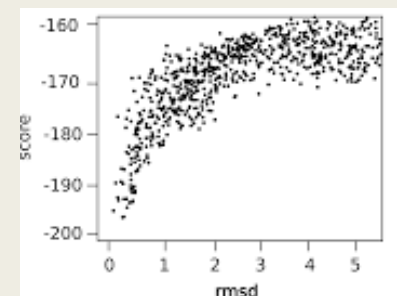


HOMOLOGY MODELLING

MCDB187

Marcus Gallagher-Jones
marcusgj@chem.ucla.edu



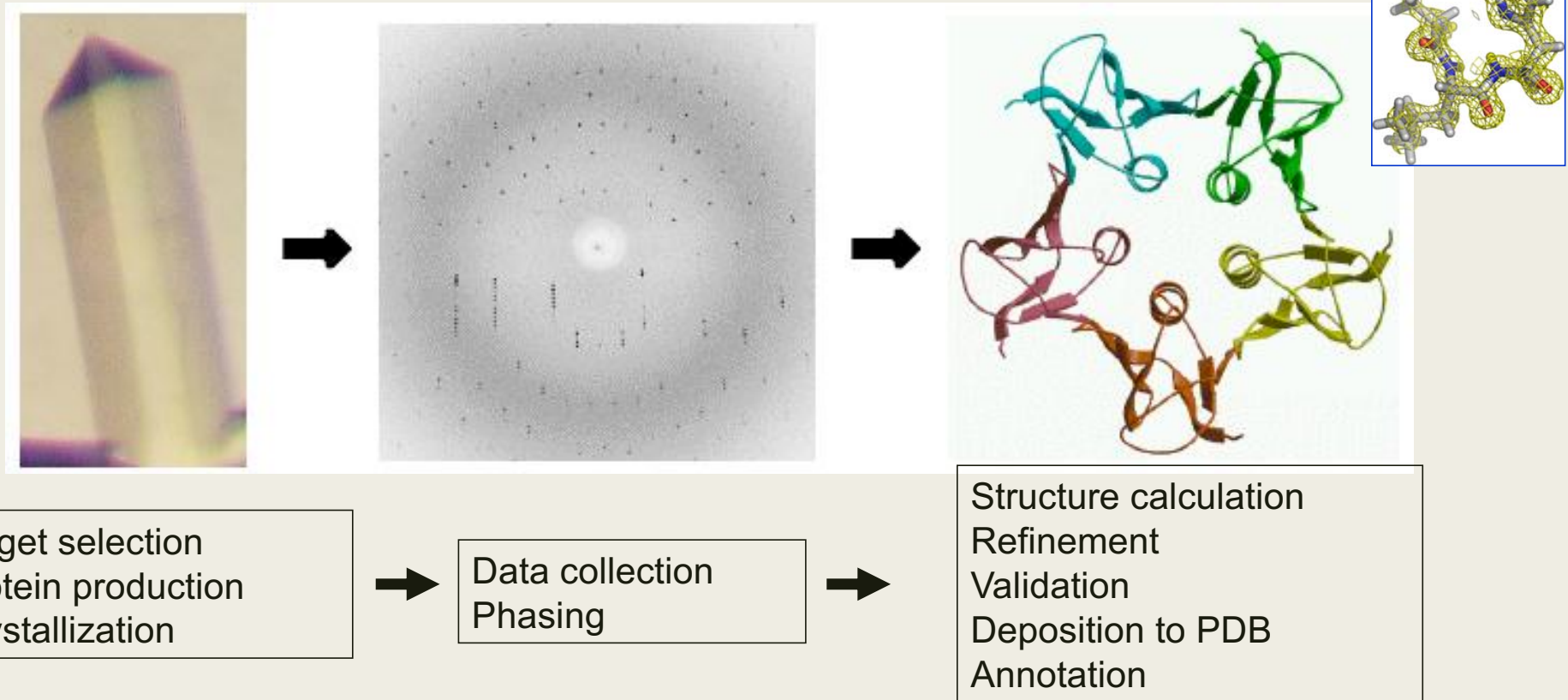
Why do we care about structure?

Structure/Function relationship

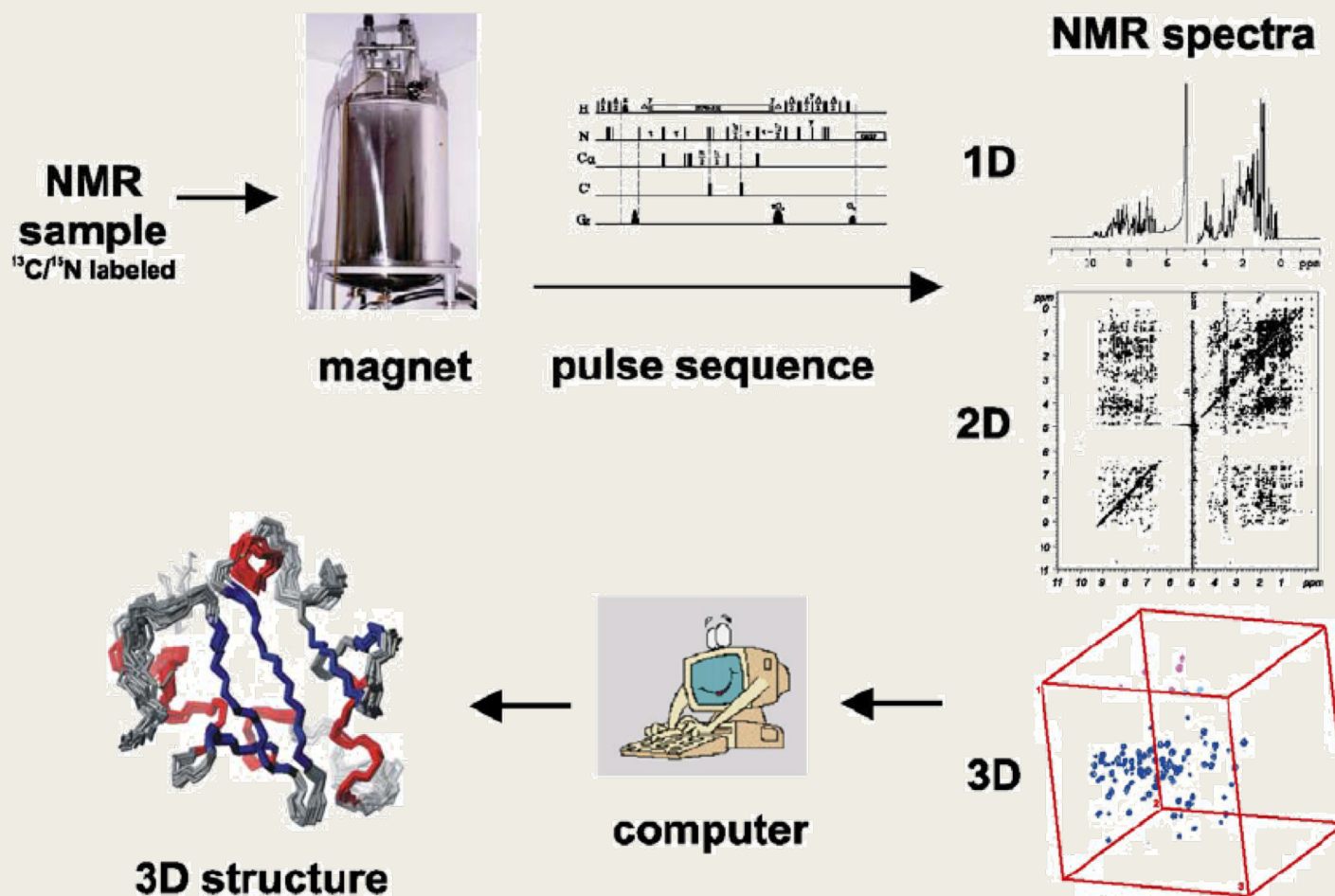
- You may already know about your targets function – Structure can provide the why of function.
- Atomic coordinates allow quantitative predictions to be made.
- Being able to look at your model can provide new insights.
- Knowledge of a structure can allow us to make beneficial changes.

How do we currently get structures?

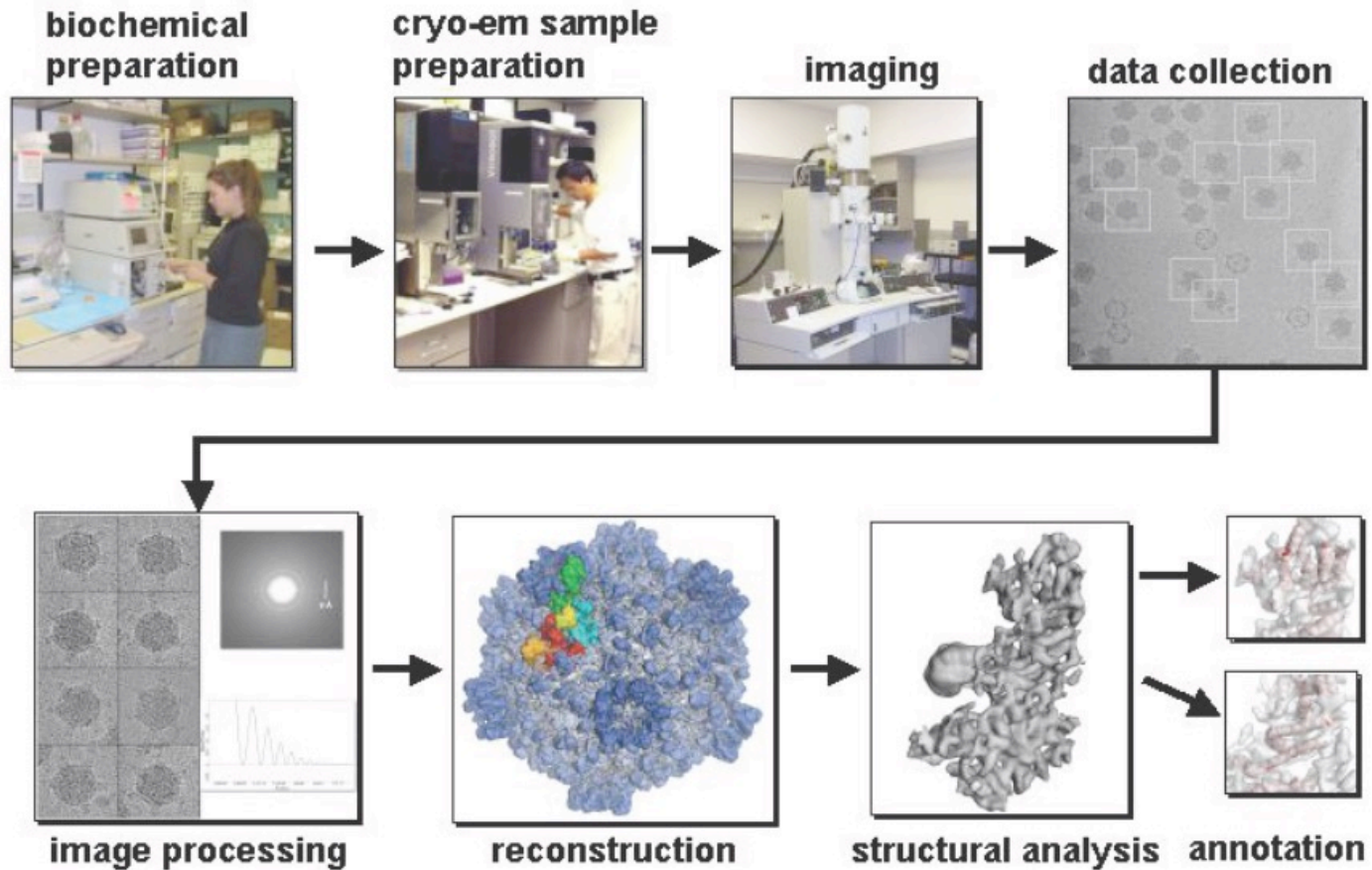
X-ray crystallography



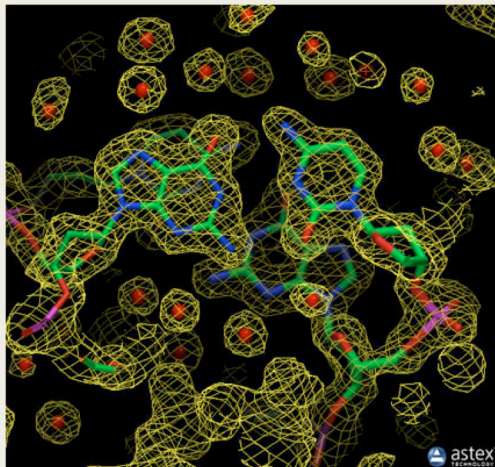
Nuclear Magnetic Resonance (NMR)



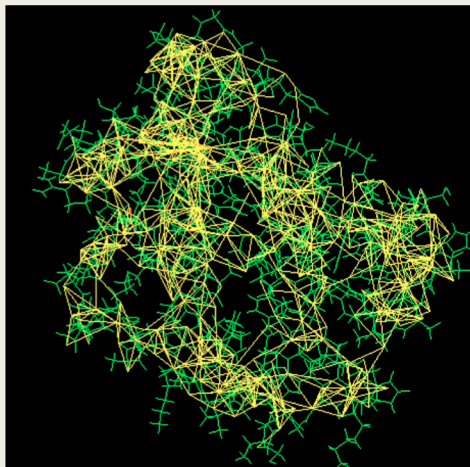
Cryo-Electron Microscopy (Cryo-EM)



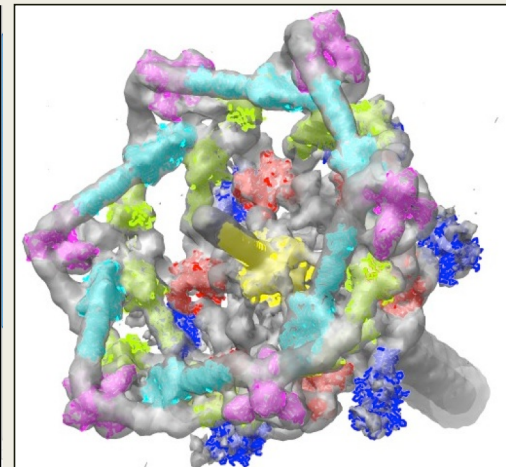
Each method provides different set of experimental data.



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.



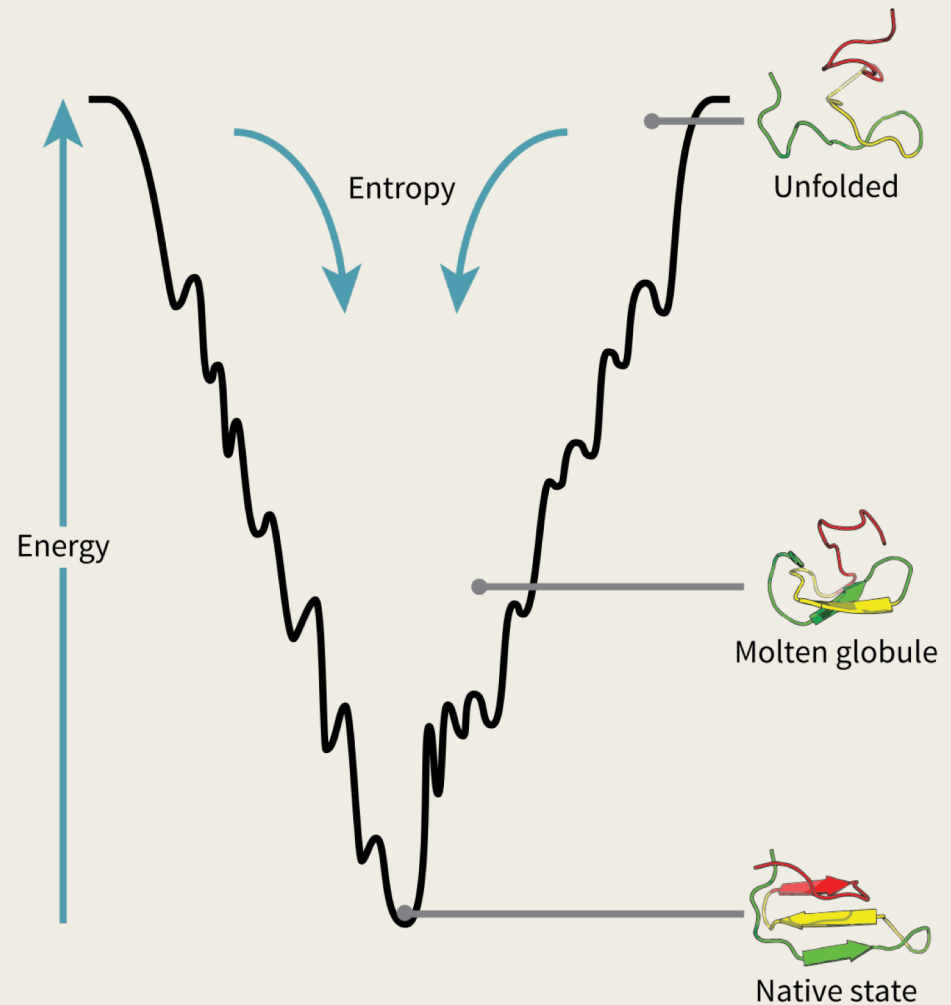
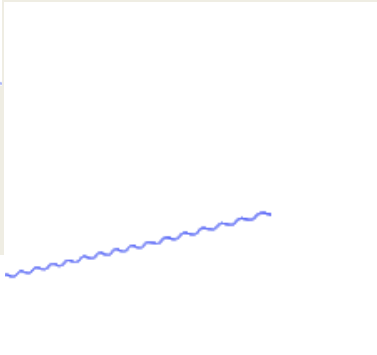
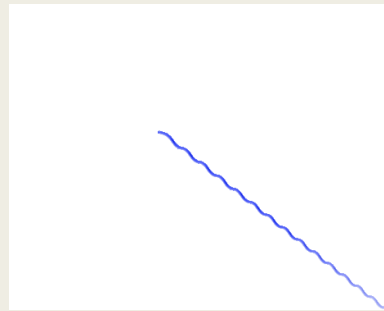
EM: Tail of the T4 bacteriophage. Surface rendering of the EM data (emd-1048) with atomic coordinates from PDB entries 1pdf, 1pdi, 1pdl, 1pdm, 1pdp, and 2fl8.

Obtaining structure is challenging

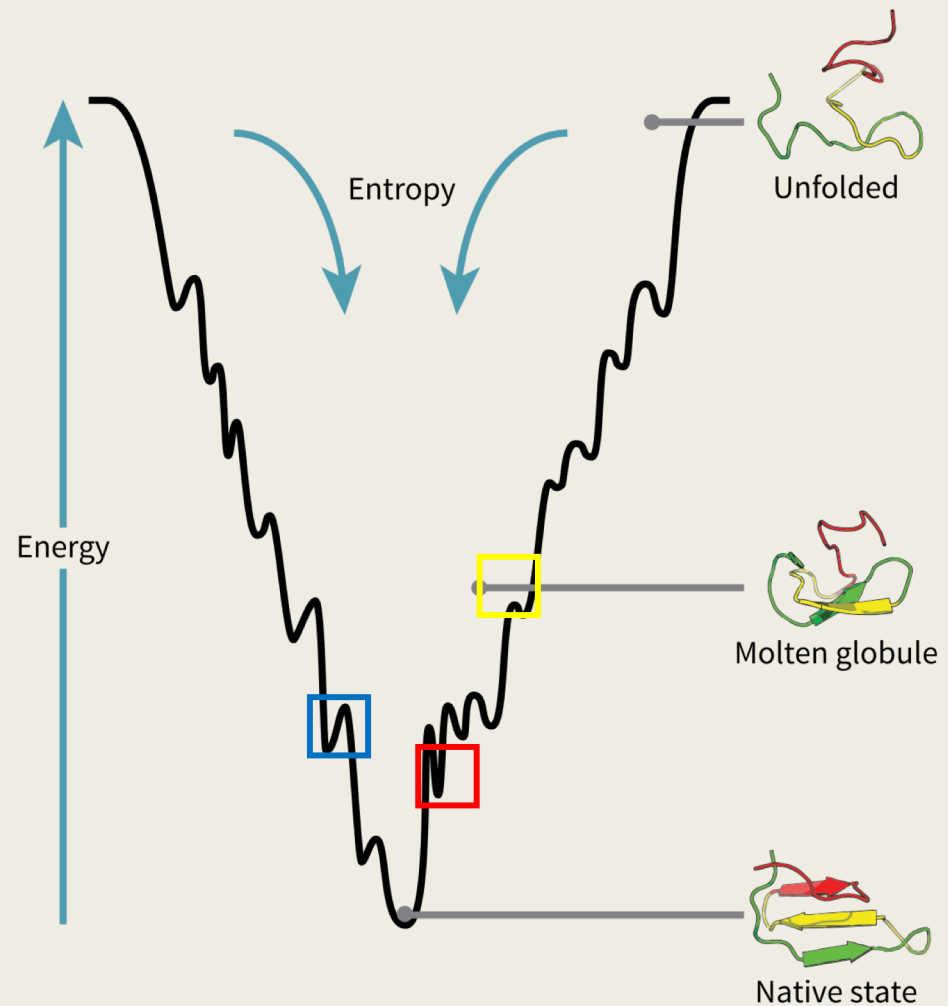
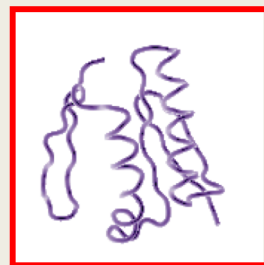
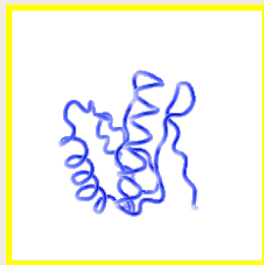
- Obtaining structure experimentally is challenging, time consuming and expensive.
- Experimental methods in structural biology are currently lagging far behind sequencing.
 - *PDB* has 139,717 entries (as of 04/25/2018), however only 77,678 are unique.
 - By contrast UniProt contains over 114,000,000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



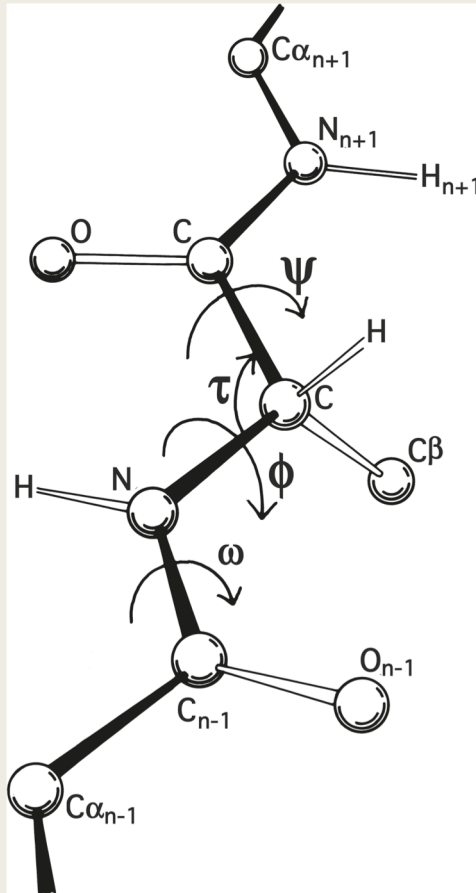
The challenge – folding landscapes are complex



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

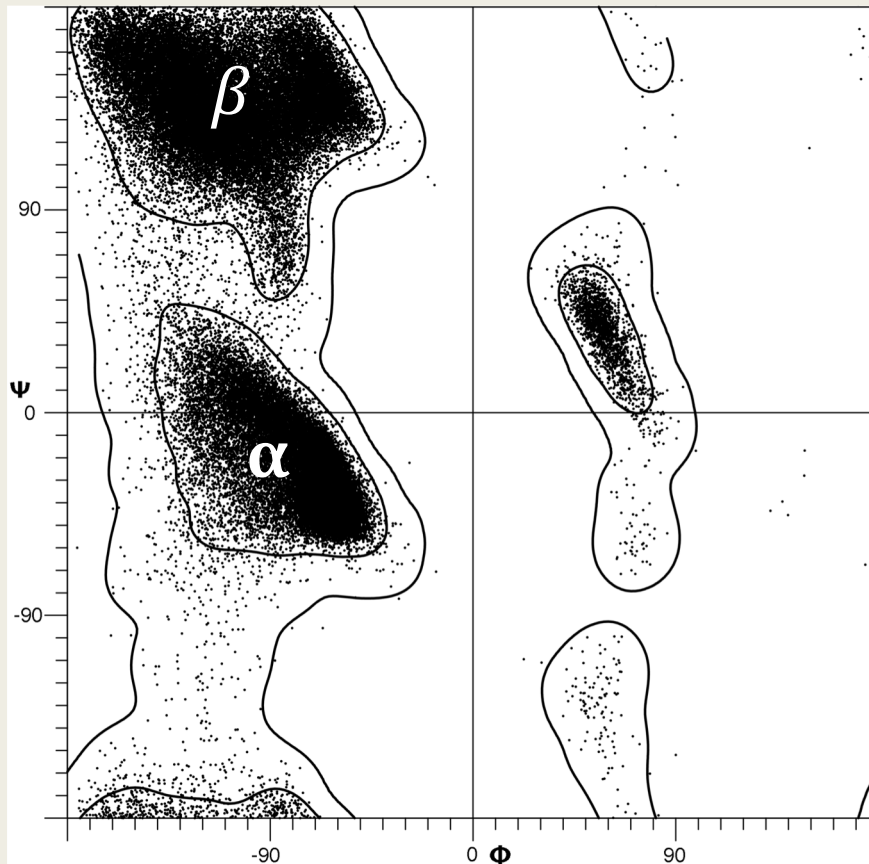


Dihedral angles of the peptide backbone are the source of almost all the interesting variability in protein conformation.

Of these phi and psi (either side of $\text{C}\alpha$) are the most important.

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

Dihedral angles and Ramachandran plot



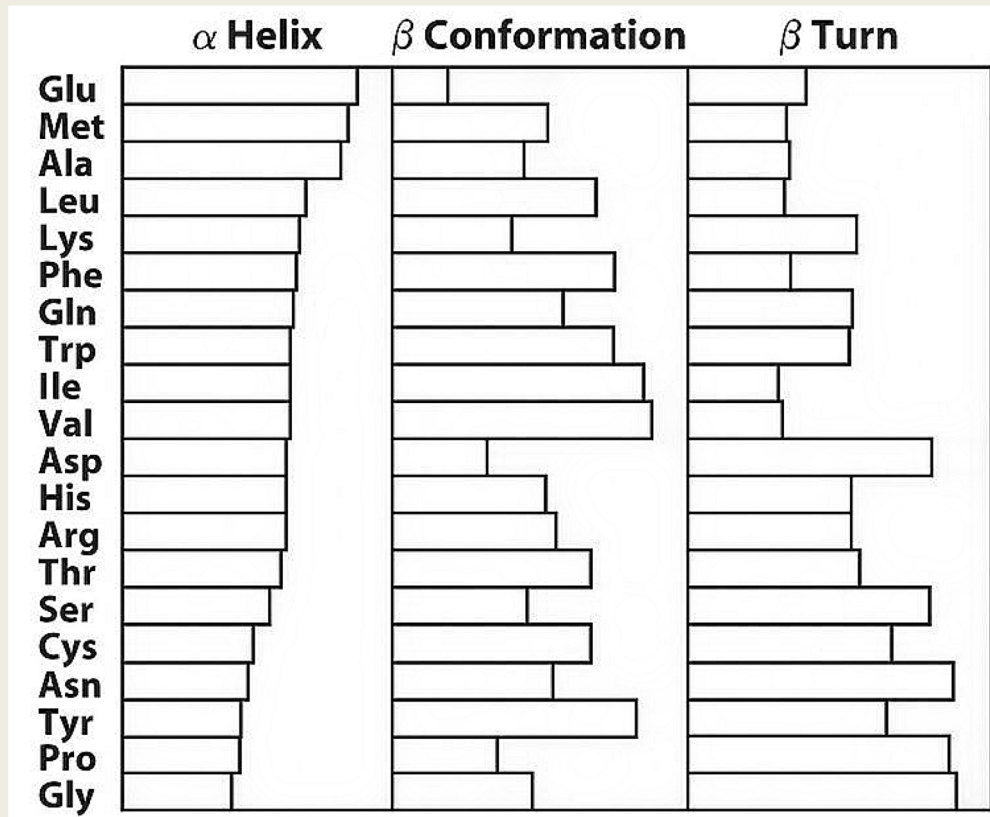
Lovell et al. 2003 *Proteins* 50:437,

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 C α) are the most important.

For amino acids other than glycine 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

The PSIPRED Protein Sequence Analysis Workbench

The PSIPRED Protein Sequence Analysis Workbench aggregates several UCL structure prediction methods into one location. Users can submit a protein sequence, perform the predictions of their choice and receive the results of the prediction via e-mail or the web.

For a summary of the available methods you can read [More...](#)

NOTE: users who need to run our methods on a large number of proteins should consider downloading our software using the menu on the left (Server Navigation -> Software Download).

The PSIPRED Team

Current Contributors David T. Jones, Daniel Buchan, Domenico Cozzetto & Kevin Bryson

Previous Contributors Tim Nugent, Federico Minneci, Anna Lobley, Sean Ward, Liam J. McGuffin

For queries regarding PSIPRED: psipred@cs.ucl.ac.uk

Input

Sequence Filter

Choose Prediction Methods

☒ PSIPRED v3.3 (Predict Secondary Structure)

☐ DISOPRED3 (Disorder Prediction)

☐ pGenTHREADER (Profile Based Fold Recognition)

☐ MEMSAT3 & MEMSAT-SVM (Membrane Helix Prediction)

☐ BioSerf v2.0 (Automated Homology Modelling)

☐ DomPred (Protein Domain Prediction)

☐ FFPred 3 (Eukaryotic Function Prediction)

☐ GenTHREADER (Rapid Fold Recognition)

☐ MEMPACK (SVM Prediction of TM Topology and Helix Packing)

☐ pDomTHREADER (Fold Domain Recognition)

☐ DomSerf v2.0 (Automated Domain Modelling by Homology)

[Help...](#)

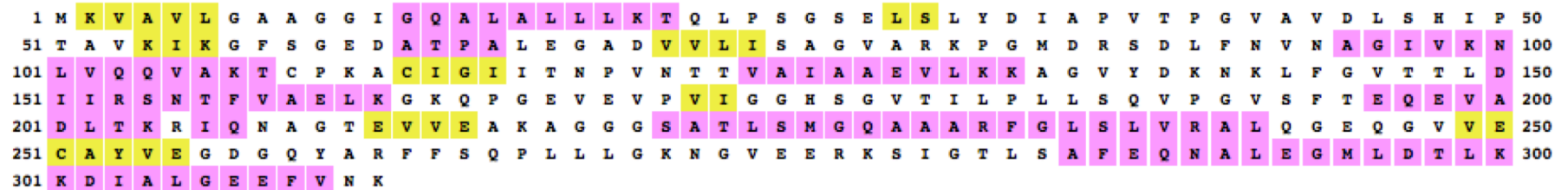
Input Sequence (Single sequence or Multiple Sequence alignments; as raw sequence or fasta format)

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

Psipred results page



Secondary Structure Map

Feature predictions are colour coded onto the sequence according to the sequence feature key shown below.





KEY Helix Sheet Disordered

Annotations M L E

Conf:  

Pred: CEEEEEECCCCCHHHHHHHHHHCCCCCEECCCCCCCCC
AA: MKVAVLGAAGGIGQALALLLKTQLPSGSELSLYDIAPVTP

10 20 30 40

Conf:  

Pred: CCCCCCCCCCCCCCEEECCCCCHHHHCCCCCEEECCCCC
AA: GVAVDLSHIPTAVKIKGFSGEDATPALEGADVVLISAGVA

50 60 70 80

Disordered protein binding

Domprcd Boundary

DomSSEA Boundary

E

A

D

Conf:] [

Pred: —————→

Pred: CCCCCCCCCCCCCCHHHHHHHHHHHHHHHHCCCCEEEECCCC

AA: RKPGMDRSDLFNVNAGIVKNLVQQVAKTCPKACIGIITNP

90 100 110 120

Conf:] [

Pred: —————→

Pred: CCCCHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCHHHHHHHHHHHHH

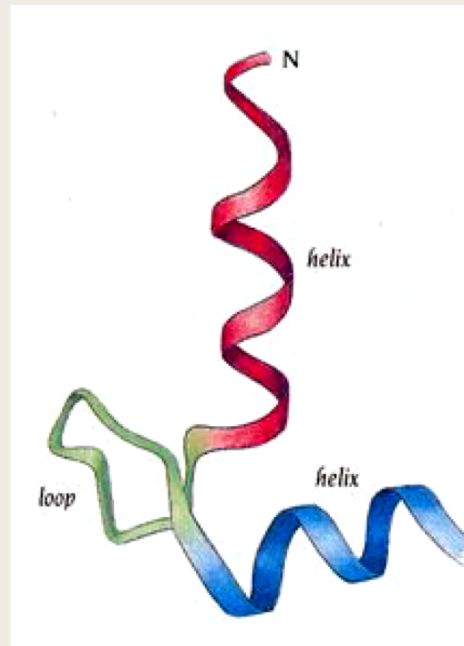
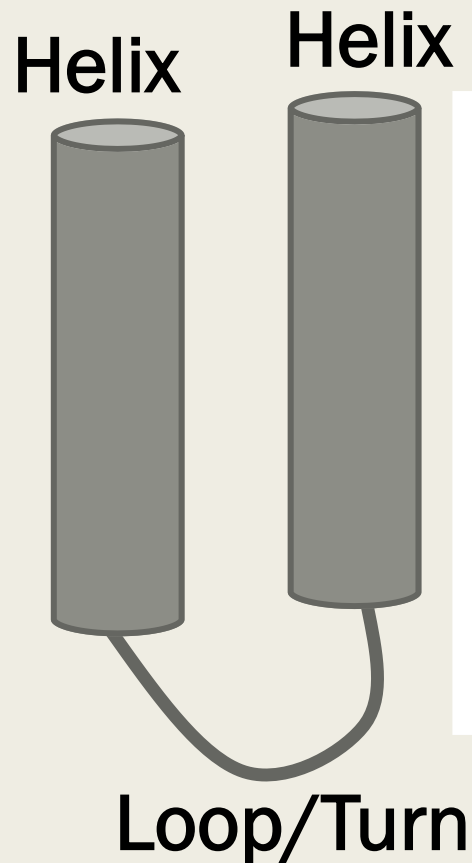
AA: VNTTVAIAAEVLKKGAVYDKNKLFGVTTLDIIRSNTFVAE

130 140 150 160



Conserved motifs

Alpha Helical Motifs



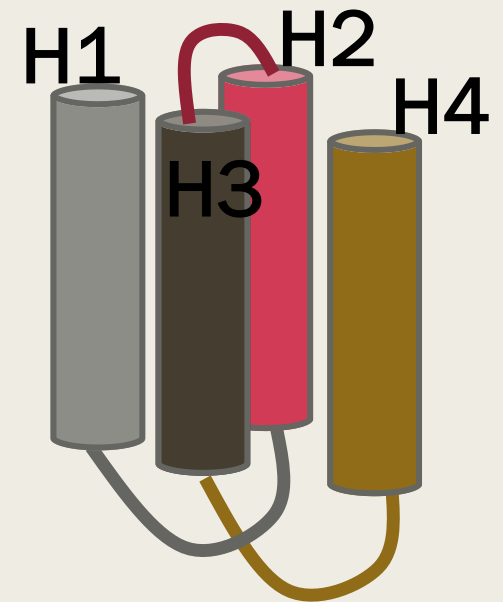
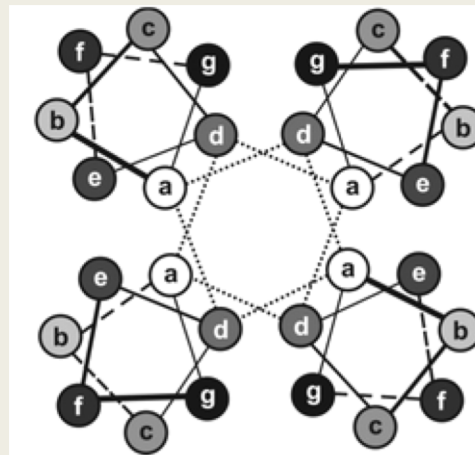
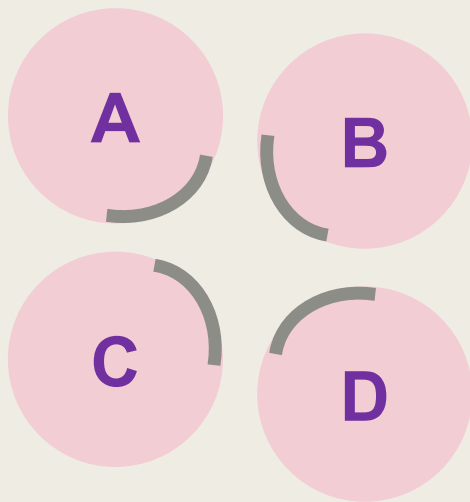
Stretches of residues with propensity to form alpha helices can be mapped onto a helical wheel.

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

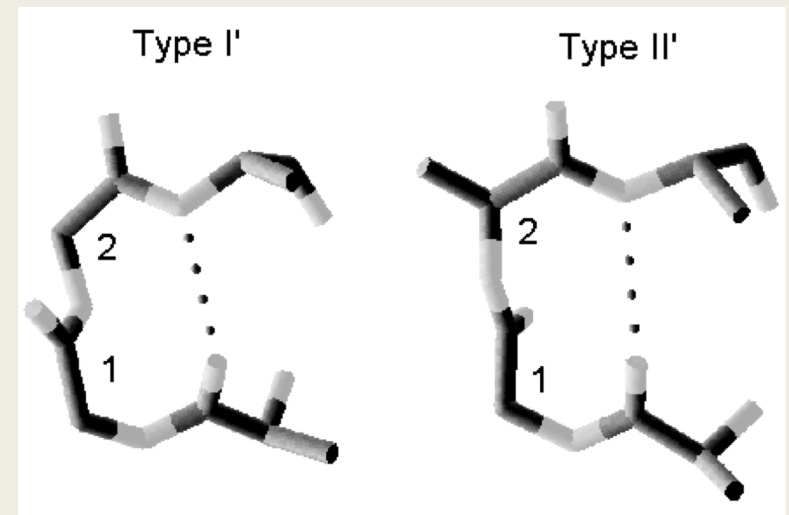
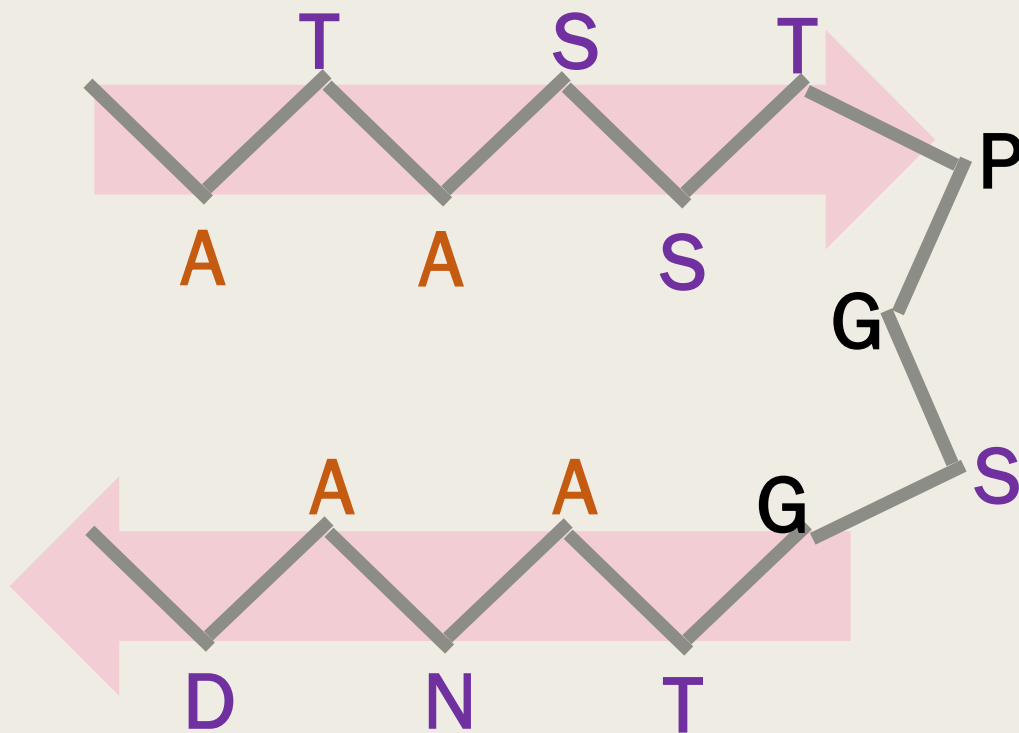
Helices can display different properties on different faces.

Helix – Helix Combinations

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



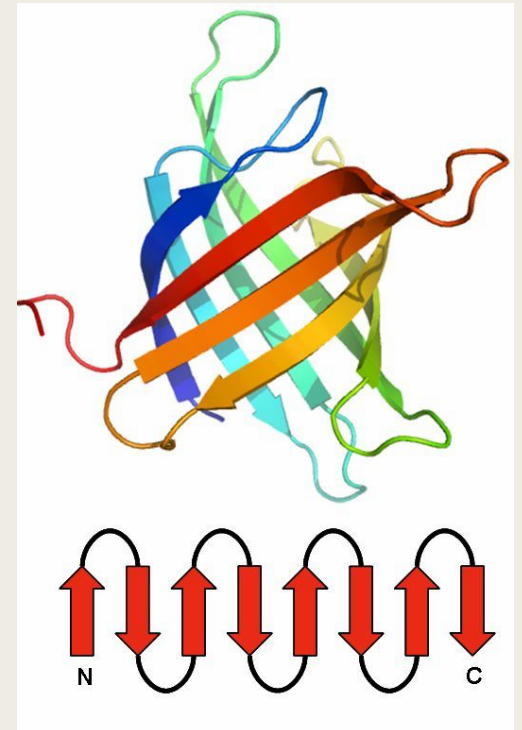
Beta Turn (Hairpin)



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.

The diagram shows a series circuit with a yellow battery at the bottom. The circuit is divided into two parallel branches. The left branch contains three red resistors, and the right branch contains three blue resistors. A label 'C' is at the top left, and a label 'N' is at the bottom center, pointing to the negative terminal of the battery.

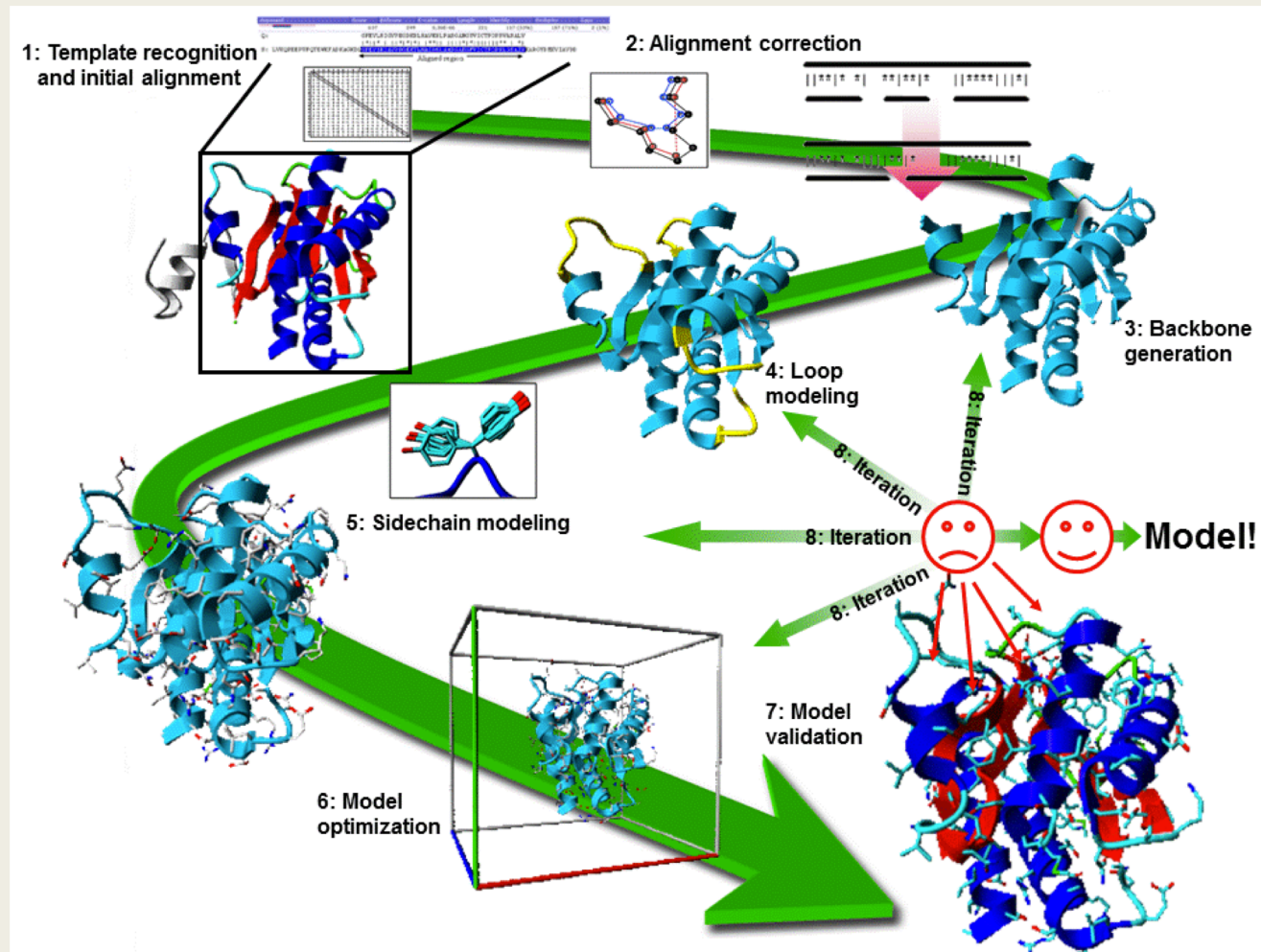
3D ribbon diagram of the protein structure, highlighting the Gly-rich loop (yellow), the H1 helix (green), and the beta1 strand (green). The beta2-Asp strand is shown in pink. A chemical structure of a ligand is shown bound to the Gly-rich loop.

**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.
- Exploite existing 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



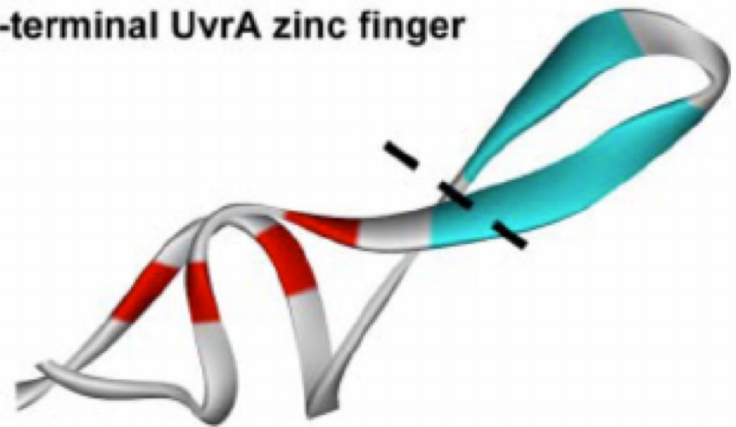
An example alignment

A: ECOLI	GRFSFNVRGGRCEACQGDGVIKVEMHFLPDIYVP---CDQCKGKRYNRETLE
RHIME	GRFSFNVKGGRCEACQGDGVIKIEMHFLPDVYVT---CDVCHGKRYNRETLD
TREPA	GRFSFNVPGGRCEHCKGDGVITIEMNFLPDVYIT---CDVCHGTRFNRETLA
HELPJ	SRFSFNVKGGRCEKCKQGDGDIKIEMHFLPDVLVQ---CDSCKGAKYNPQTLE
BCACA	GRFSFNVKGGRCEACHGDGIIKIEMHFLPDVYVP---CEVCHGKRYNRETLE
ZnG A	GRFSFNVKGGRCEACHGDGII-----G-----VP---CEVCHGKRYNRETLE
Ydj1	GRGGKKGAVKKCTSCNGQGIIKFVTRQMGPIMIQRFQTECDVCHGTGDIIDPKD

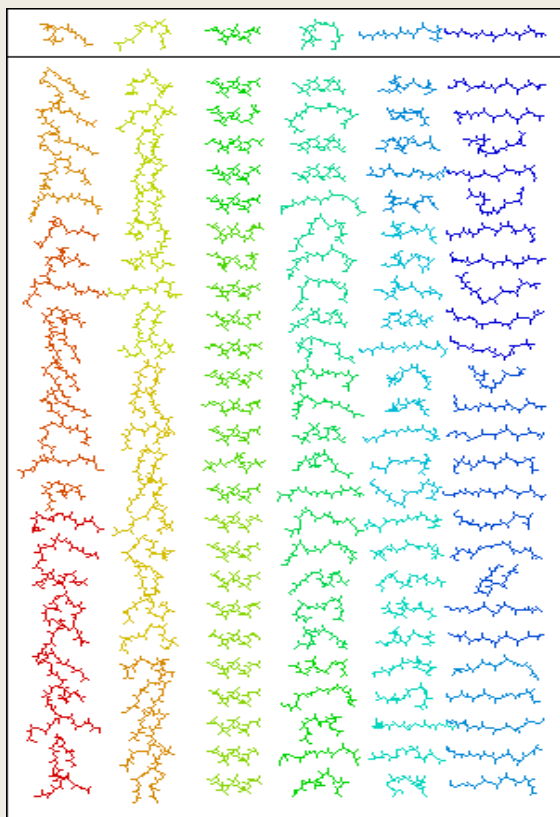
B: Ydj1



C: C-terminal UvrA zinc finger



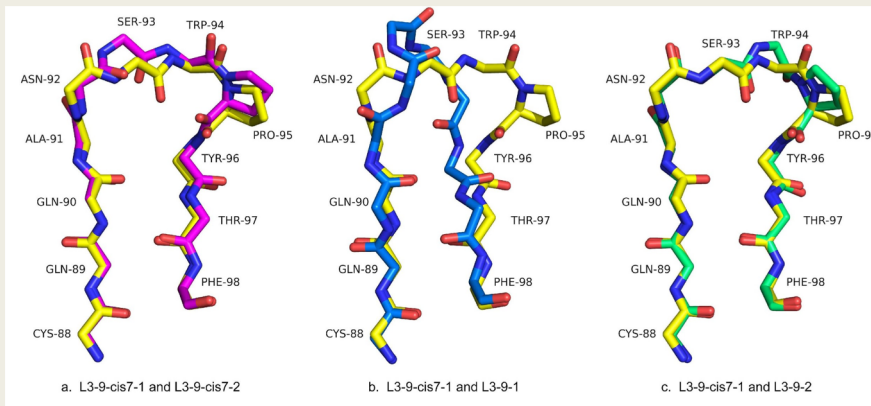
Fragment insertion to fill in gaps



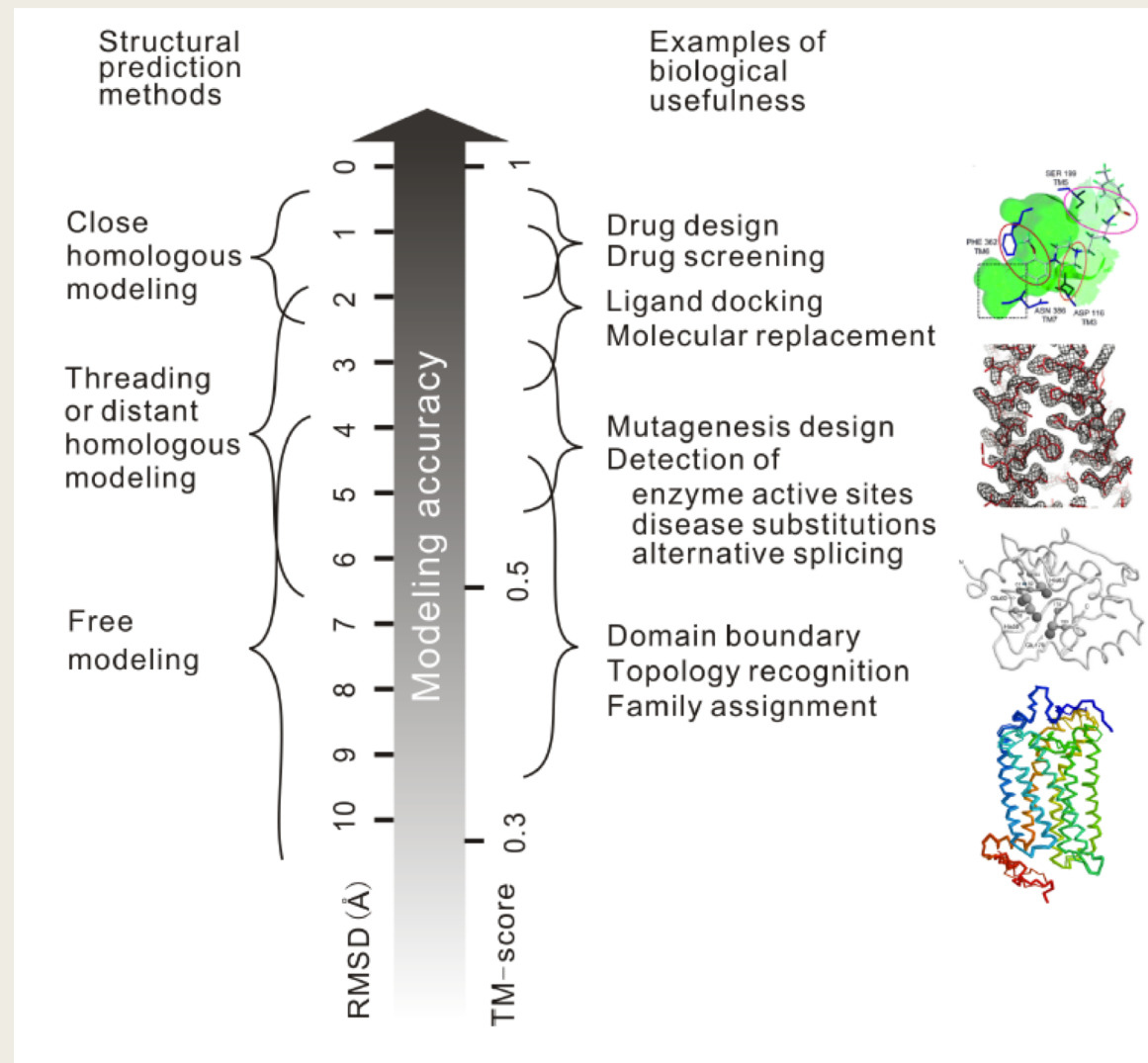
Local interactions: fragments

- 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!



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.*