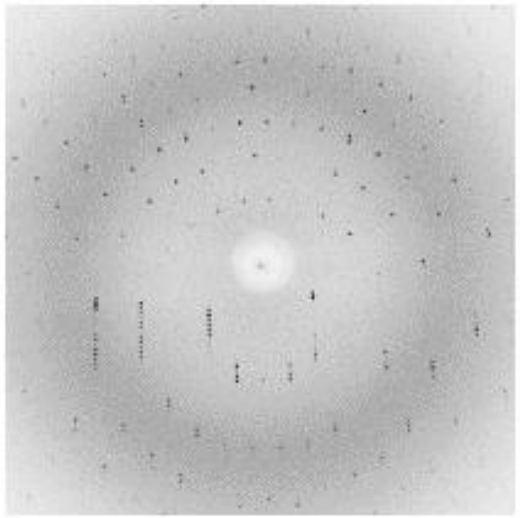


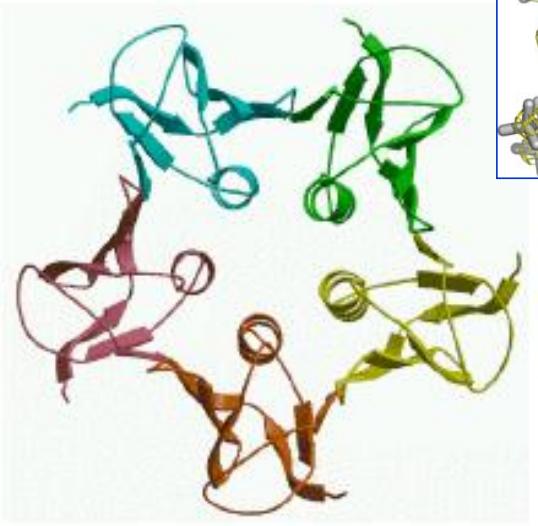
Why do we care about structure? Structure/Function relationship

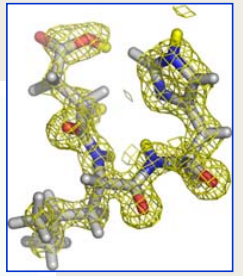
- You may already know about your targets function – Structure can provide the why of function.
- Atomic coordinates allow quantitative predictions to be made.
- Being able to look at your model can provide new insights.
- Knowledge of a structure can allow us to make beneficial changes.

How do we currently get structures?

X-ray crystallography







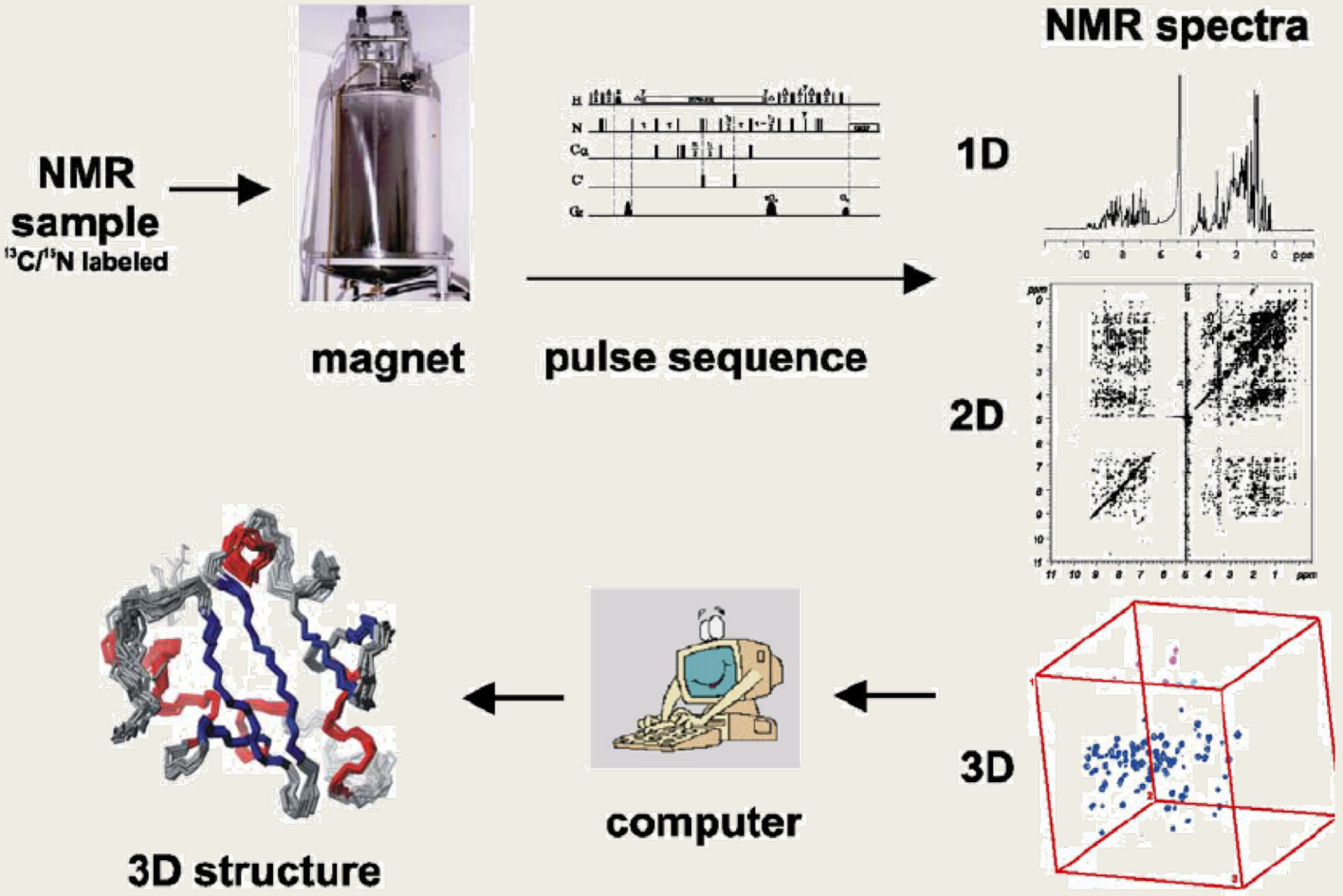
Target selection
Protein production
Crystallization

Data collection
Phasing

Structure calculation
Refinement
Validation
Deposition to PDB
Annotation

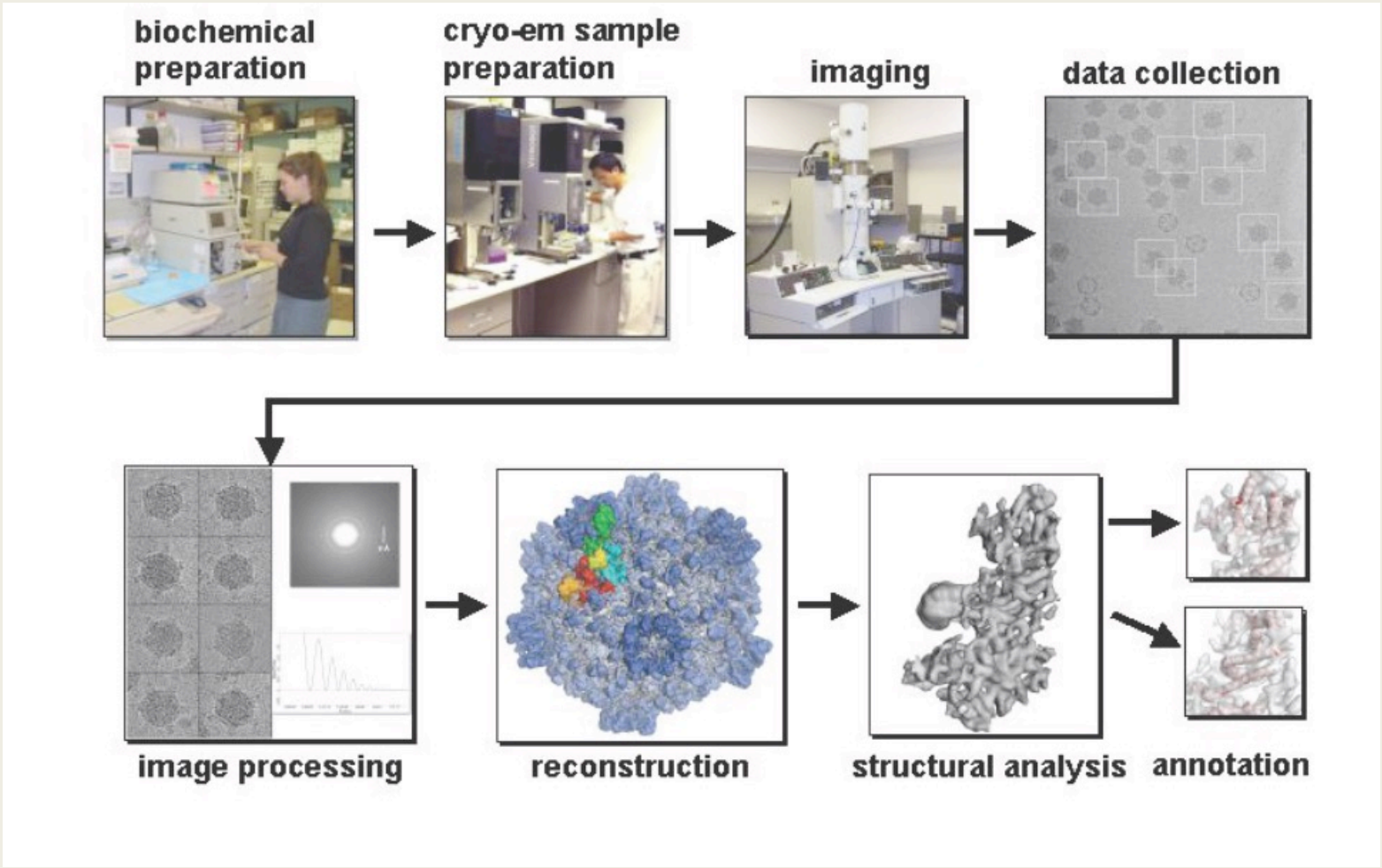
http://www-structmed.cimr.cam.ac.uk/Course/Overview/Overview.html#methods

Nuclear Magnetic Resonance (NMR)

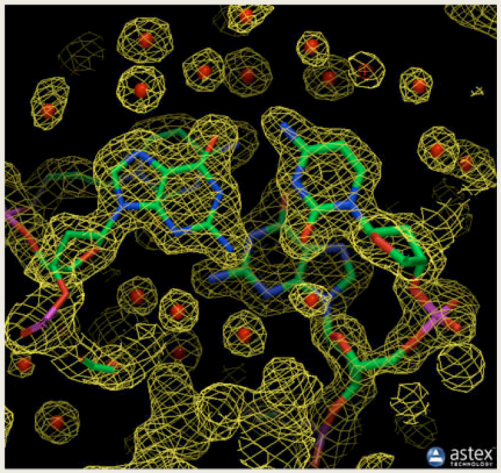


http://www.embl.de/nmr/gattler/teaching

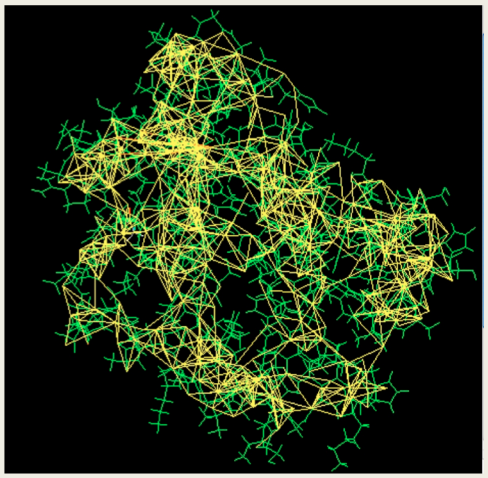
Cryo-Electron Microscopy (Cryo-EM)



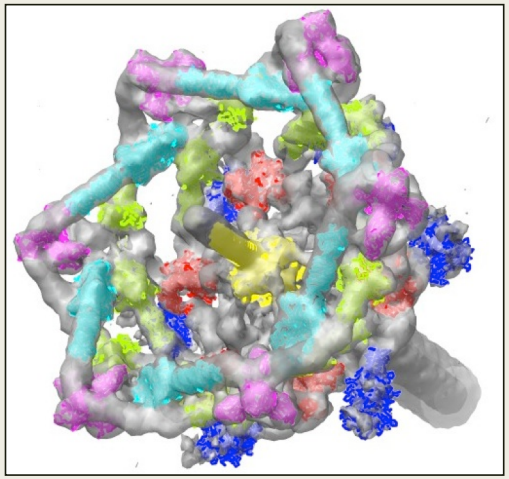
Each method provides different set of experimental data.



X-ray Crystallography: Electron density from a structure of DNA is shown here (PDB entry 196d), along with the atomic model



NMR: Restraints used to solve the structure of a small monomeric hemoglobin. The protein (1vre and 1vrf) is shown in green, and restraints are shown in yellow.



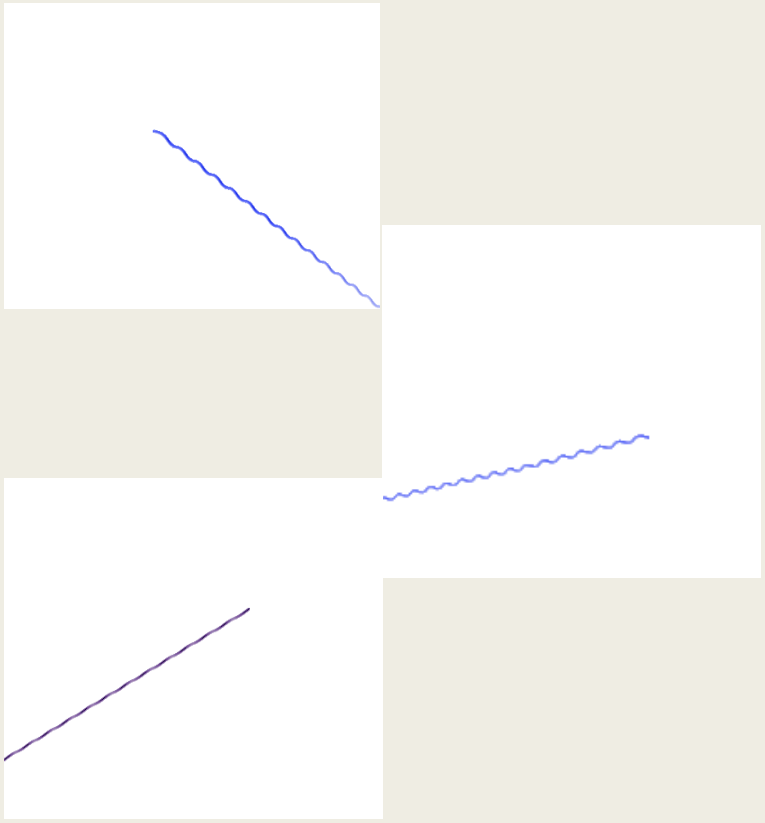
EM: Tail of the T4 bacteriophage. Surface rendering of the EM data (emd-1048) with atomic coordinates from PDB entries 1pdf, 1pdi, 1pdl, 1pdm, 1pdp, and 2fl8.

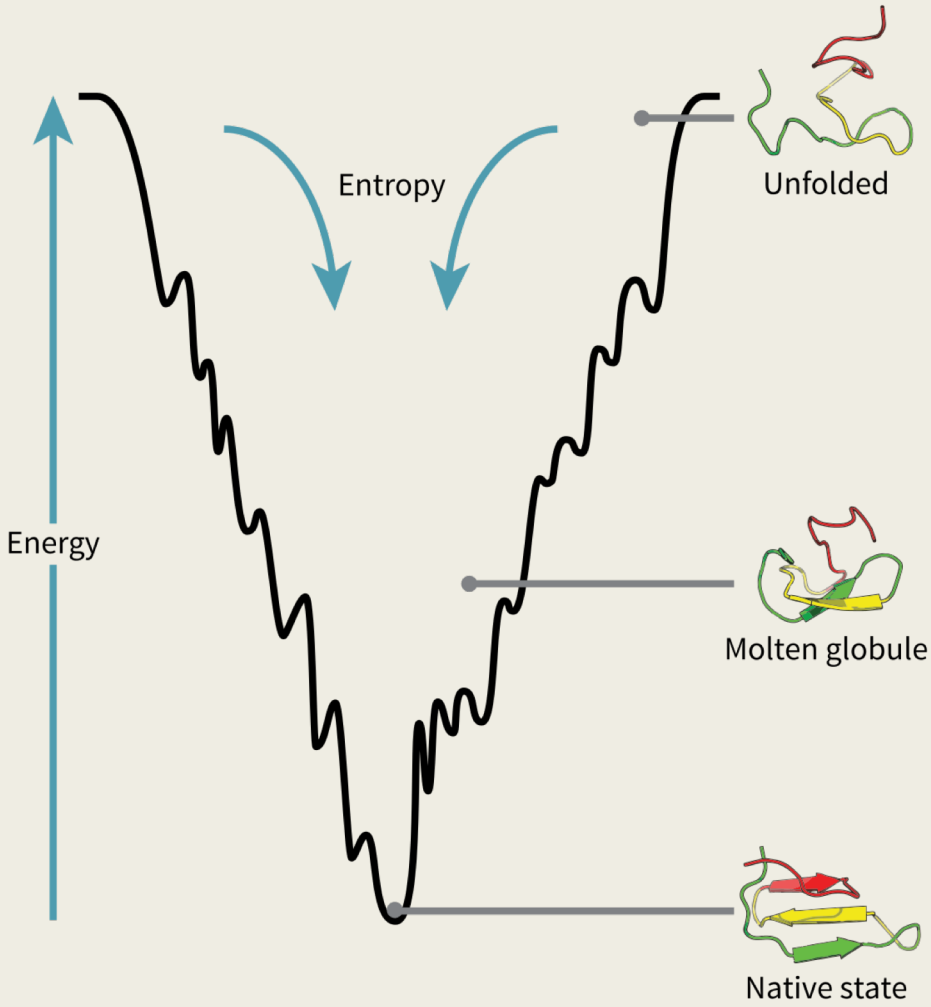
Obtaining structure is challenging

- Obtaining structure experimentally is challenging, time consuming and expensive.
- Experimental methods in structural biology are currently lagging far behind sequencing.
  - *PDB has* 139,717 entries (as of 04/25/2018), however only 77,678 are unique.
  - By contrast UniProt contains over 114,000,000 unique entries (as of 04/25/2018).
- Definite need for good modelling tools.

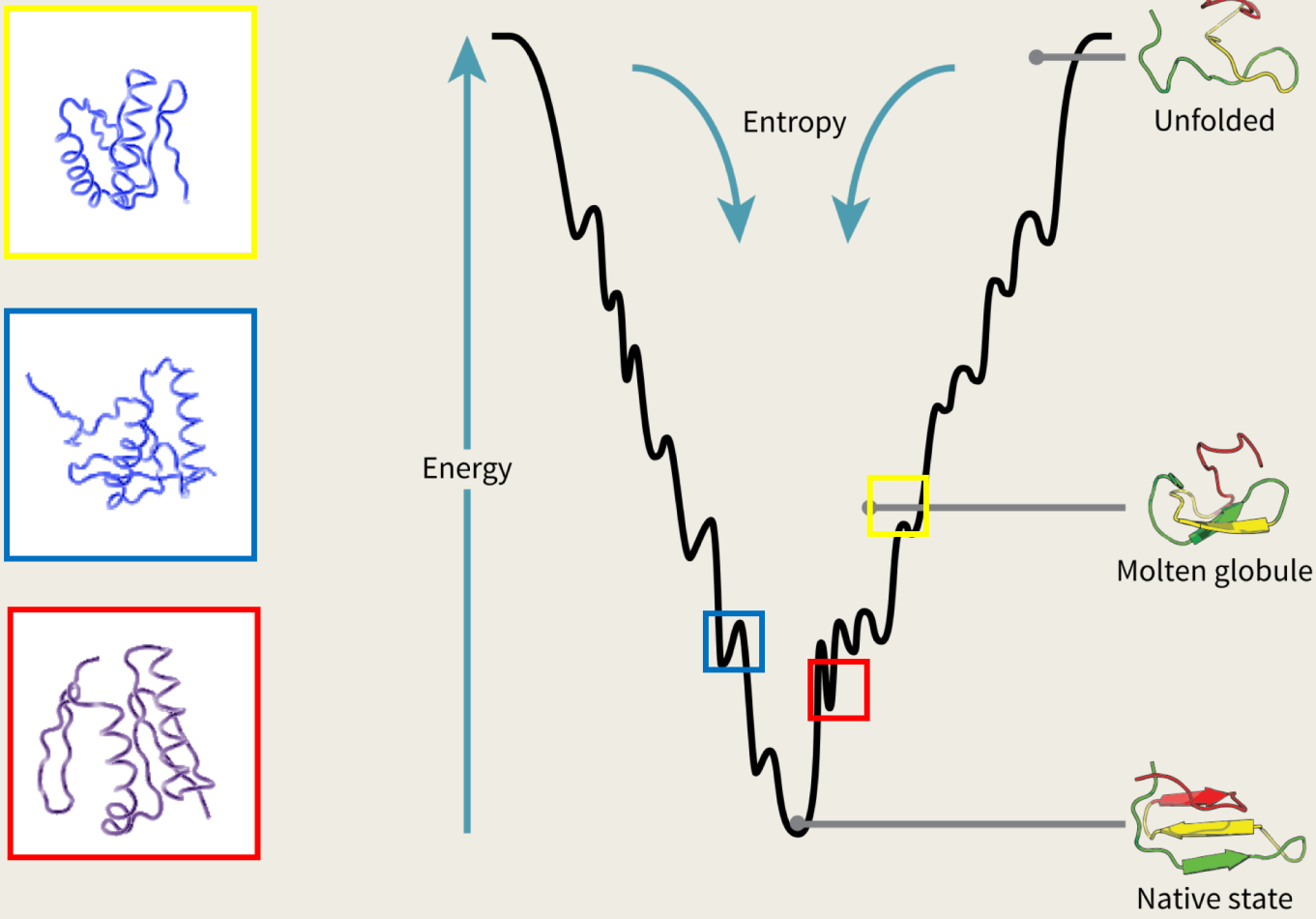
Can we predict there dimensional protein structure from sequence alone?

The challenge – folding landscapes are complex





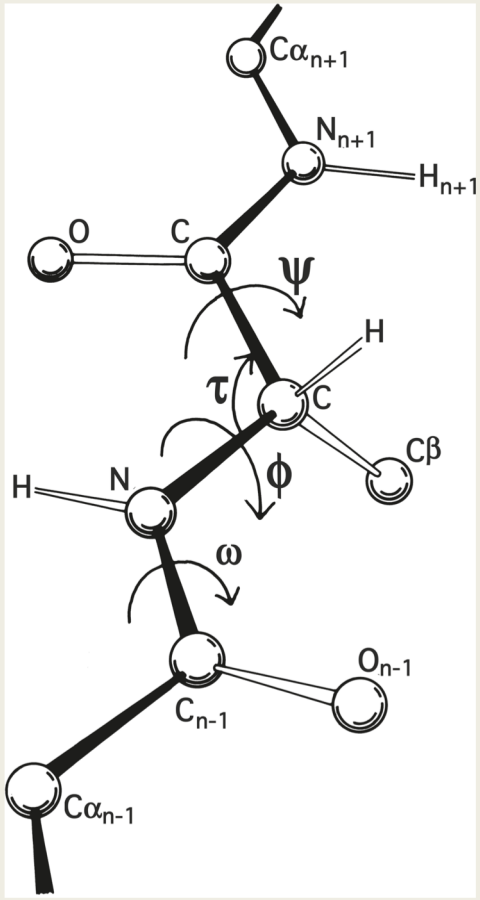
The challenge – folding landscapes are complex



Folding has a large solution space with many local minima.

Can we leverage what we already know about structure to constrain the number of possible solutions?

Dihedral angles and Ramachandran plot



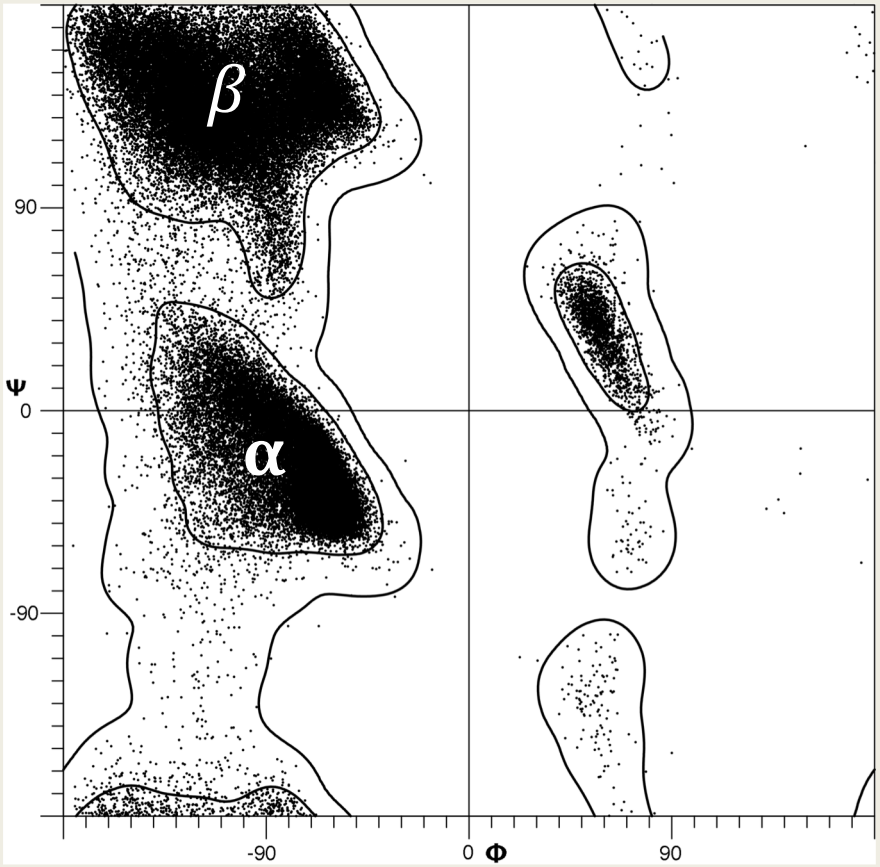
Dihedral angles of the peptide backbone are the source of almost all the interesting variability in protein conformation.

Of these phi and psi (either side of C**α**) are the most important.

For amino acids other than glyceine and proline the number of possible phi/psi angles is limited.

Richardson. 1981, *Advances in Protein Chemistry,* **34**, 167 -339

Dihedral angles and Ramachandran plot



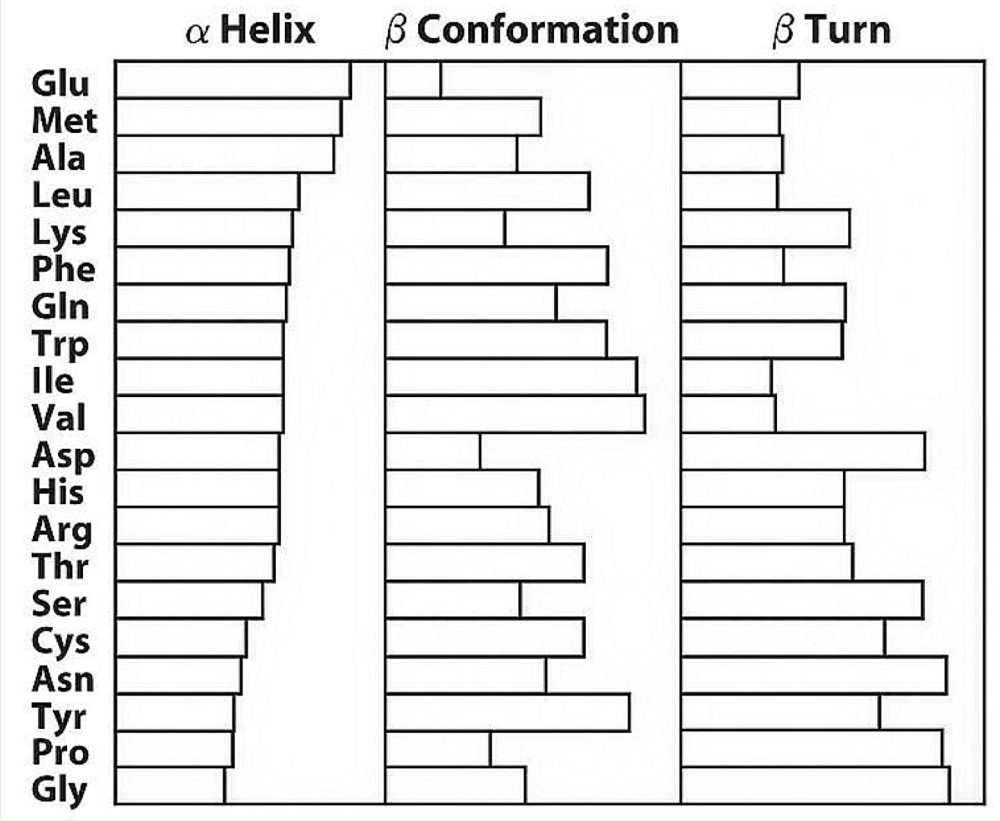
Lovell et al. 2003 *Proteins* 50:437,

Dihedral angles of the peptide backbone are the source of almost all the interesting variability in protein conformation.

Of these phi and psi (either side of Cα) are the most important.

For amino acids other than glyceine and proline the number of possible phi/psi angles is limited.

Propensity for Secondary Structure



Not all amino acids favor the same secondary structure.

Destabilize Helices

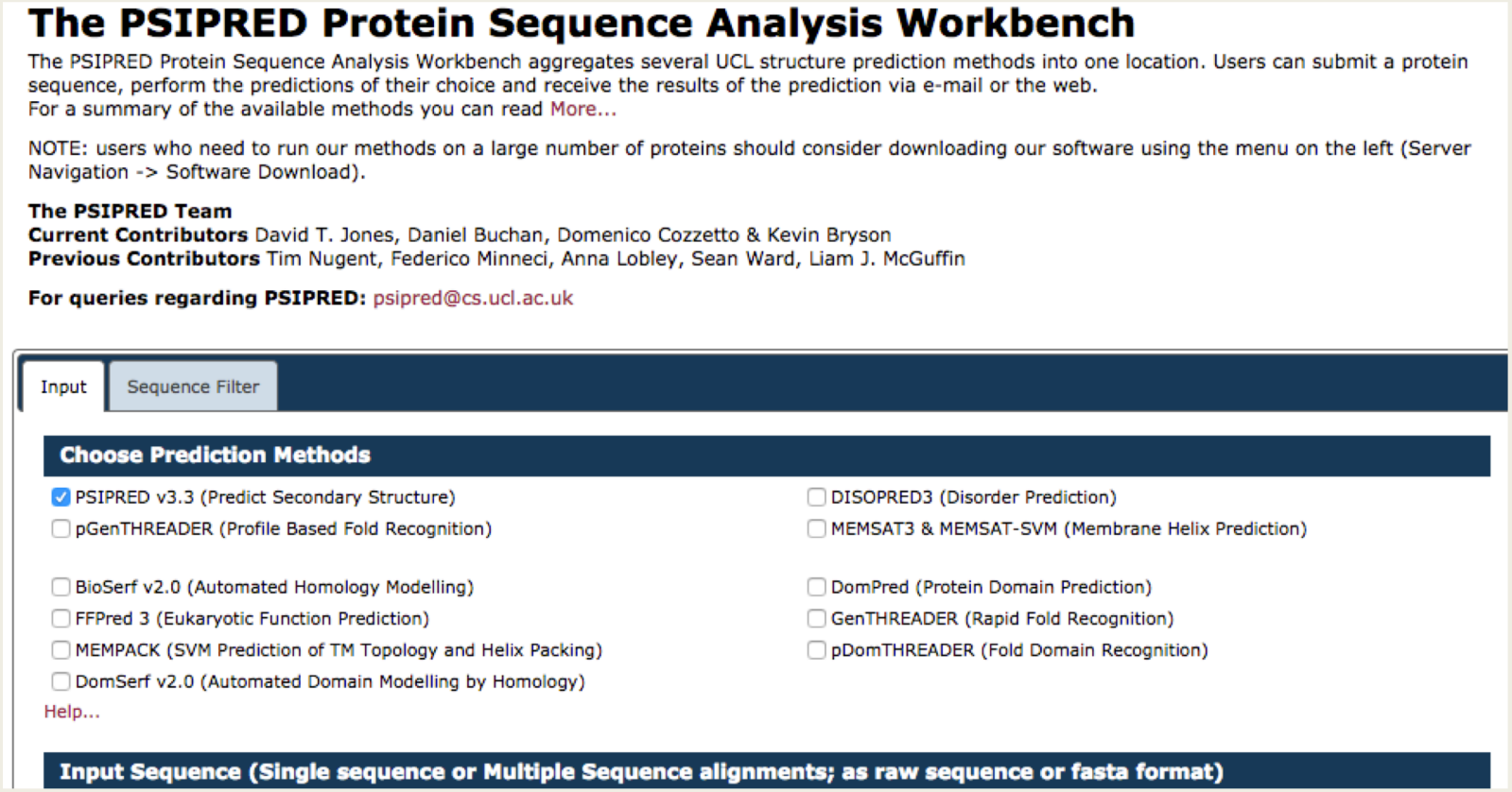
Valine, Isoleucine: Branched at C_β position.

Serine, Aspartate, Asparagine: Compete for main chain H-bonds.

Destabilize Sheets

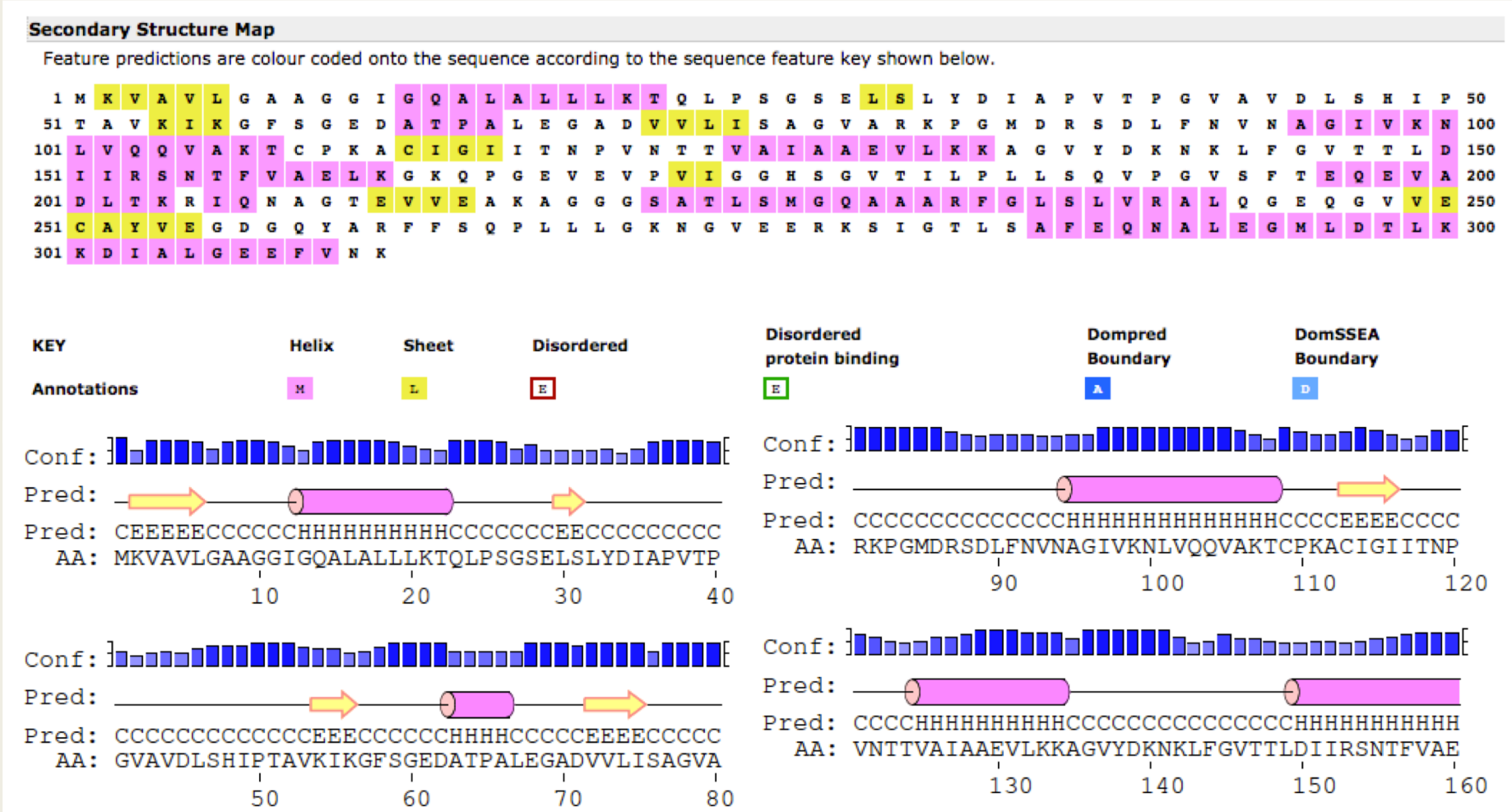
Proline, Glycine: Too flexible or rigid, can kink main chain.

Secondary structure prediction - psipred



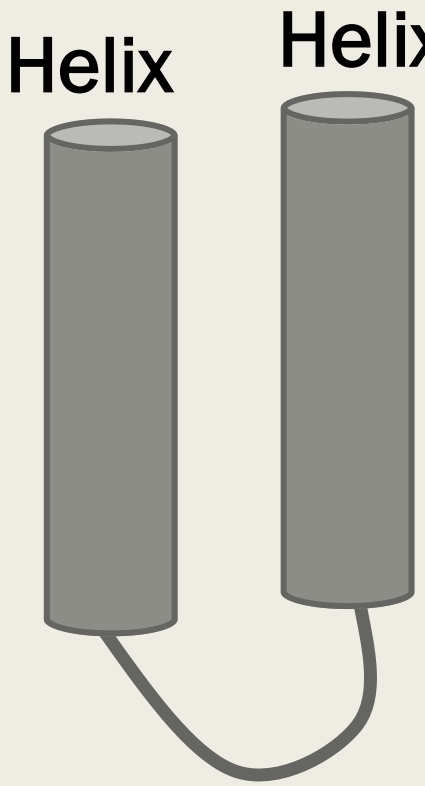
http://bioinf.cs.ucl.ac.uk/psipred/

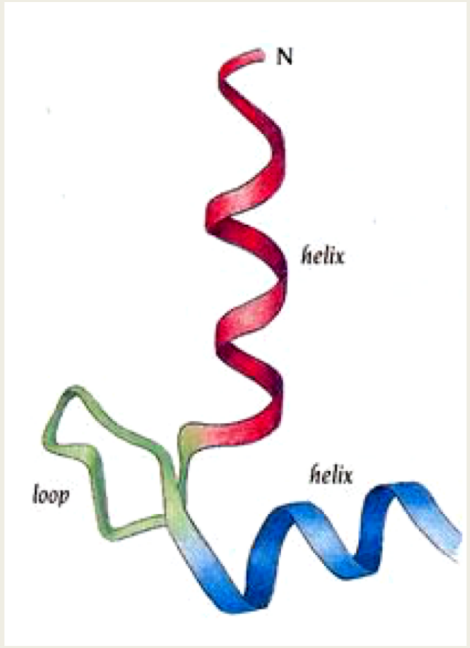
Psipred results page



Conserved motifs

Alpha Helical Motifs





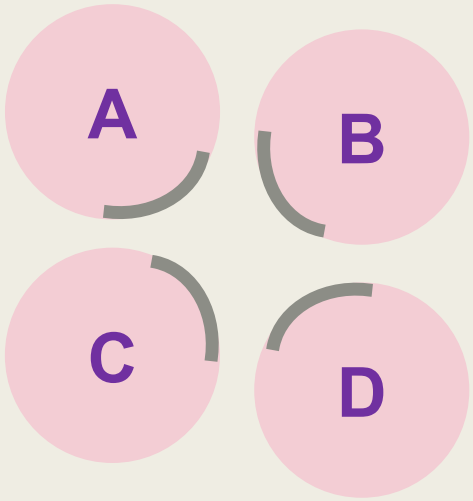
Stretches of residues with propensity to form alpha helices can be mapped onto a helical wheel.

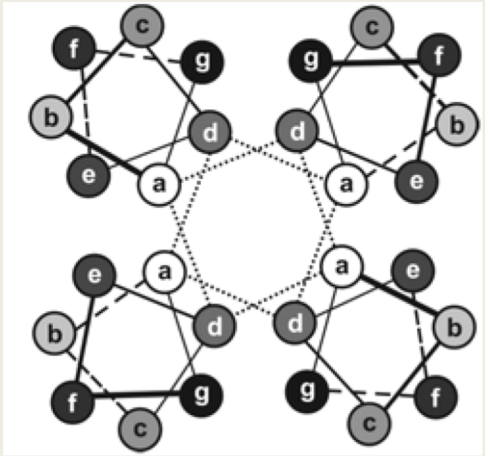
Residues are mapped on the helix from N-to-C terminus (helices show right handedness) with 3.7 residues per turn.

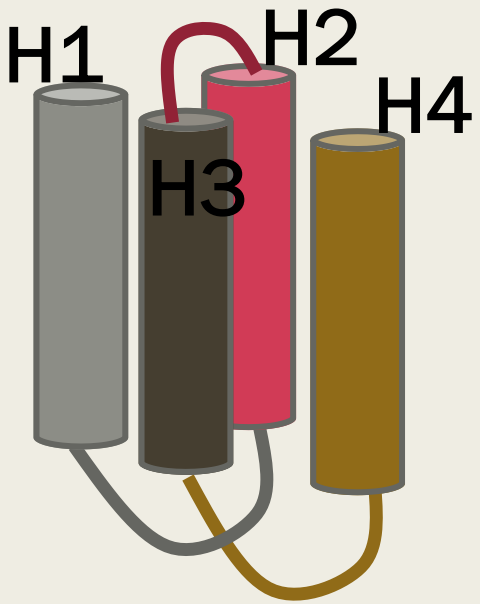
Helices can display different properties on different faces.

Helix – Helix Combinations

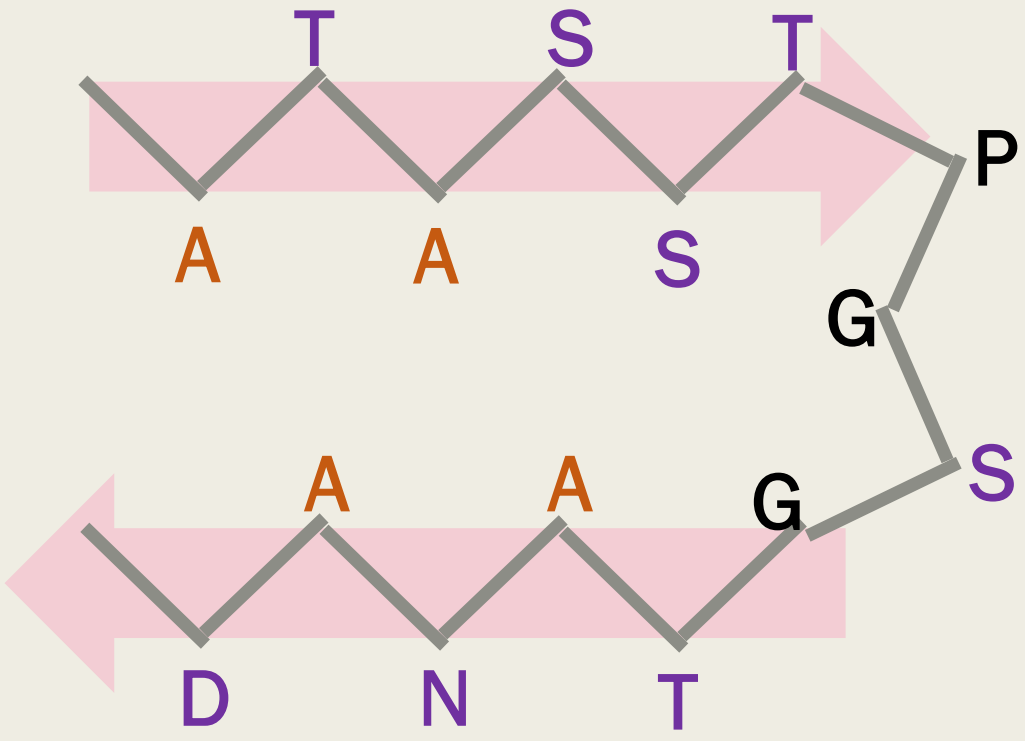
The four helix bundle: Helices wrap around each other, minimizing exposure of hydrophobic residues to aqueous environment.

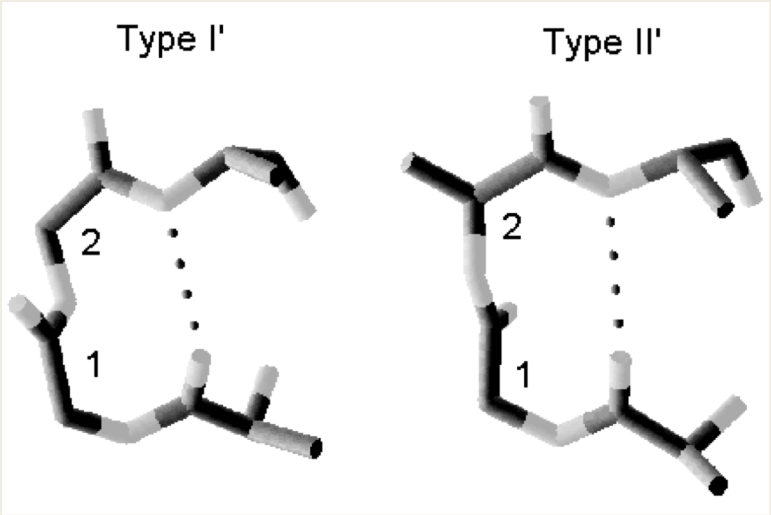






Beta Turn (Hairpin)

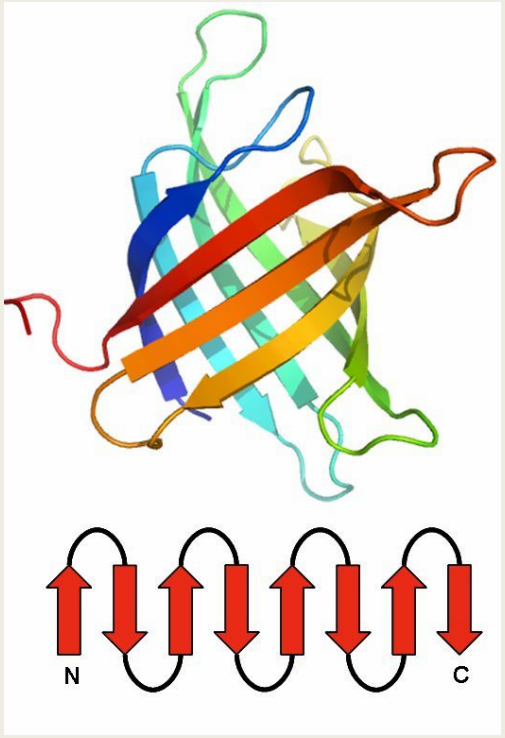




Turn motifs often have prolines, glycines, or other residues that facilitate kinked backbones.

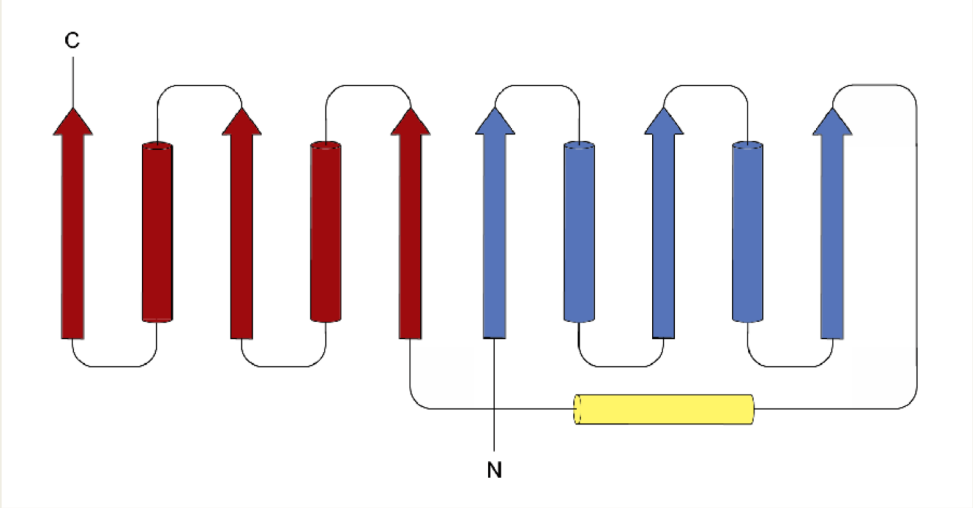
Sheets, Keys and Barrels

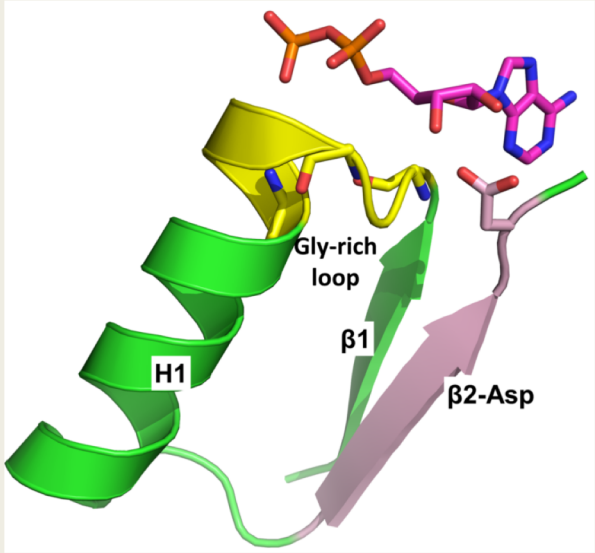
Sheets have a natural twist and can wrap to create more complex folds. Motifs vary based on the way a series of beta strands are linked:



A beta barrel formed by anti-parallel strands.

Secondary Structure Combinations
The Rossmann Fold





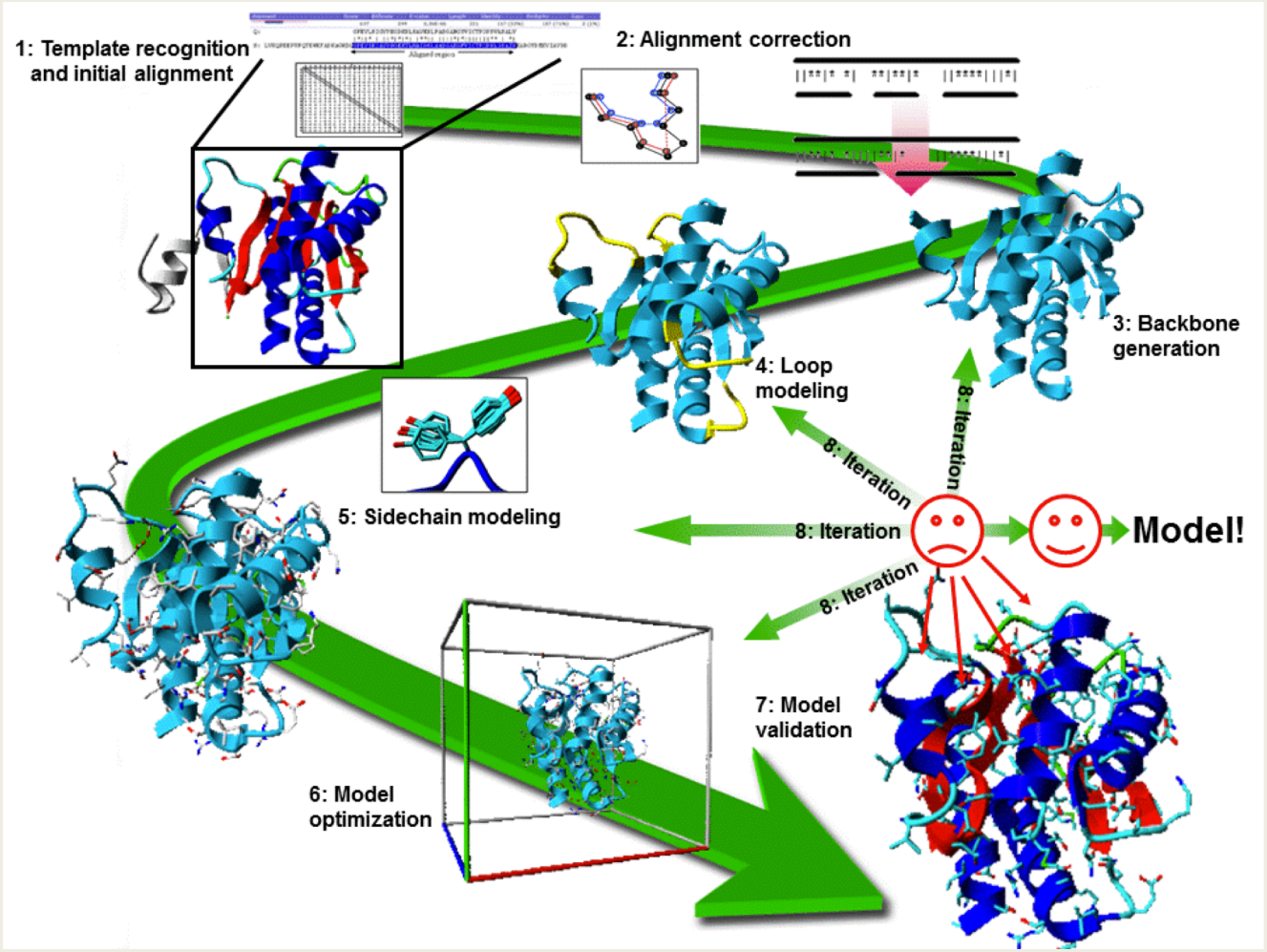
Motif composed of alternating strand – helix pairs often giving rise to a central sheet against which helices stack.

Rossmann fold - a putative minimal element in cofactor utilizing enzymes. (Laurino 2016)

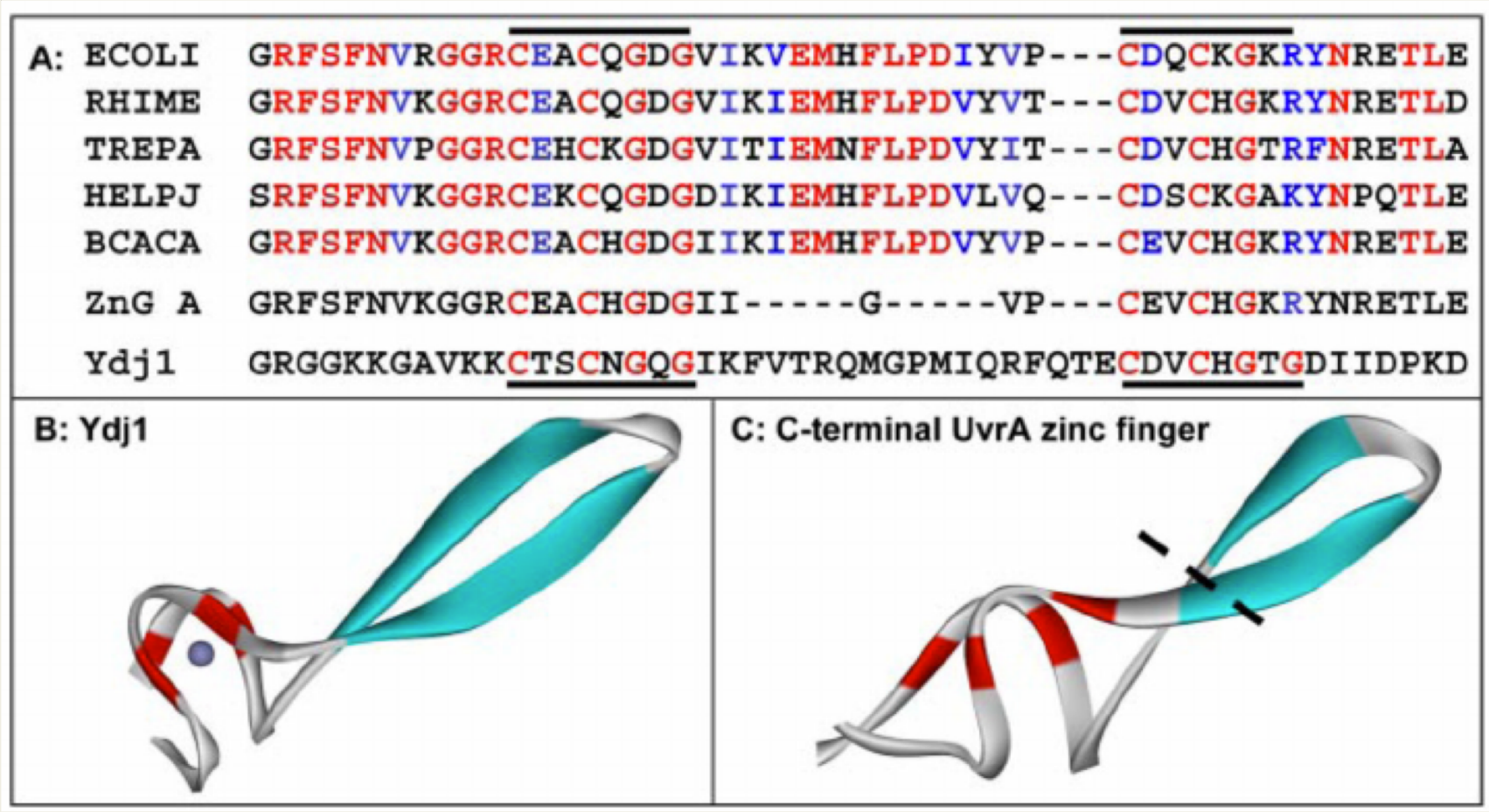
Homology modelling is the logical extension of this

- Proteins found in nature generally exist in their lowest energy possible conformation.
- Exploitexisting structural information to make inferences about unknown macromolecules.
- Alignment of an unknown structure (target) to a similar known structure (template).
- Typically this is an evolutionarily related protein/molecule.
- Ultimate goal is to **predict** a structure with accuracy **comparable to experimental methods.**

Basic steps of homology modelling

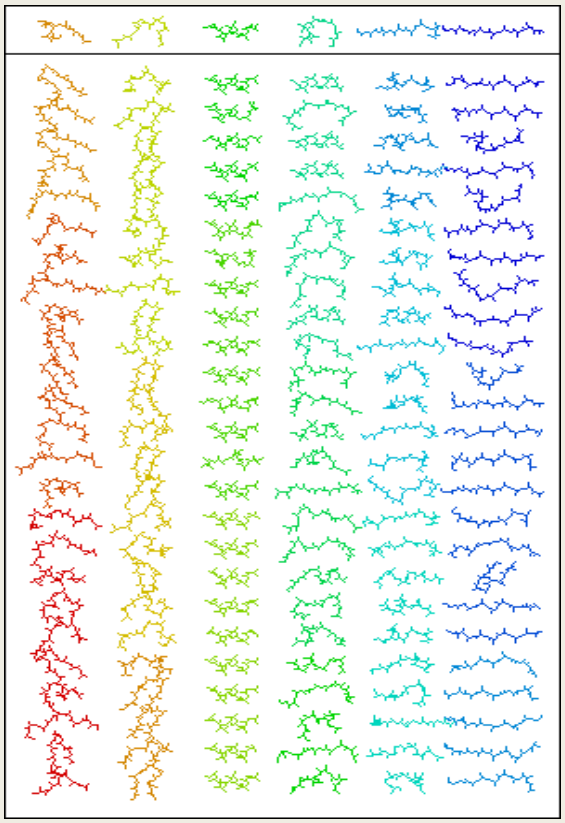


An example alignment



Croteau et al. Journal of Biological Chemistry 281(36):26370-8

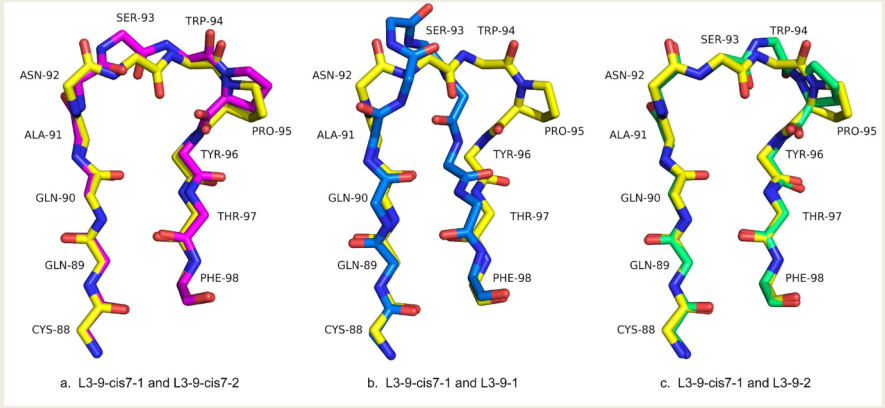
Fragment insertion to fill in gaps

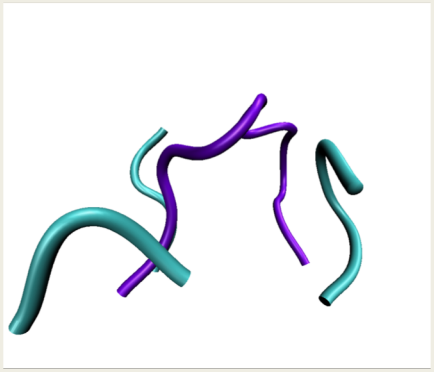


Local interactions: fragments

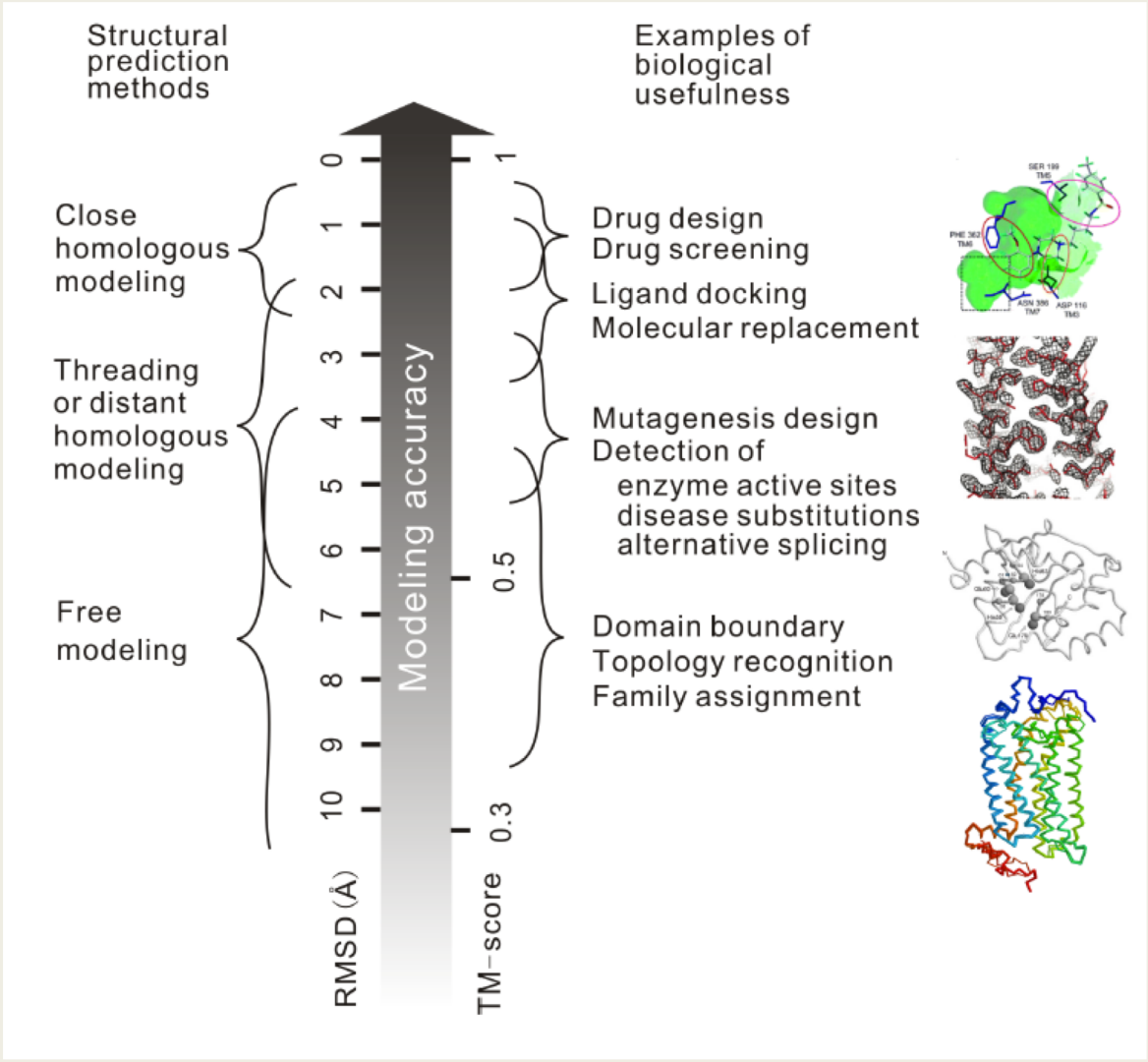
- Derived from known structures
- *Sampled for similar sequences/secondary structure propensity*
- *Fragment library represents accessible local structures for short sequence*

Loop modelling – The bane of homology modelling





- Loops are a challenge for homology modelling – few rules.
- Important for understanding biological function, interfaces and interactions.
- With current methods it is hard to accurately predict loops longer than 12 – 14 residues.
- Problem is not a sampling one
  - *Can computationally sample all loop conformations but discrimination is difficult.*
- Some methods of loop modelling are inspired by robotics and video games!



Zhang. Curr Opin Struct Biol. 2009 April ; 19(2): 145–155

Homology modelling best practices

- Garbage in, garbage out – the converse is also true.
- Sequence identity is generally a good predictor of how well things will work.
- However, not all regions of protein are equal.
  - *Be mindful of secondary structure.*
- Generally sequences well matching in the hydrophobic core can give good models – more evolutionarily conserved.
- Quality of the template model is also very important
  - *You can only be as confident in homology model as you are in template.*