Biology: Bigger structures, Greater confusion.

Dr David C Briggs
Biophysics Section, Department of Life Sciences
Imperial College London

Running Order
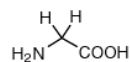
1. Proteins

1. Dealing with Size and Complexity

1. Dealing with different sources of information

1. Conclusions
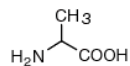
Proteins: Nature's Chemists.
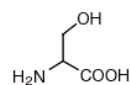
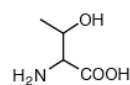Most of structural biology is concerned with proteins.

Small

Glycine (Gly, G)
MW: 57.05

Alanine (Ala, A)
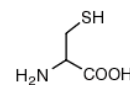MW: 71.09
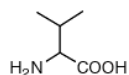
Nucleophilic

Serine (Ser, S)
MW: 87.08, pK_a ~ 16
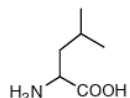
Threonine (Thr, T)
MW: 101.11, pK_a ~ 16
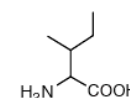
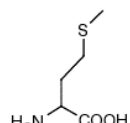Cysteine (Cys, C)
MW: 103.15, pK_a = 8.35
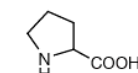
Hydrophobic

Valine (Val, V)
MW: 99.14

Leucine (Leu, L)
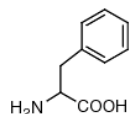MW: 113.16

Isoleucine (Ile, I)
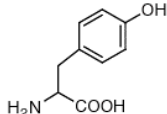MW: 113.16

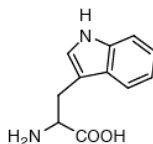Methionine (Met, M)
MW: 131.19

Proline (Pro, P)
MW: 97.12

Aromatic
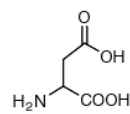
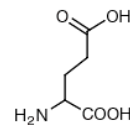Phenylalanine (Phe, F)
MW: 147.18

Tyrosine (Tyr, Y)
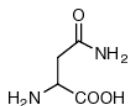MW: 163.18

Tryptophan (Trp, W)
MW: 186.21

Acidic

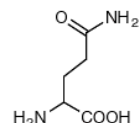Aspartic Acid (Asp, D)
MW: 115.09, pK_a = 3.9

Glutamic Acid (Glu, E)
MW: 129.12, pK_a = 4.07

Amide

Asparagine (Asn, N)
MW: 114.11

Glutamine (Gln, Q)
MW: 128.14

Basic
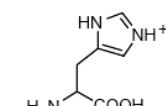
Histidine (His, H)
MW: 137.14, pK_a = 6.04
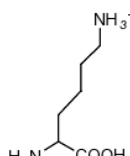
Lysine (Lys, K)
MW: 128.17, pK_a = 10.79
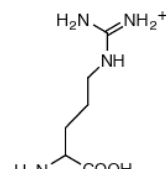
Arginine (Arg, R)
MW: 156.19, pK_a = 12.48
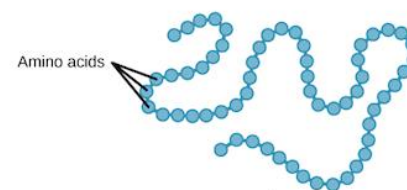
Amino acids

Primary protein structure
sequence of a chain of animo acids

Pleated sheet — Alpha helix —
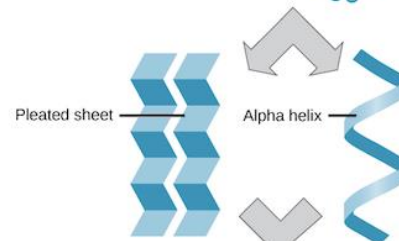
Secondary protein structure
hydrogen bonding of the peptide backbone causes the amino acids to fold into a repeating pattern

Tertiary protein structure
three-dimensional folding pattern of a protein due to side chain interactions
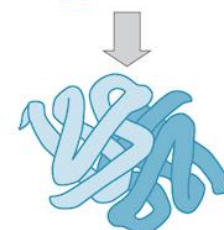
Quaternary protein structure
protein consisting of more than one amino acid chain

Protein size is not compatible with all atom representation
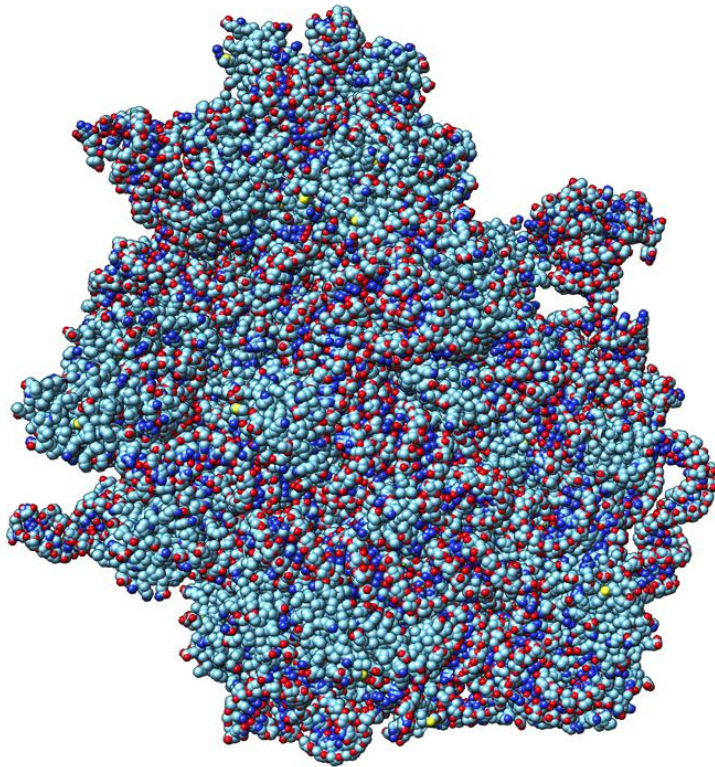
Staph Aureus Ribosome
2.3 MDa.
3 rRNA chains.
~50 protein chains.
>57000 non-H atoms
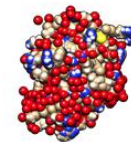
Lysozyme
14.3 kDa
1 protein chain – 129 amino acids

1001 non-H atoms

Challenge 1: Simplifying things.

1. VDW radius
2. Stick
3. Wireframe
4. Cartoon
5. Cartoon mainchain
   Stick sidechains

There is no way of displaying an entire with structure whilst preserving information content
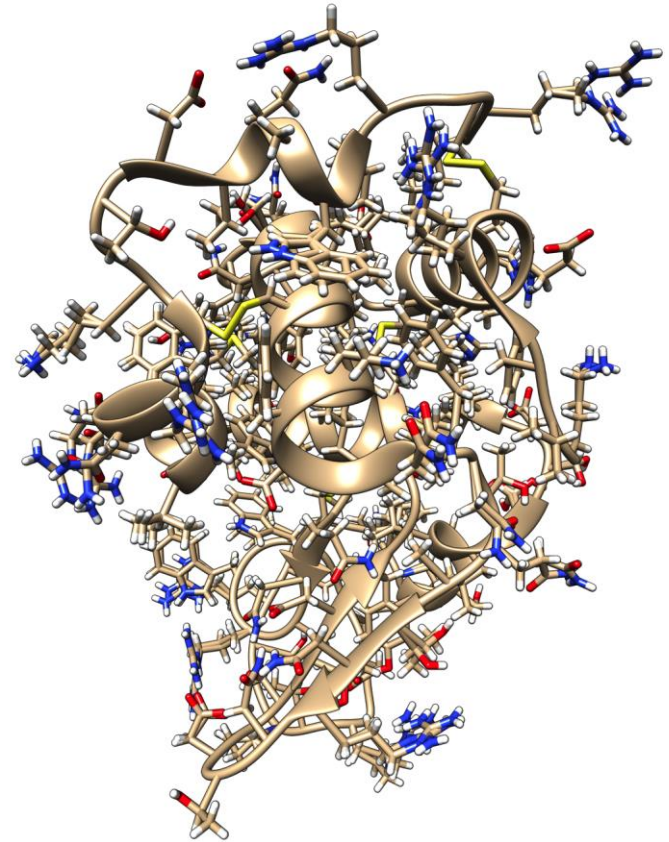
Challenge 1: Simplifying things.

Solution 1: Sacrifice information content for clarity.

Challenge 1: Simplifying things.

Solution 2: Schematics
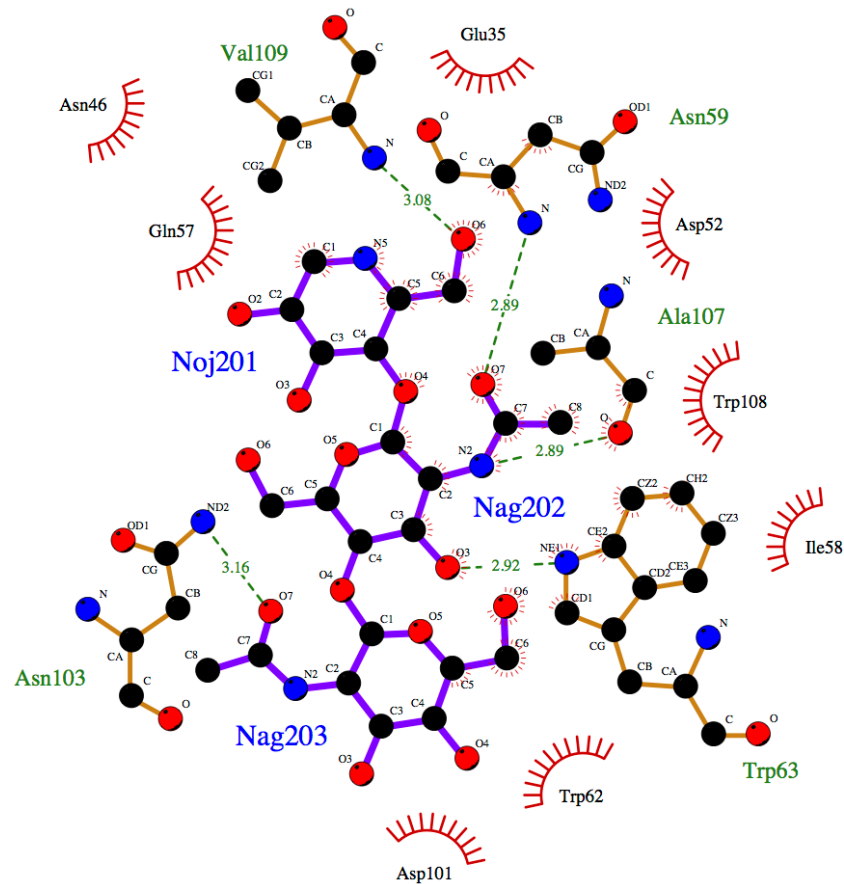
LigPlot (shown), FLEV

Challenge 2: Buried active sites.
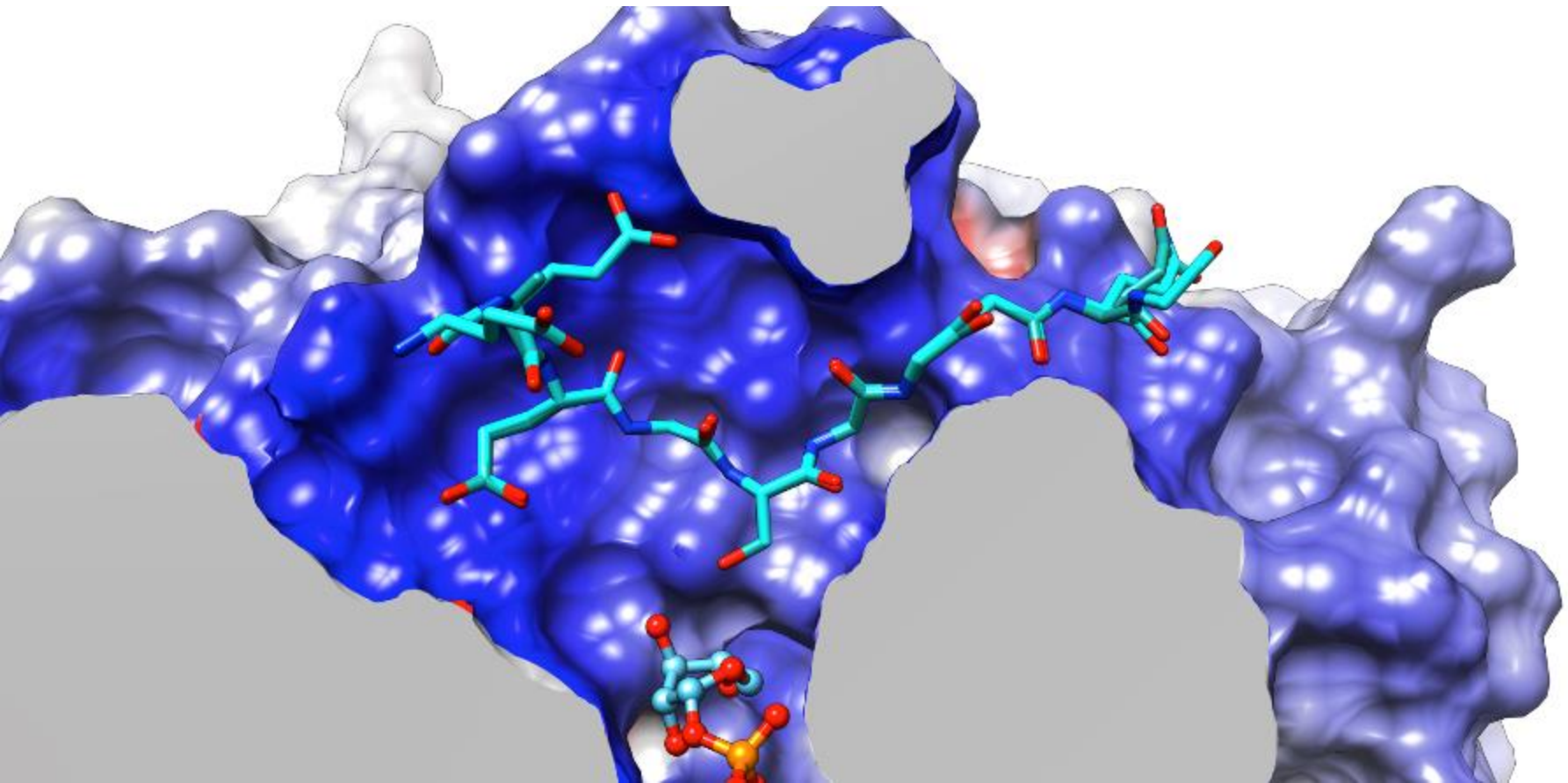
Something a little more difficult…

Nascent Peptide-mediated Ribosome Stalling with Antibiotics

MDFF model of ERY-bound ribosome with the ErmBL nascent chain

Erythromycin (ERY)

Telithromycin (TEL)

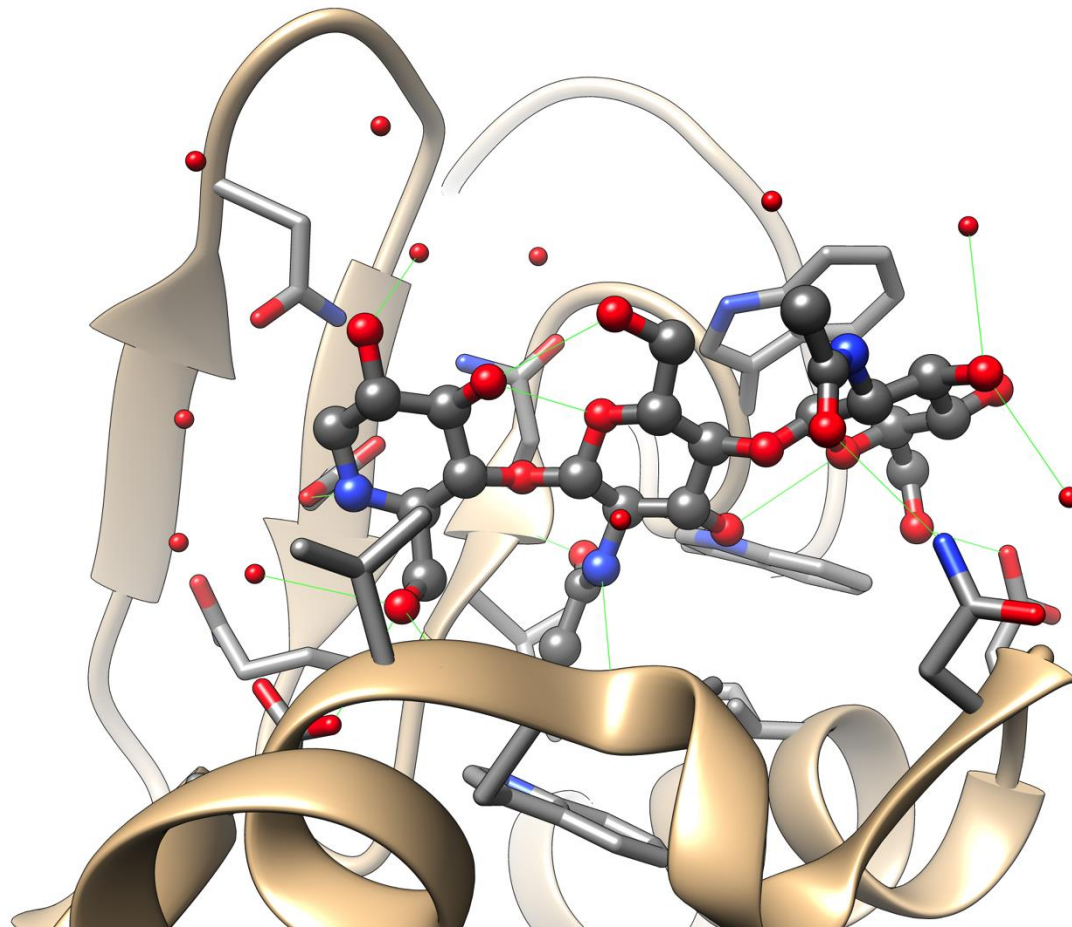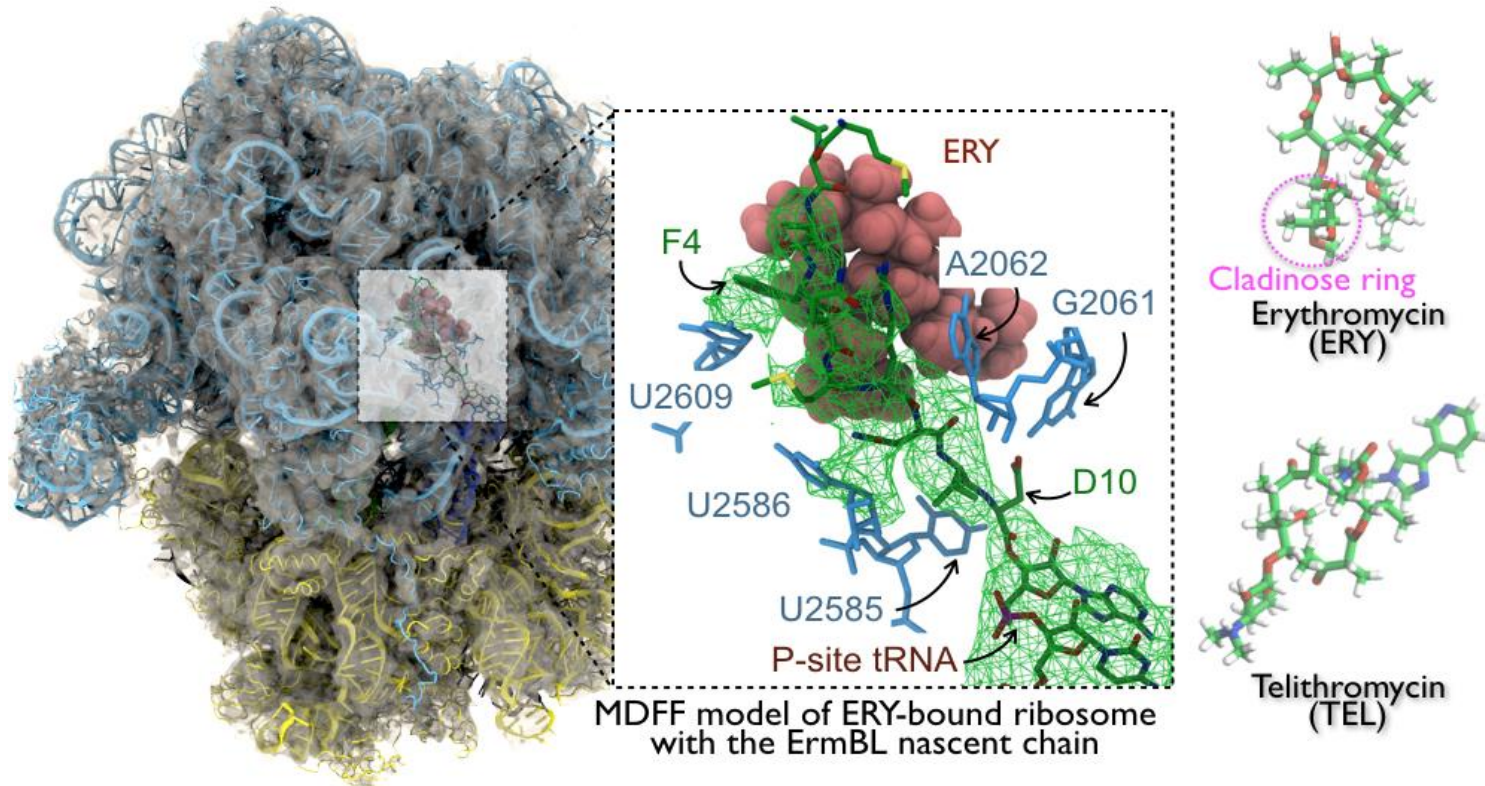*Combination of drugs and Single Mutations at the C-terminus of nascent ErmBL protein tunes stalling:*

	Asp (native)	Glu	Tyr
TEL	Stall	Stall off	Stall
ERY	Stall	Stall	Stall off

https://www-s.ks.uiuc.edu/Research/ribosome/

Information type

- The Protein Data Bank
- ~130,000 entries.
- Structures determined by multiple techniques
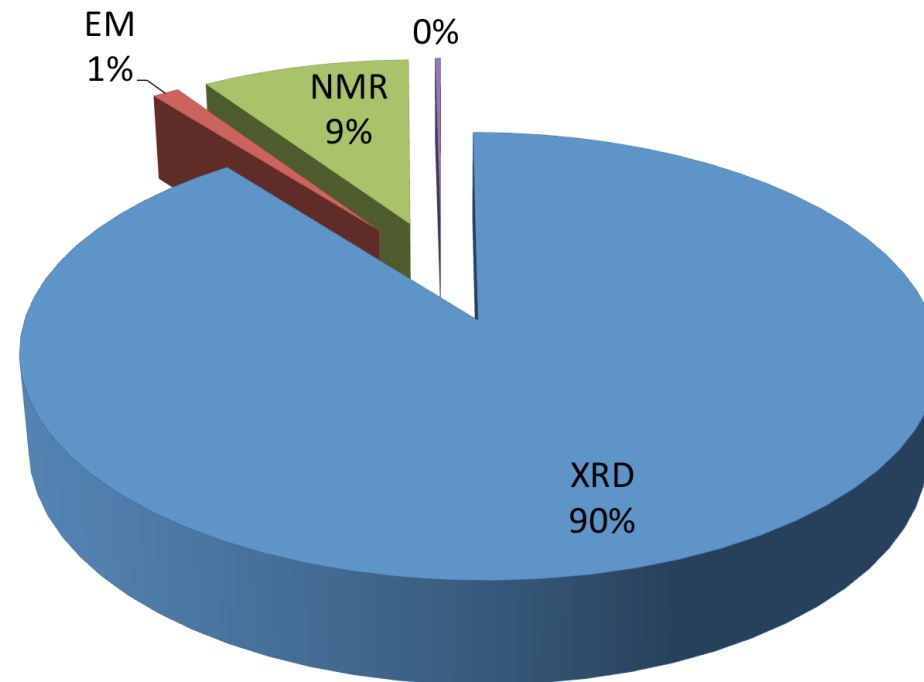
PDB entries by technique

The Protein Data Bank

Different techniques = Different information content.

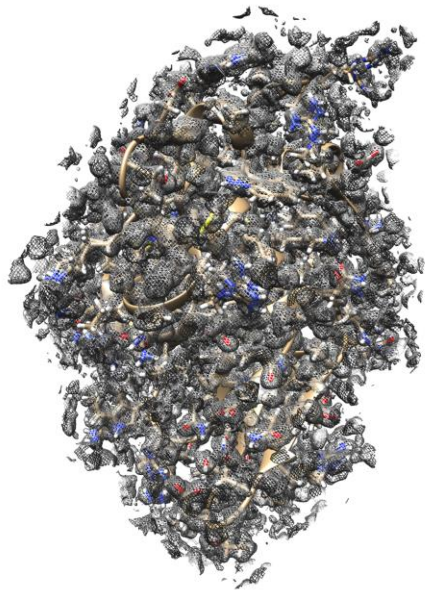| Technique	Pros	Cons
XRD	Best resolution (0.48Å, mean ~2.5Å) Best observation : parameter ratio	Constrained by packing Crystals
NMR	Easier to model dynamics In solution	Low obs:param ratio Size constraints
EM	Resolution improving rapidly! (<3Å)	Easier with bigger molecules
SAXS	In solution	Low resolution (~20Å)

The effect of resolution...

0.7Å
Outstanding XRD
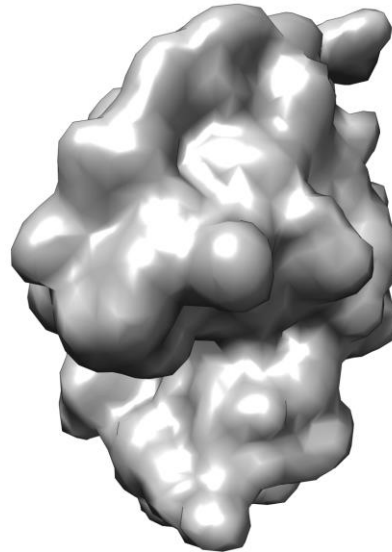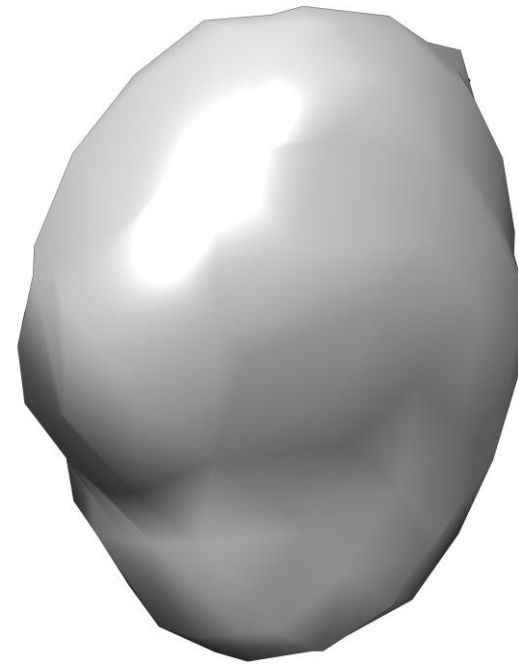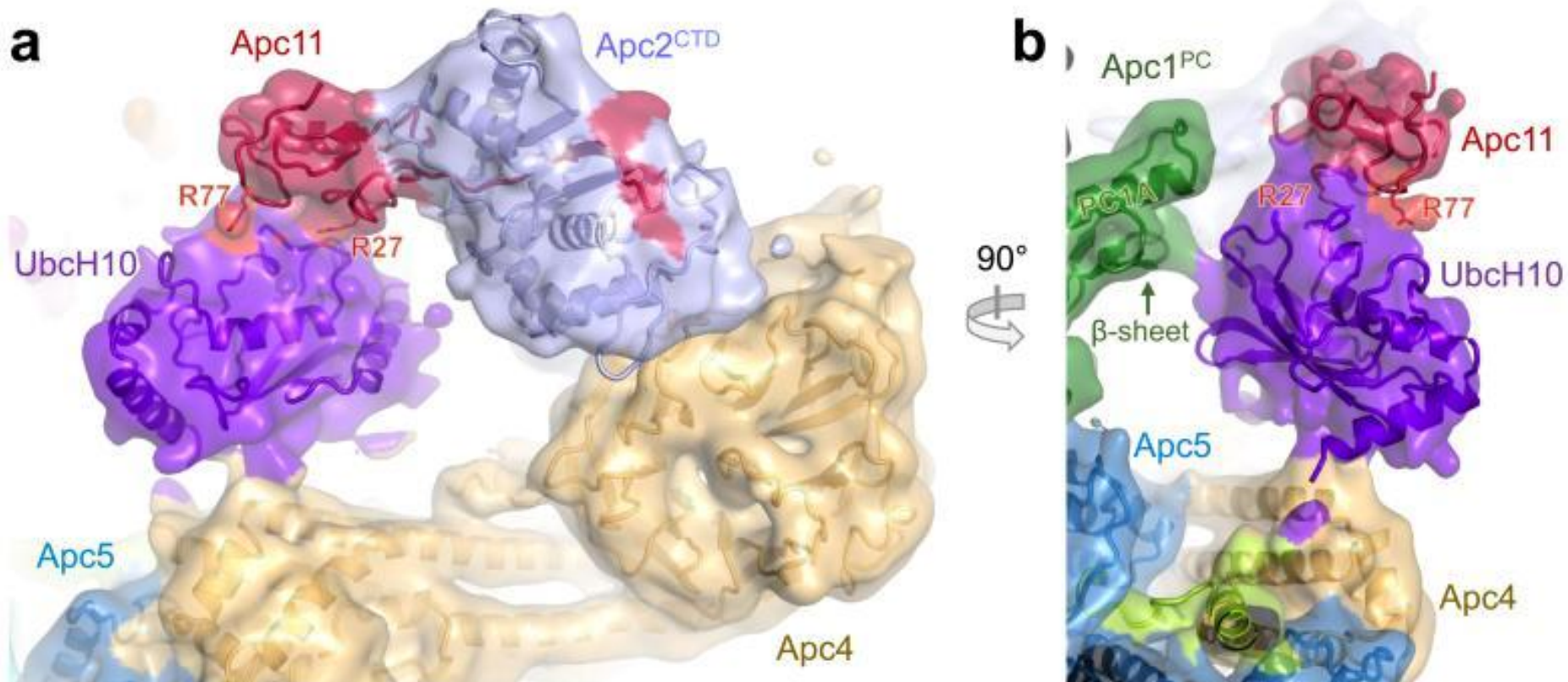
5Å
(NMR or 'new' EM)

20Å
(SAXS or 'old' EM)

Integrating different types of information.
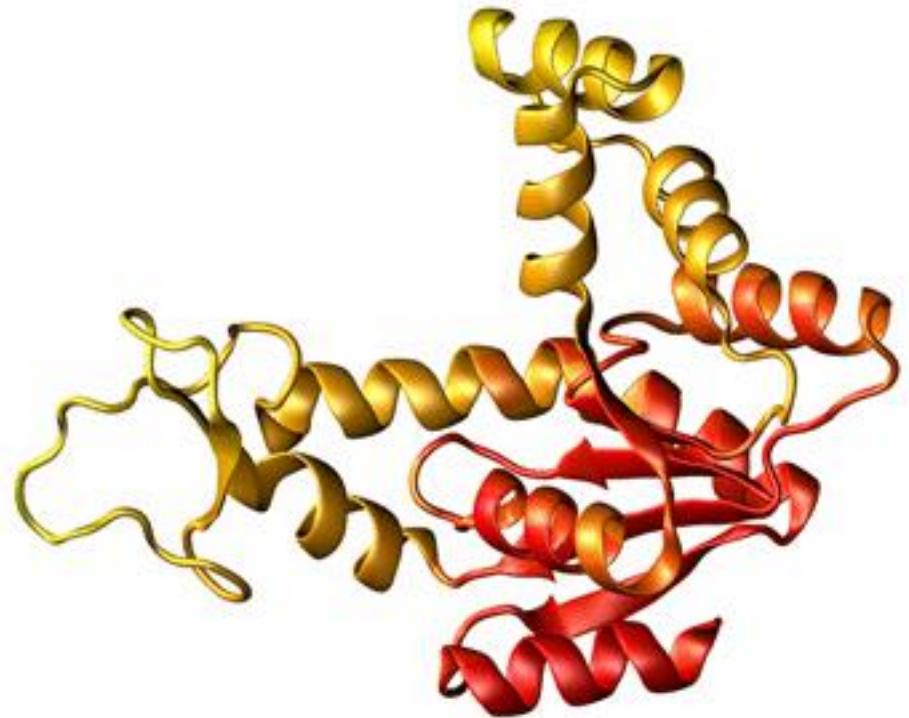
Challenge 2: Dynamics

Proteins 'breathe'

They are not rigid, static objects.

Most techniques lose information
About dynamism.

Challenge 2: Ensembles

Biology…

More Complexity…

Size
Inherent in the system

Resolution
A technical issue
&
inherent in the system

Dynamics
inherent in the system
&
data not always
retained

…Less Complexity

Limited chemical
repertoire
20 amino acids
5 nucleotide bases

Limited conformational
repertoire
e.g. Ramachandran plot
& preferred rotamers…

Conclusions

Bad news...

Above a certain size/complexity, information content has to be sacrificed for clarity.

...Good news

... plenty of room for artistic interpretation of macromolecular structures.