UNIVERSITY OF CAMBRIDGE

Protein Structure Determination

Day 9: Thursday 30th March

Experimental structure determination

- Yesterday we heard about how to look at structures

- Today we'll learn how to experimentally determine a structure
  - Techniques for determining structures
  - How to collect and process structural data
  - How to build a new structure
  - How to judge the quality of a structure

Why study protein structure?

- Basic biology
  - Unique insights into how biological systems function

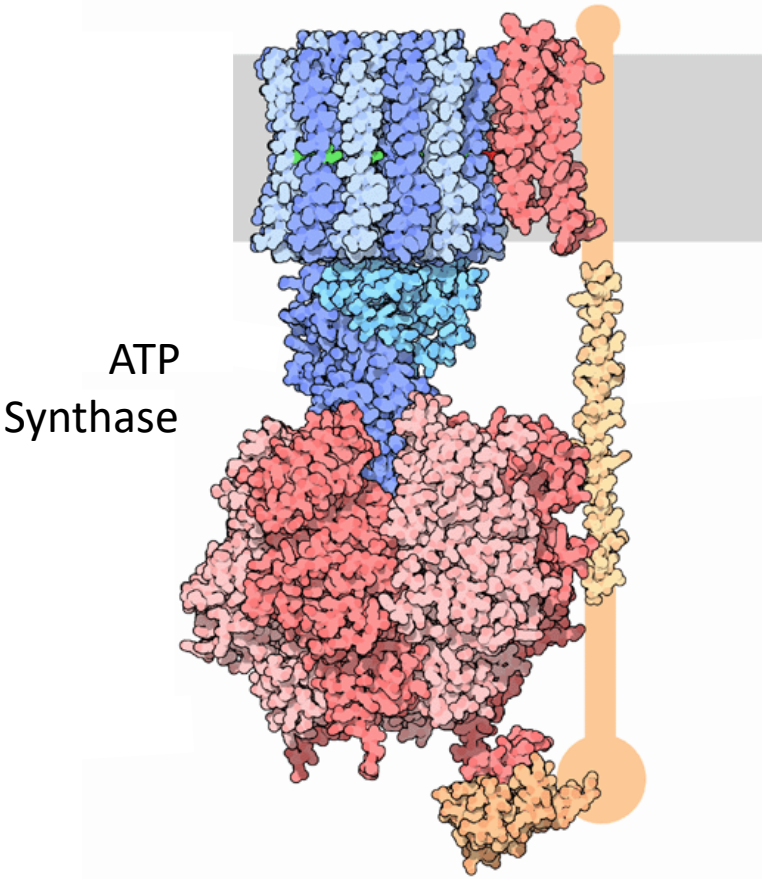
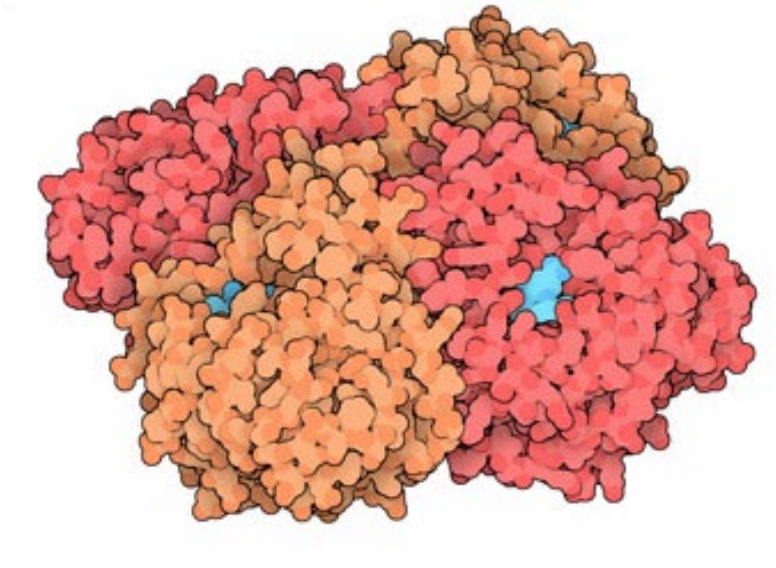


Image: PDB-101: Molecule of the Month: ATP Synthase (rcsb.org)

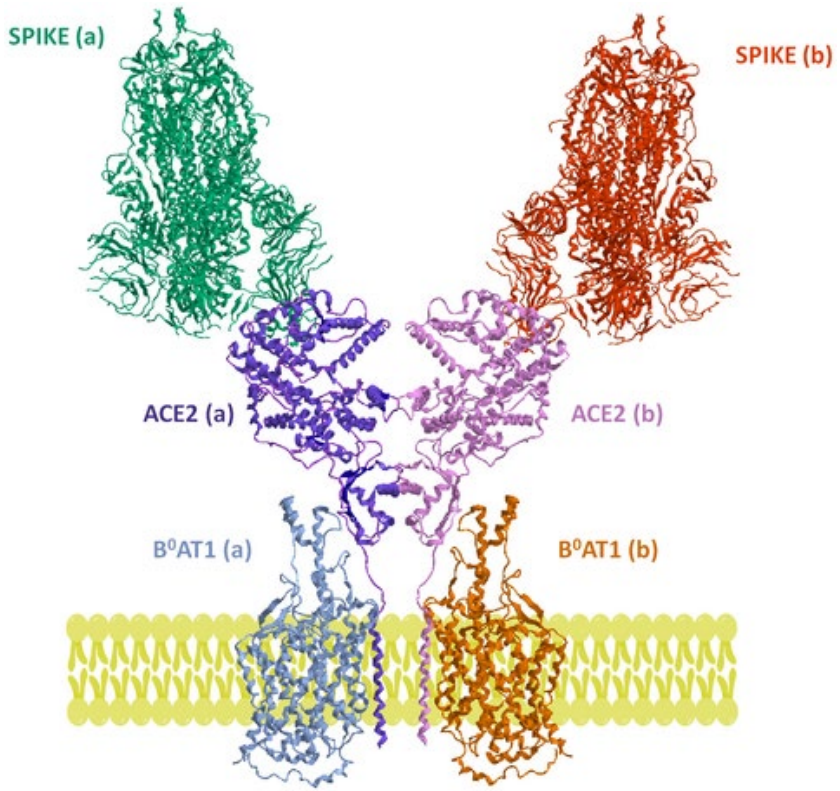
Why study protein structure?

- Basic biology
  - Unique insights into how biological systems function
- Structure-based drug design
  - Relenza and Tamiflu were designed using structure of flu protein (neuraminidase)



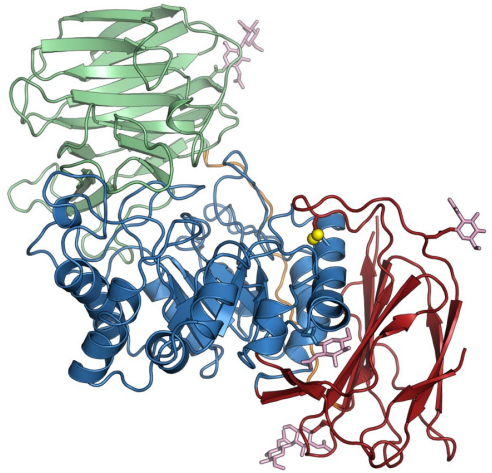
Why study protein structure?

- Vaccine development
  - Can understand how new sequence variants might alter virus binding to its receptor
  - SARS-Cov2 spike protein to the ACE2 receptor

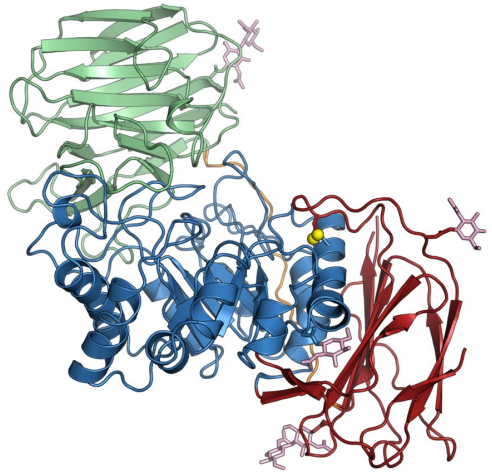


Determining a 3D structure

- Three main experimental approaches:
  - X-ray crystallography
  - Electron microscopy (cryo-EM)
  - NMR



Determining a 3D structure



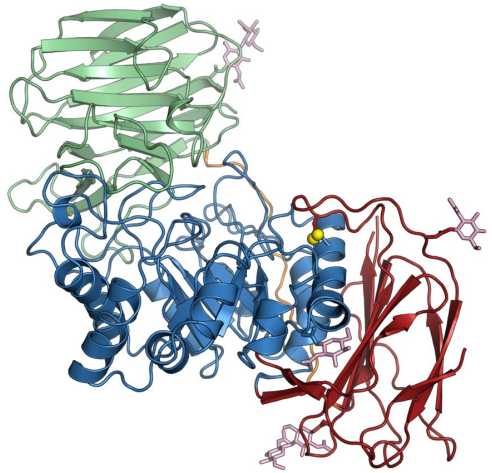
- Three main experimental approaches:
  - X-ray crystallography
  - Electron microscopy (cryo-EM)
  - NMR

- All result in atomic co-ordinates that describe the position of atoms
- These co-ordinates are deposited in a public repository called the Protein Data Bank (PDB)

RCSB PDB
PROTEIN DATA BANK

Determining a 3D structure

- Three main experimental approaches:
  - X-ray crystallography
  - Electron microscopy (cryo-EM)
  - NMR

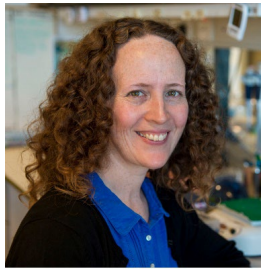


- All result in atomic co-ordinates that describe the position of atoms
- These co-ordinates are deposited in a public repository called the Protein Data Bank (PDB)
- All techniques have to tackle the problem of how *very small* proteins are!

Proteins are small

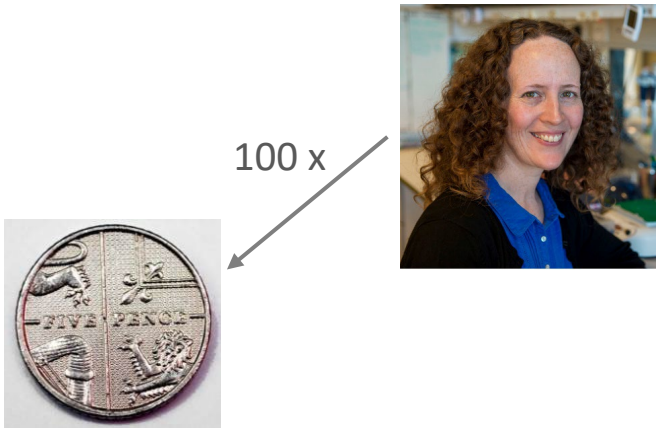
Proteins are small

- Me: ~1.5×10^0 m



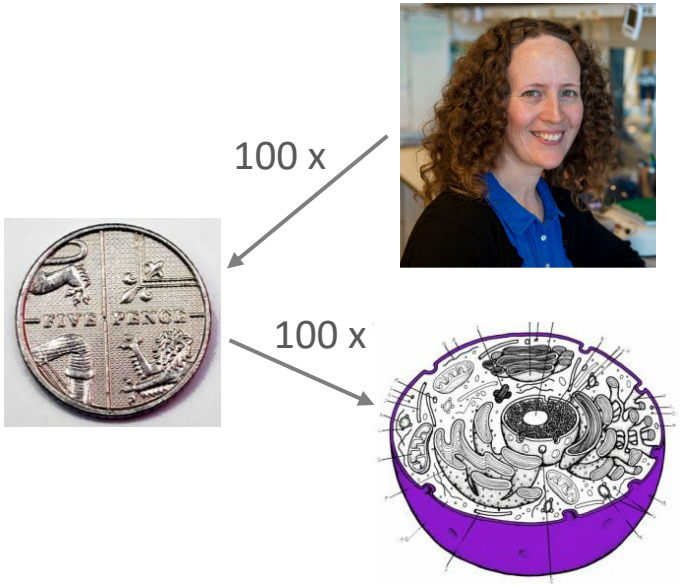
Proteins are small

- Me: ~1.5×10^{0} m
- 5p piece: ~1.8×10^{-2} m



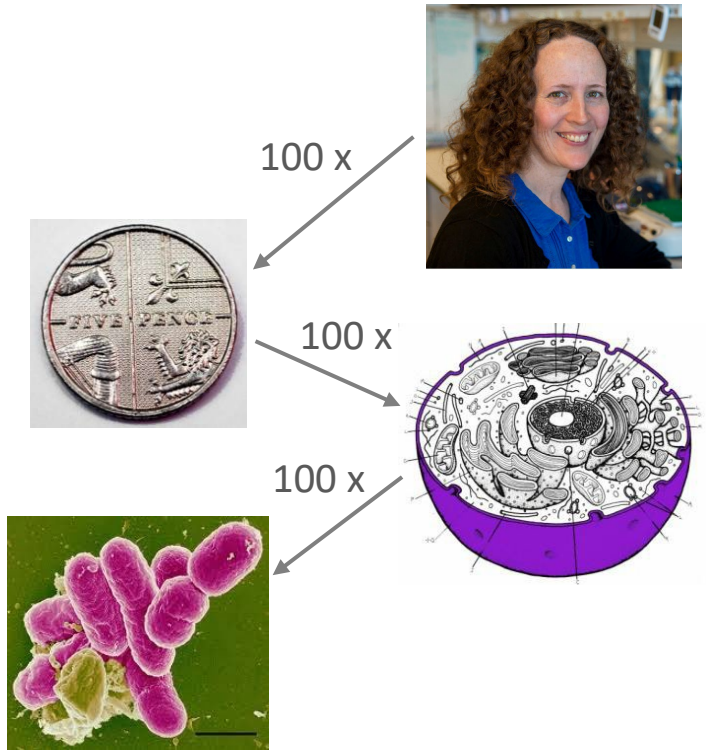
Proteins are small

- Me: ~1.5×10^{0} m
- 5p piece: ~1.8×10^{-2} m
- Eukaryotic cell: ~1×10^{-4} m



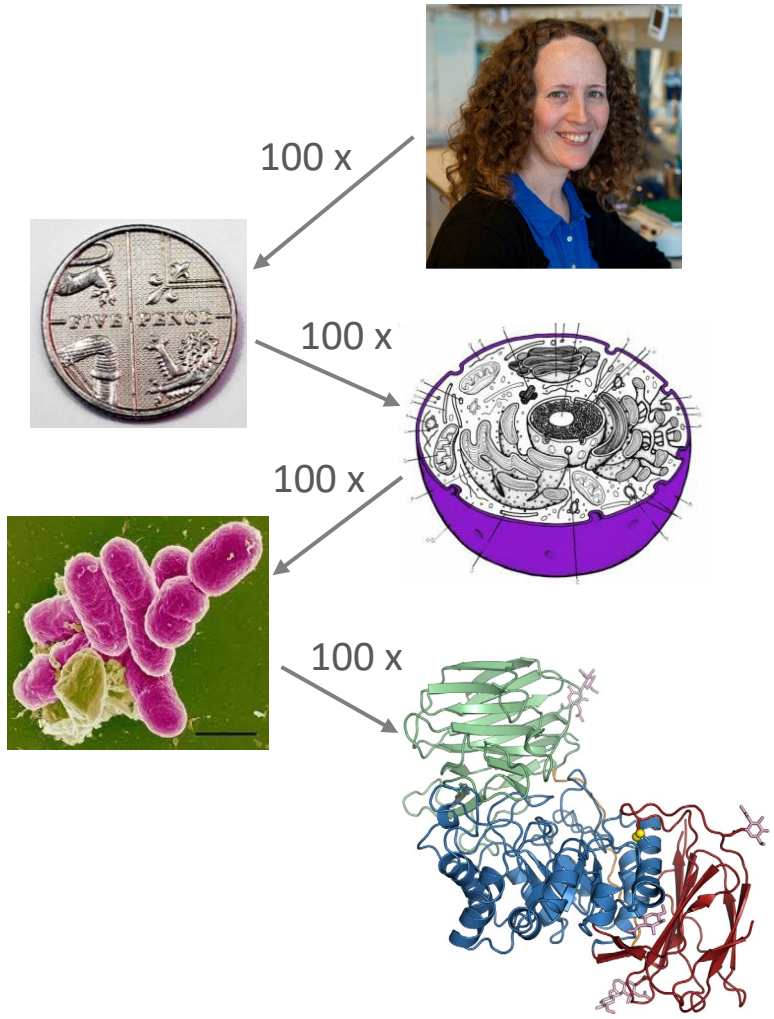
Proteins are small

- Me: ~1.5×10^{0} m
- 5p piece: ~1.8×10^{-2} m
- Eukaryotic cell: ~1×10^{-4} m
- Bacterial cell: ~1×10^{-6} m



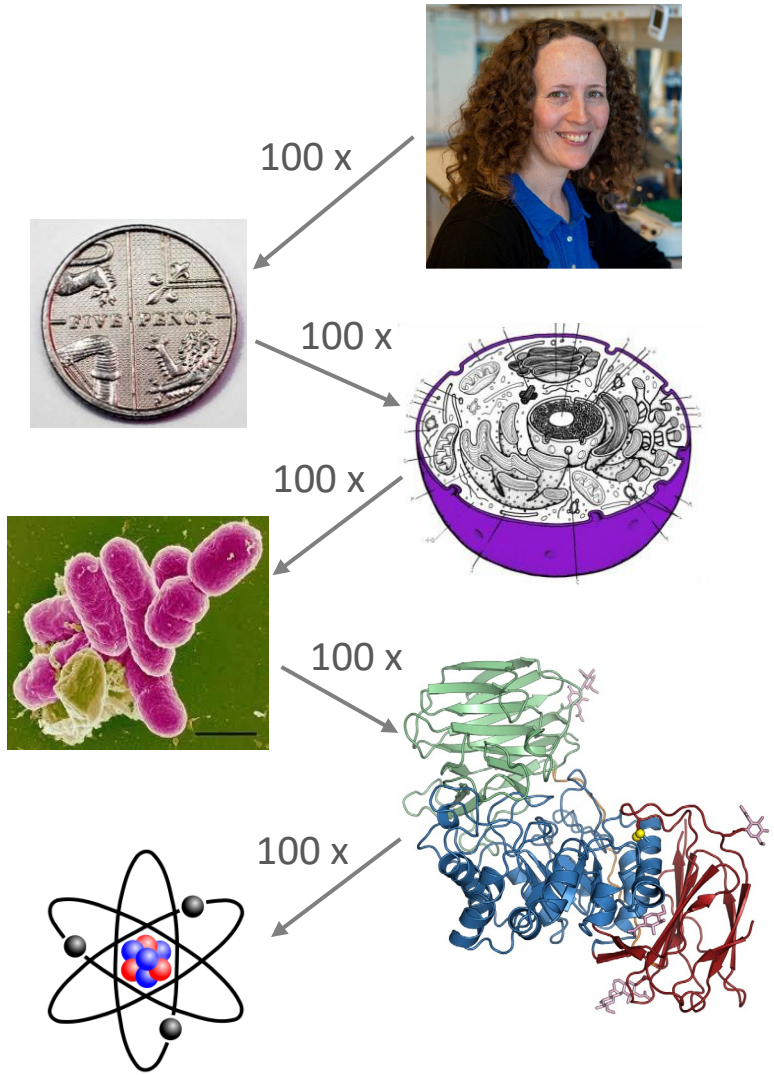
Proteins are small

- Me: ~1.5×10^{0} m
- 5p piece: ~1.8×10^{-2} m
- Eukaryotic cell: ~1×10^{-4} m
- Bacterial cell: ~1×10^{-6} m
- **Proteins: 1×10^{-9} to 1×10^{-8} m**

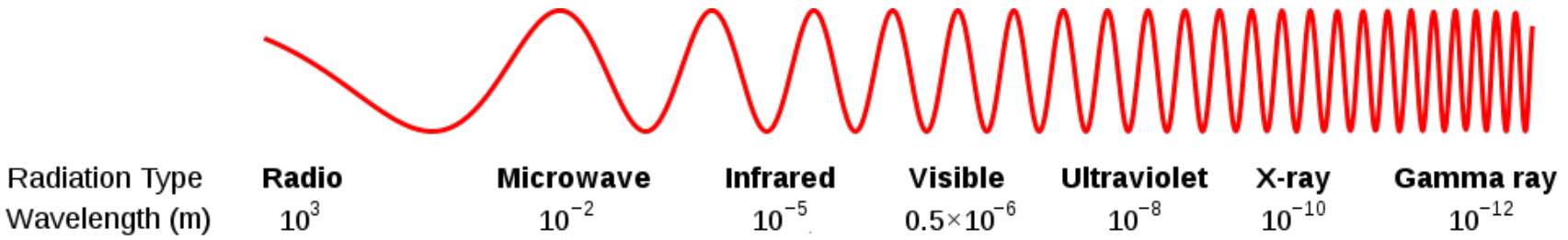


Proteins are small

- Me: ~1.5×10^{0} m
- 5p piece: ~1.8×10^{-2} m
- Eukaryotic cell: ~1×10^{-4} m
- Bacterial cell: ~1×10^{-6} m
- **Proteins: 1×10^{-9} to 1×10^{-8} m**
- Atoms: ~0.4 x 10^{-10} m



Proteins are so small we can’t use visible light to see them



- To resolve objects we need to match the wavelength of light we use to the object we want to observe

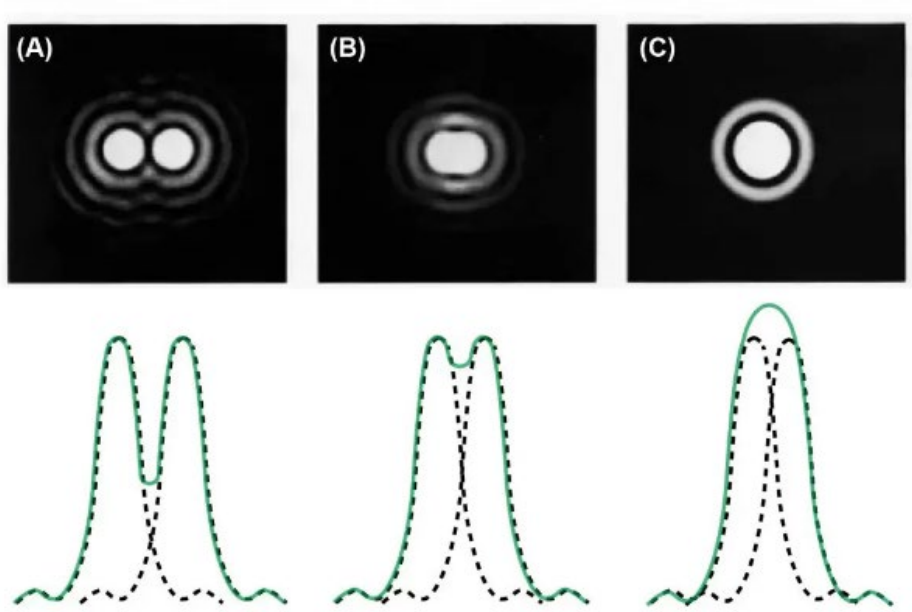


Image modified from http://hinkhousescience.weebly.com/electromagnetic-mini-unit.html

Proteins are so small we can't use visible light to see them

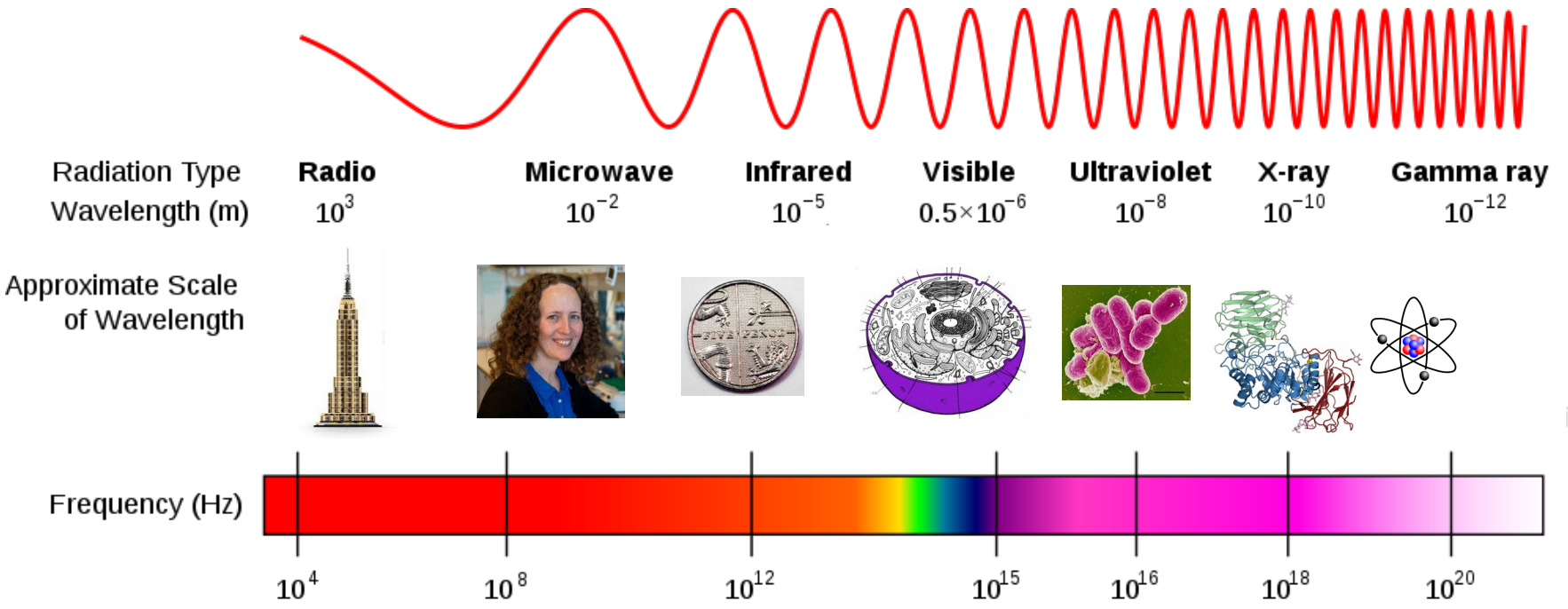


Image modified from http://hinkhousescience.weebly.com/electromagnetic-mini-unit.html

Proteins are so small we can't use visible light to see them

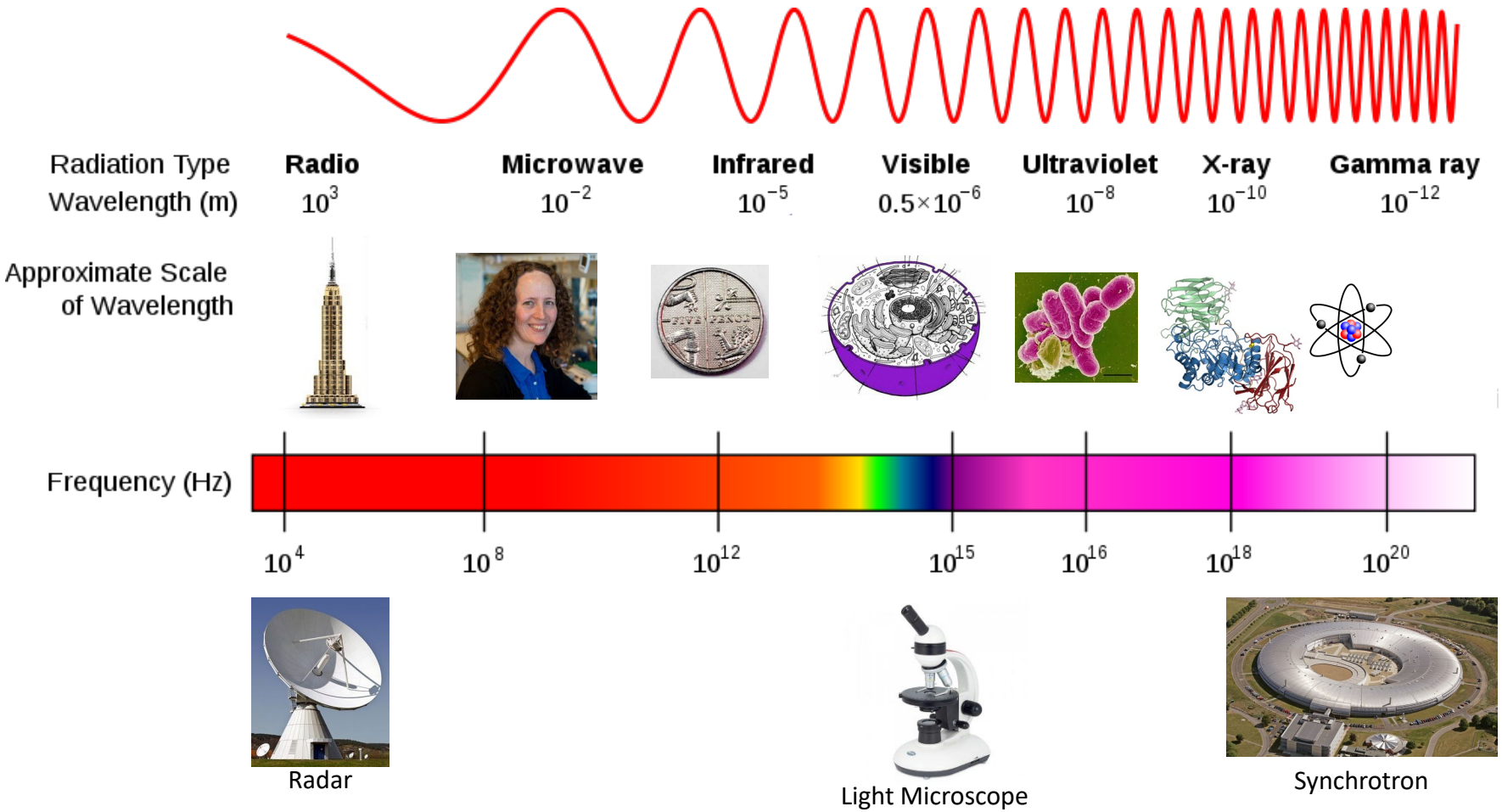
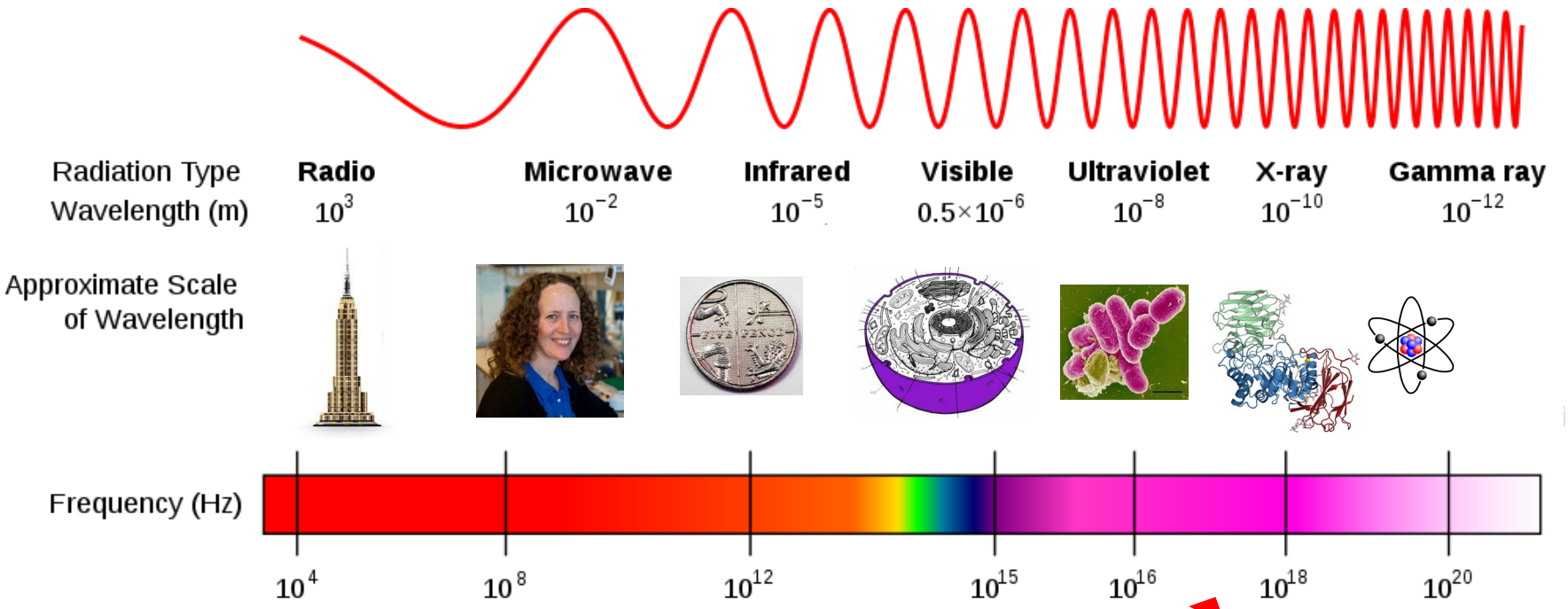


Image modified from http://hinkhousescience.weebly.com/electromagnetic-mini-unit.html

Proteins are so small we can't use visible light to see them

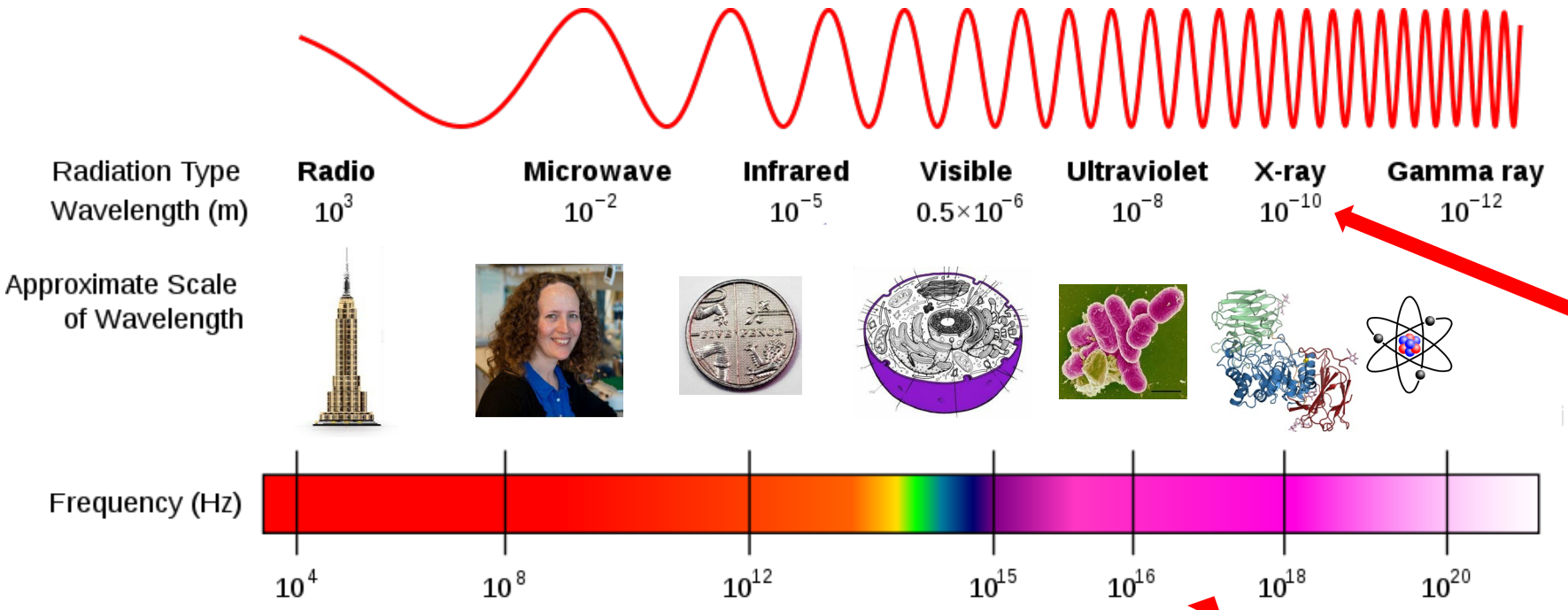


Let's start here with X-ray crystallography

Synchrotron

Image modified from http://hinkhousescience.weebly.com/electromagnetic-mini-unit.html

Proteins are so small we can't use visible light to see them



Structural work uses the term Ångström (Å) instead of 10^{-10}

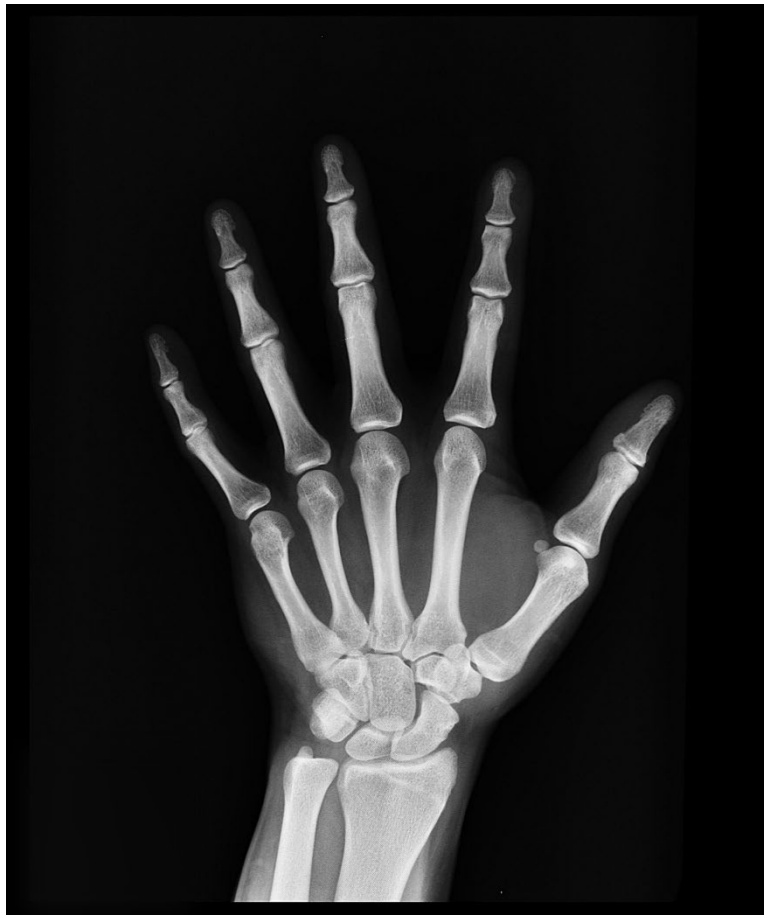
Let's start here with X-ray crystallography

Synchrotron

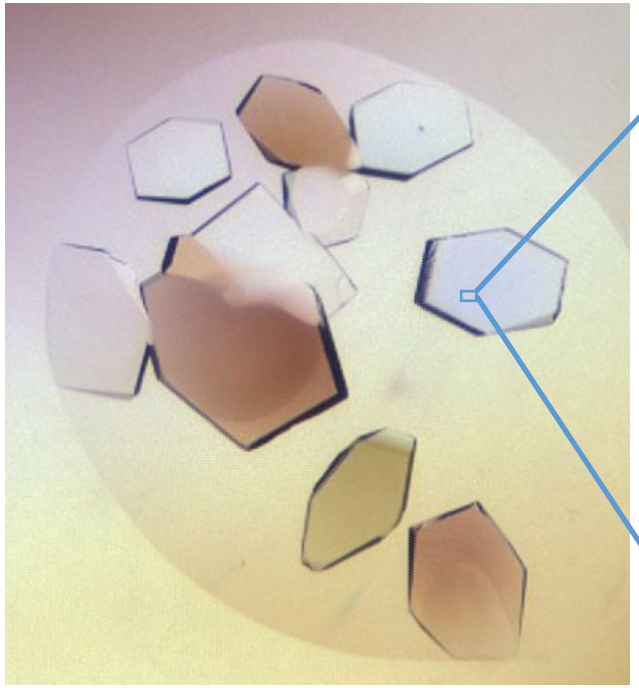
Image modified from http://hinkhousescience.weebly.com/electromagnetic-mini-unit.html

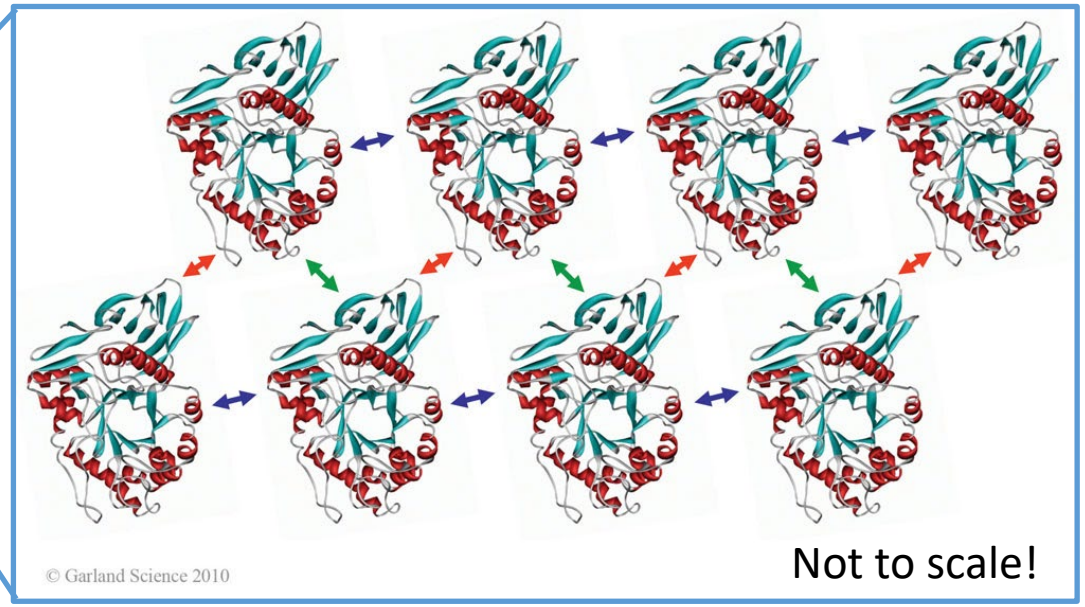
X-rays interact weakly with matter…

- They pass straight through our bodies (lots of protein!)
- We use ‘soft’ X-rays that interact more strongly with matter



...and we crystallise proteins to enhance the signal

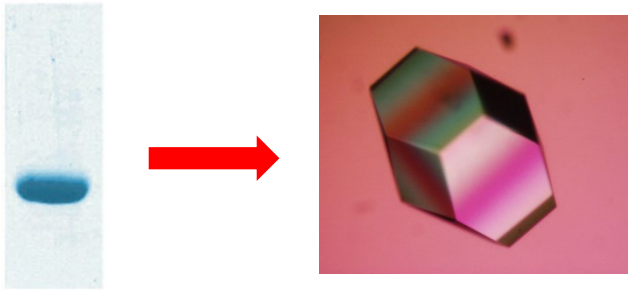




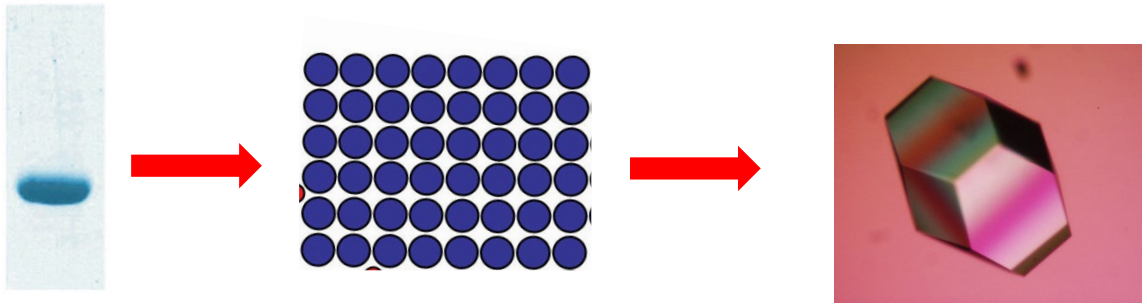
- Crystals are ordered arrays of molecules
- They diffract X-rays in phase, amplifying the signal

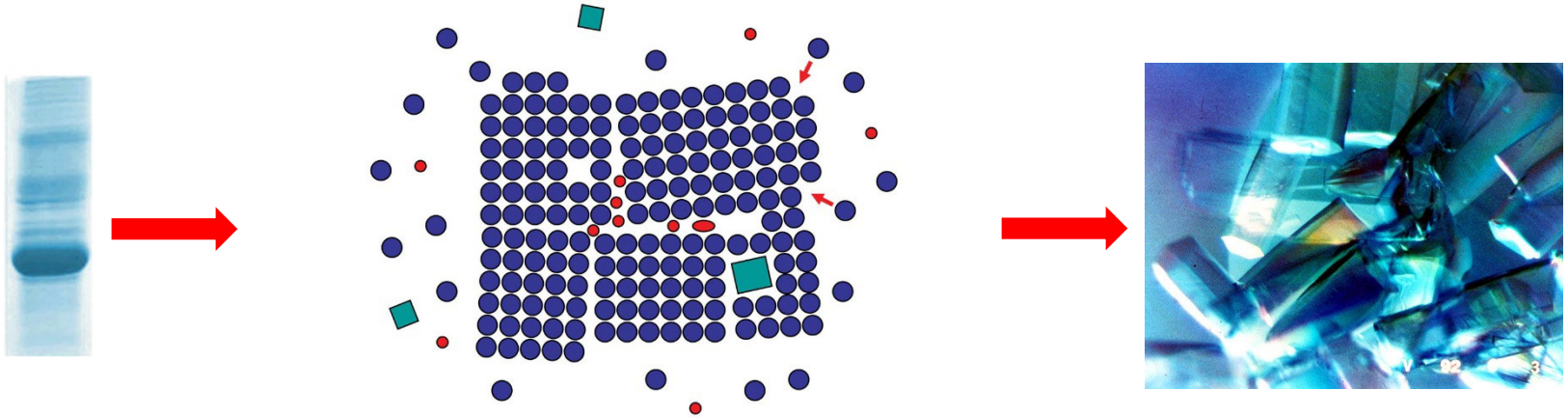
Crystallising proteins

- Requires large amounts (milligrams) of pure protein
- Exactly as we have been doing in this course



Protein purity: >95% pure

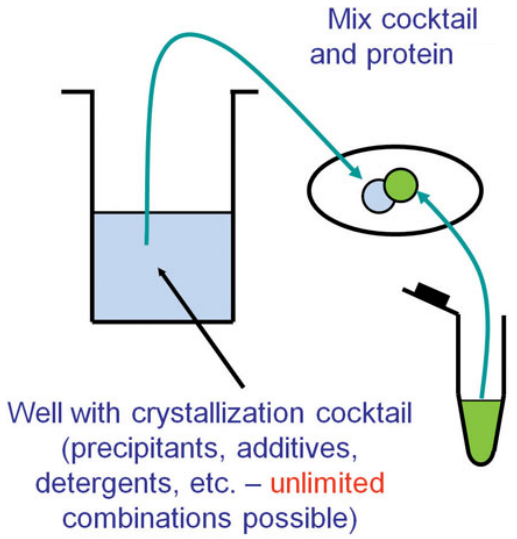


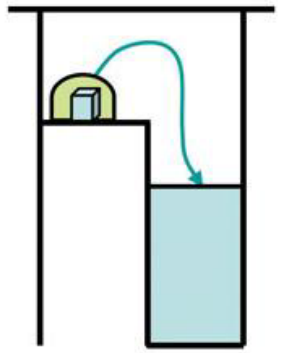


Images modified from: Bernard Rupp 2010

Crystallising proteins

- Equilibrate with chemical cocktails to promote crystallisation



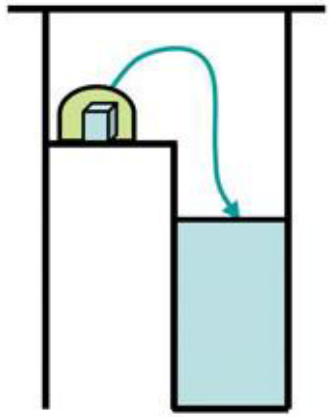


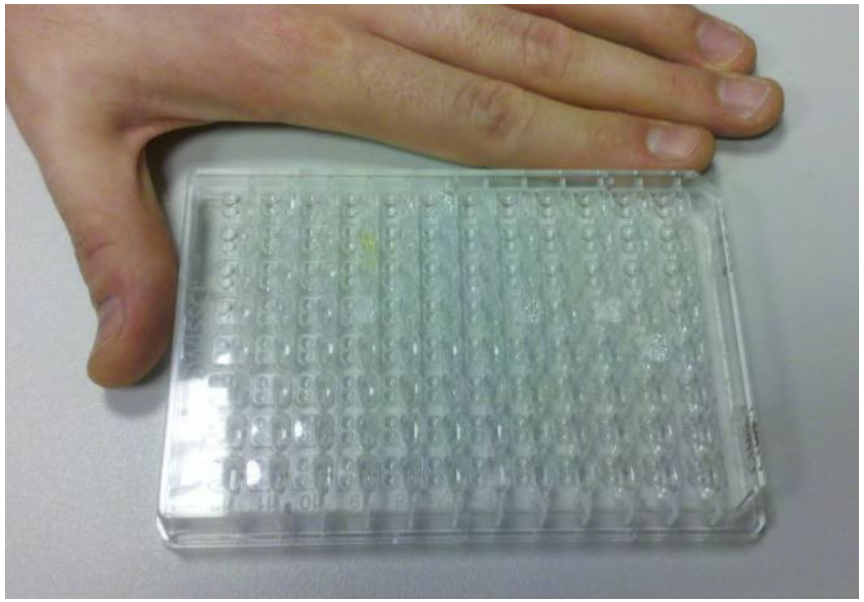
Images: Bernard Rupp 2010

Crystallising proteins

- Equilibrate with chemical cocktails to promote crystallisation

- Can’t predict conditions required to crystallise a given protein
  - Try 1000s of conditions



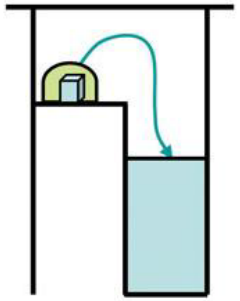


Crystallising proteins

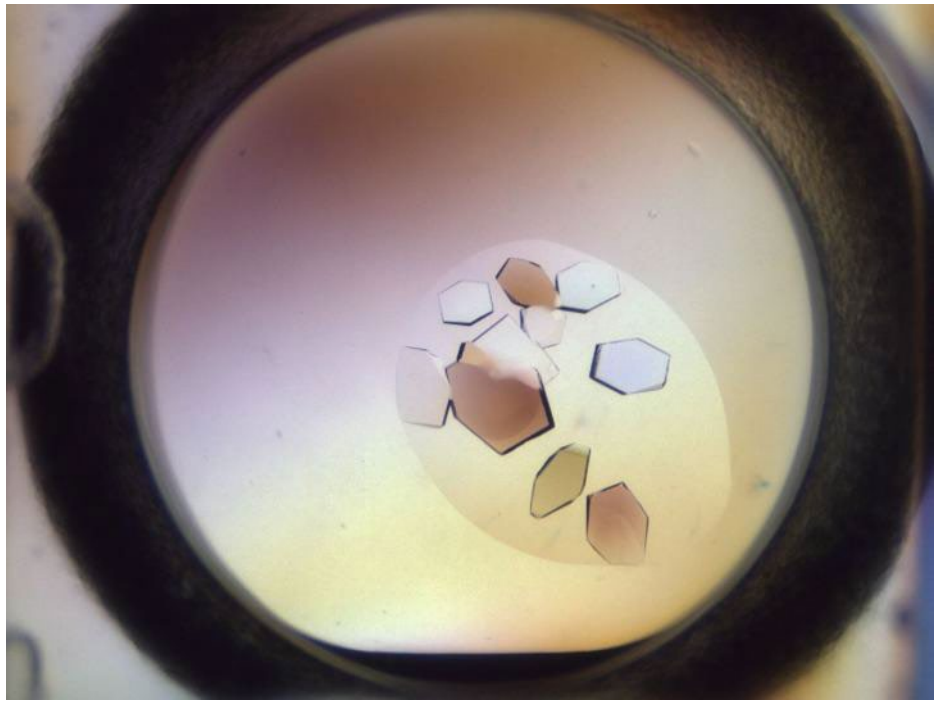
- Equilibrate with chemical cocktails to promote crystallisation

- Can’t predict conditions required to crystallise a given protein
  - Try 1000s of conditions

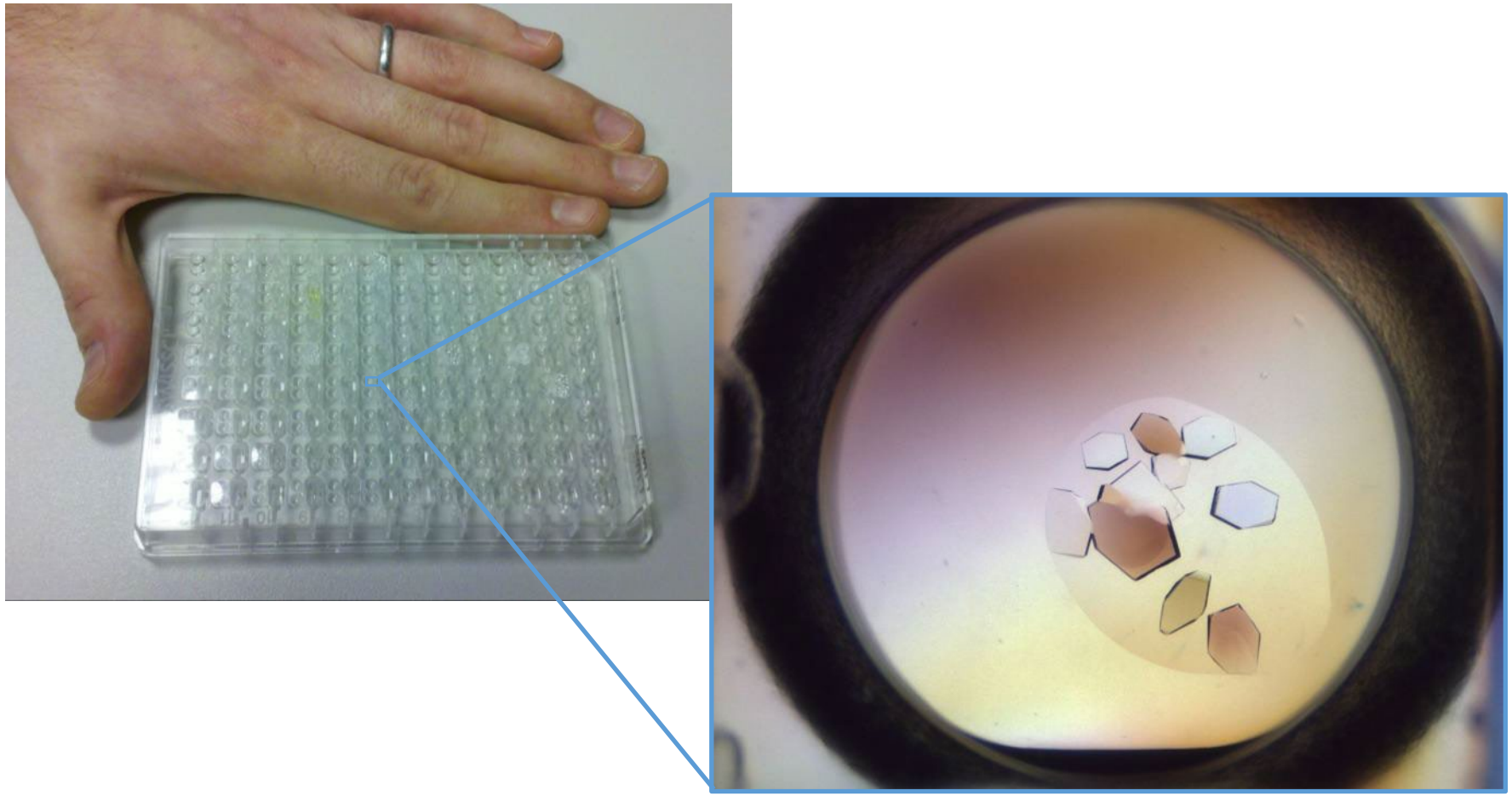
Nanolitre-scale crystallisation



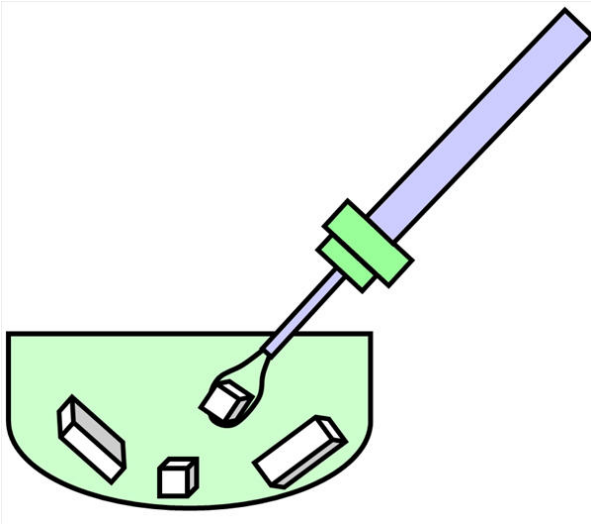
and automated visualisation

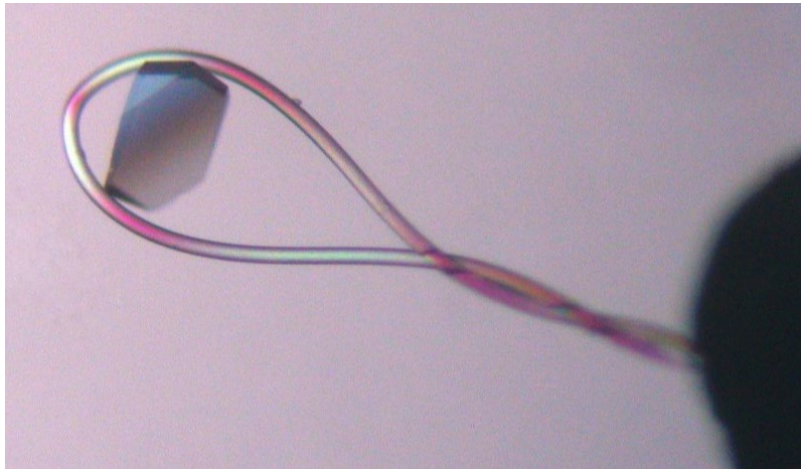


Manually harvest crystals



Harvesting and Mounting

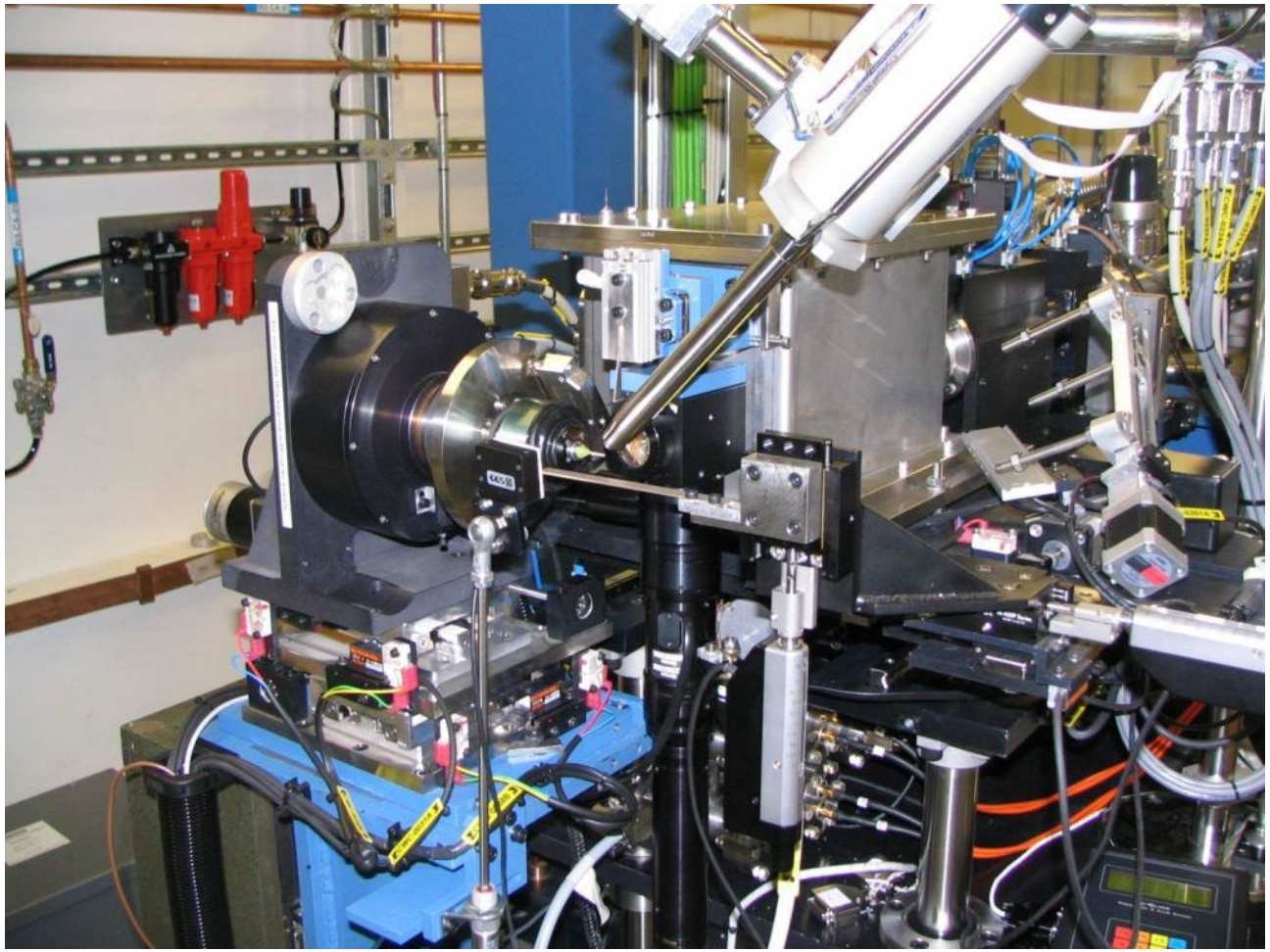




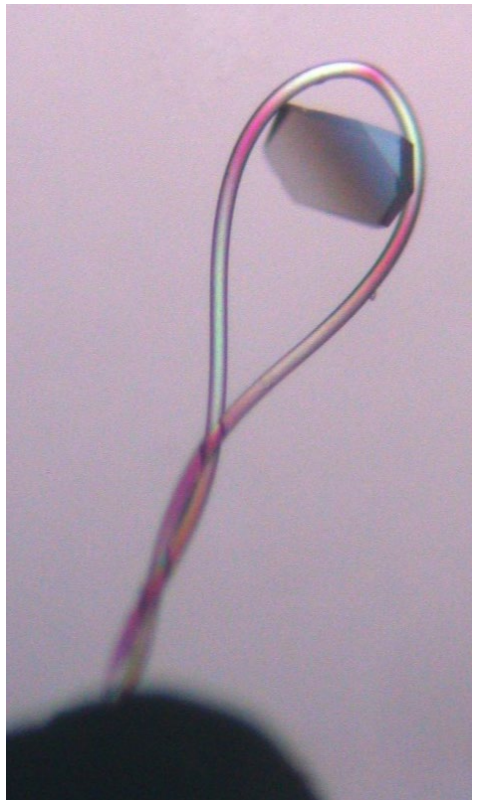
- Flash-cooled in liquid nitrogen to minimise radiation damage by intense X-rays
- Stored in a transport dewar at liquid nitrogen temperatures

Travel to a synchrotron

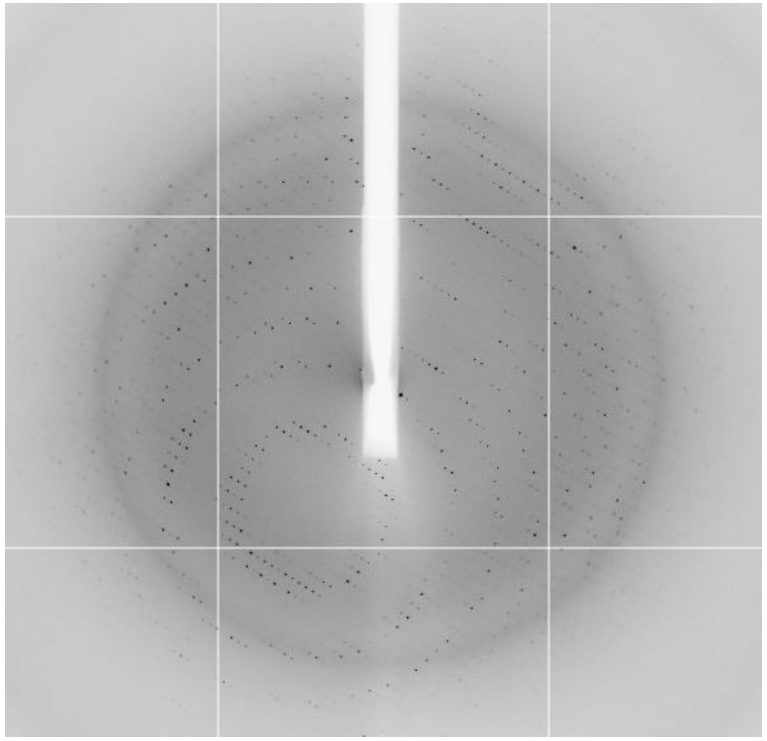
...which actually looks like this



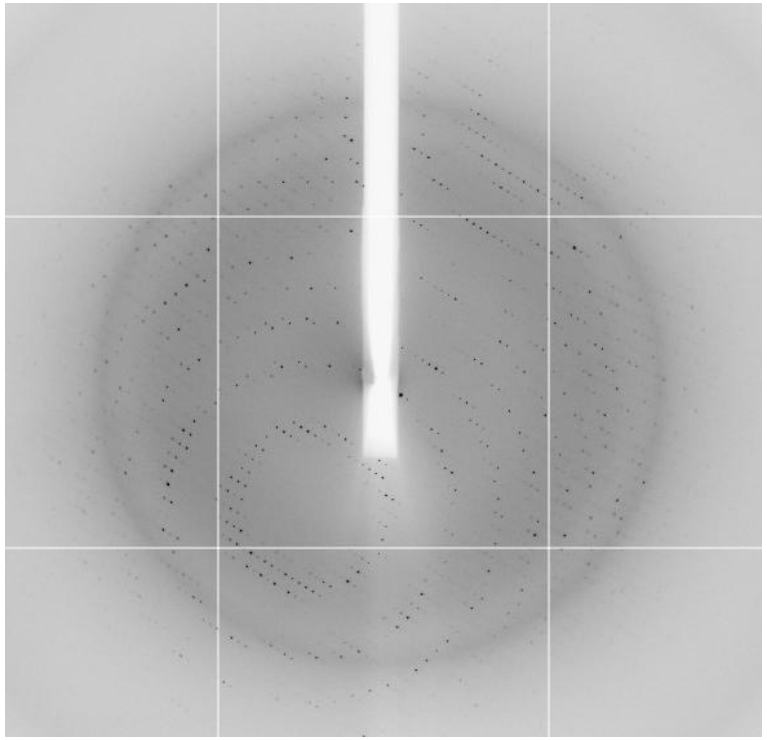
Collect diffraction data

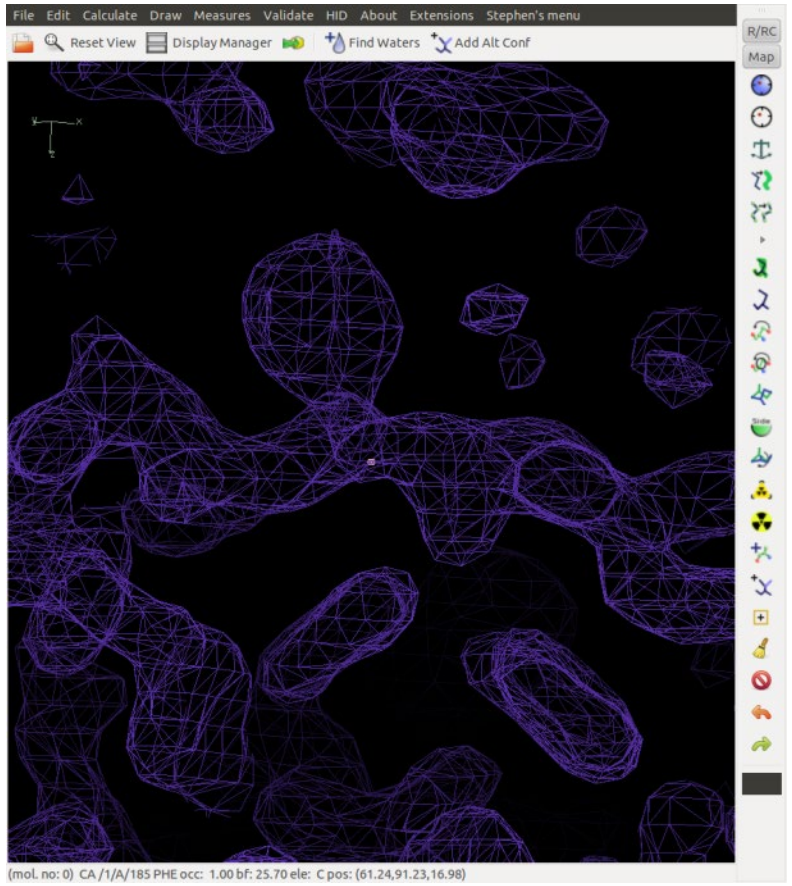


X-rays



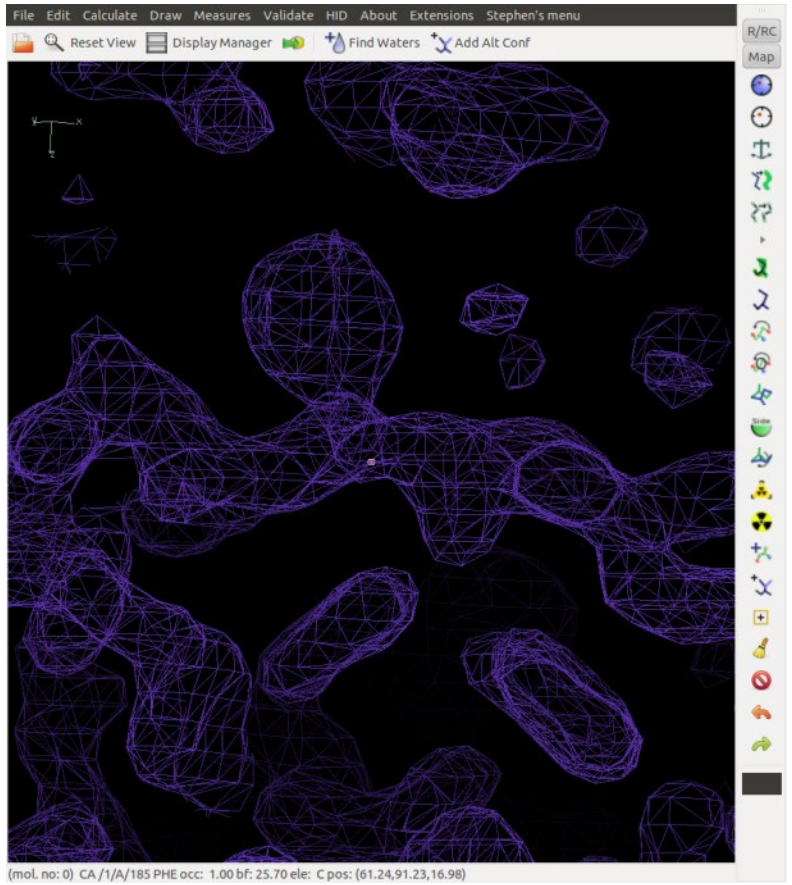
Solve structure



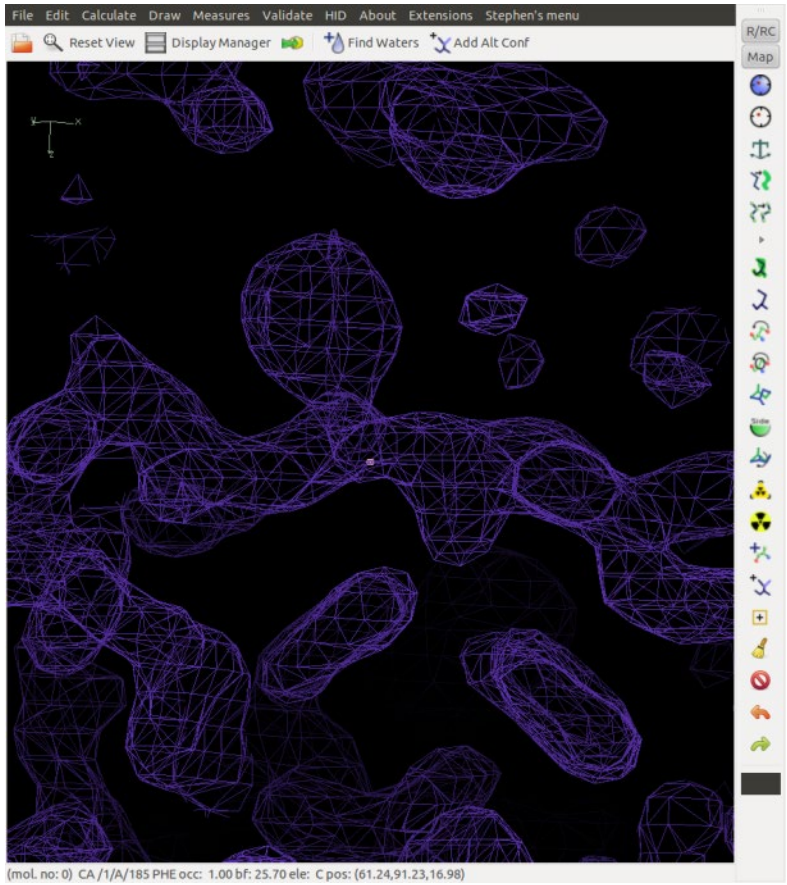


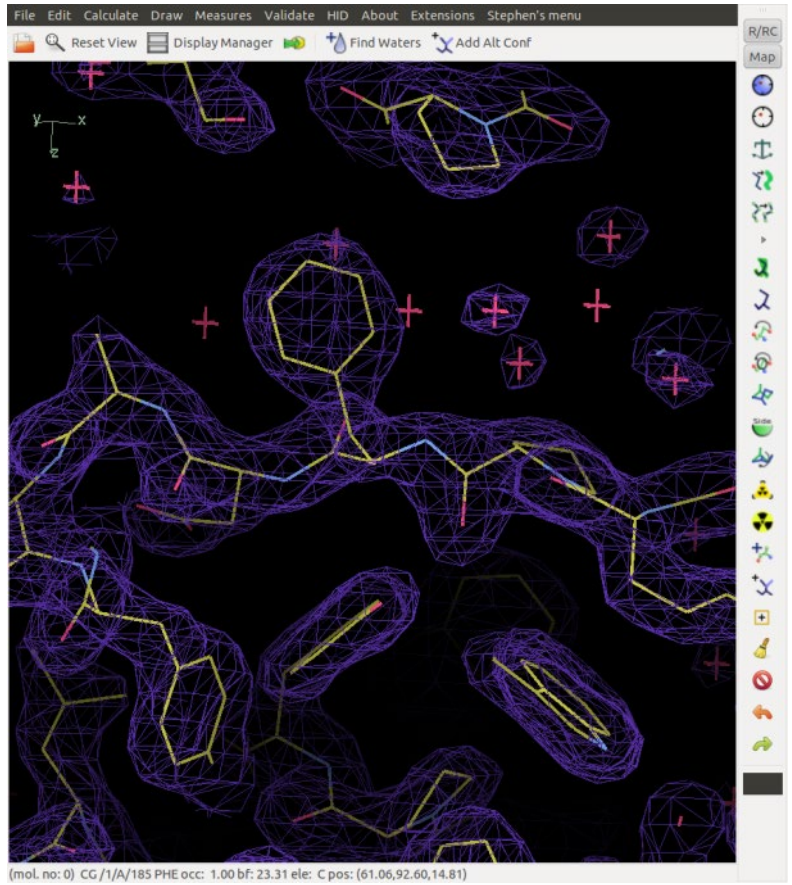
Solve structure – in a nutshell

- Several approaches
- Most common is called molecular replacement
- This uses a predicted model of the protein to “determine phase information”
- This is a whole course in itself!

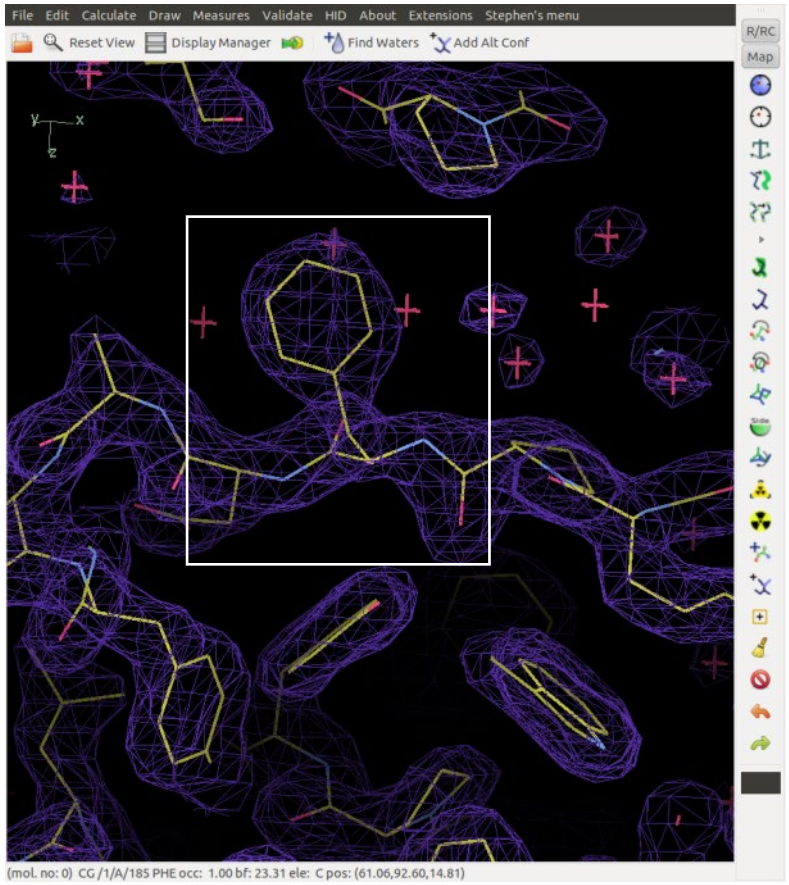


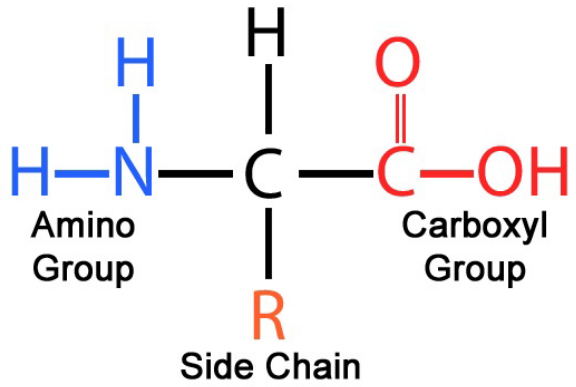
Build structure



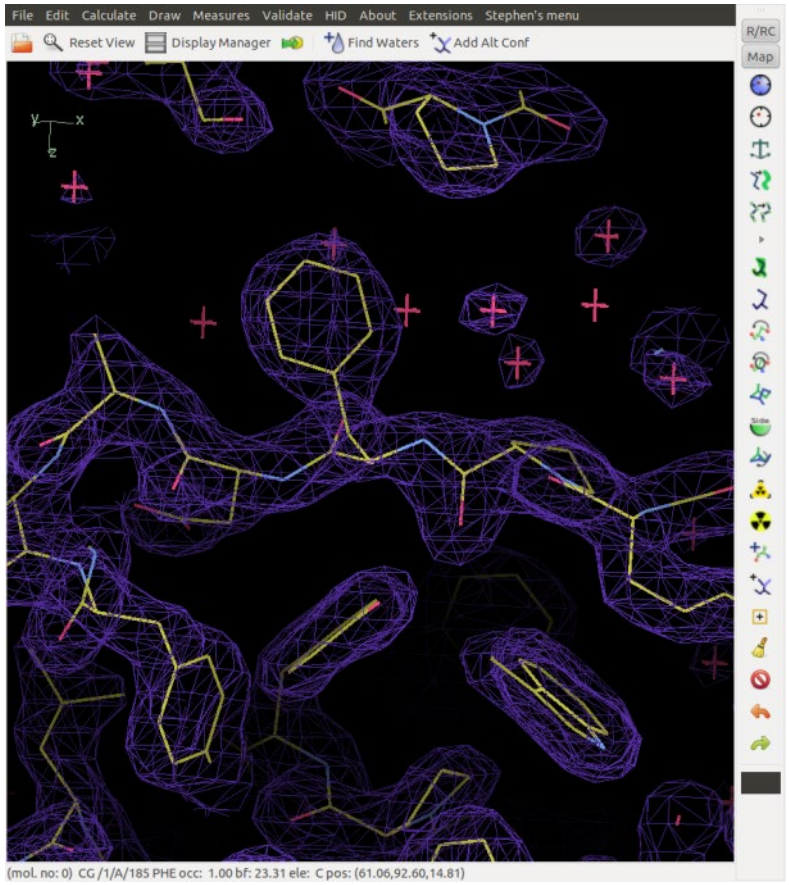


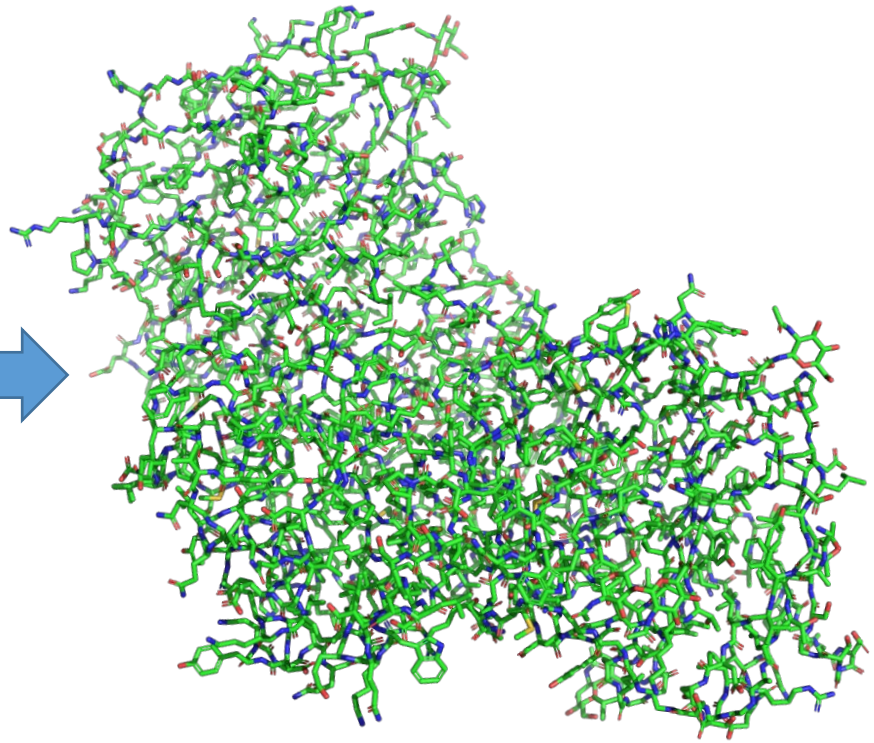
Build structure



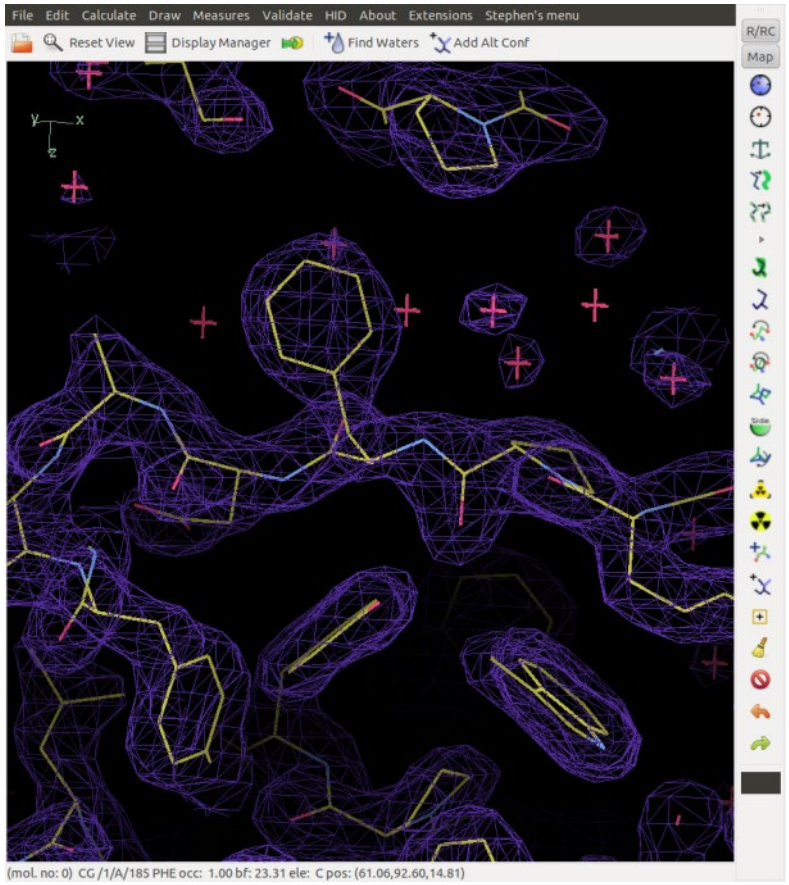


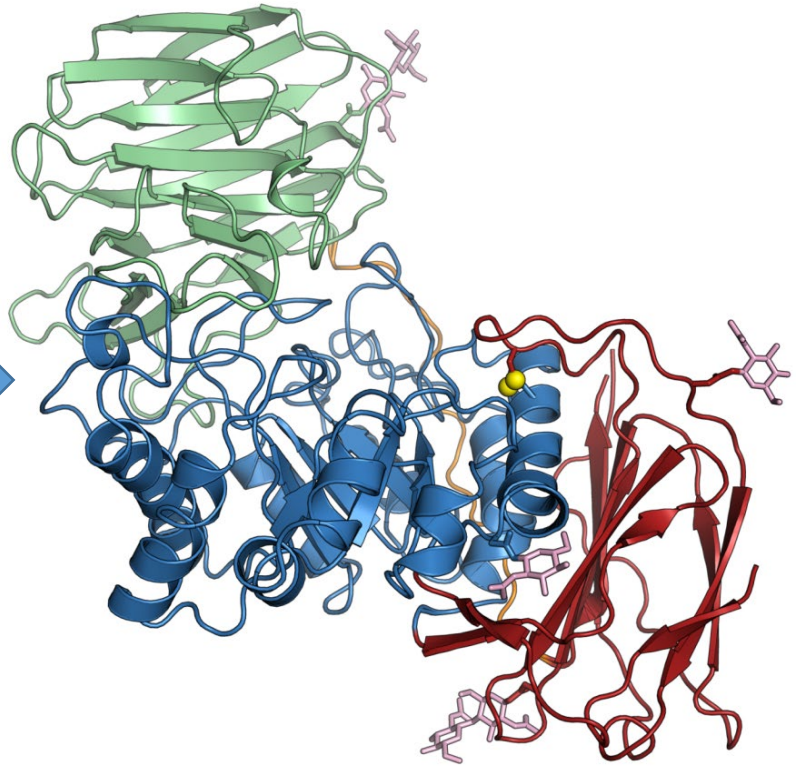
Build structure





Interpret structure and test functional hypotheses

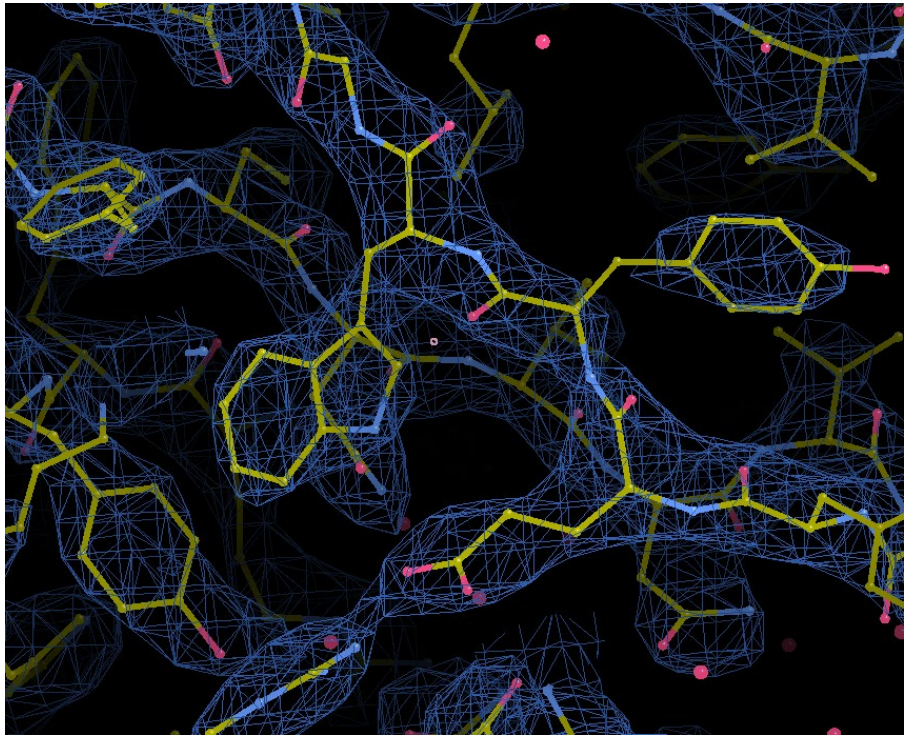




This afternoon's workshop

- We'll use Coot this afternoon to look at maps from a crystal structure

- Blue map (2Fo-Fc) is the electron density
  - This is how we see where the atoms are
  - This includes the protein and solvent

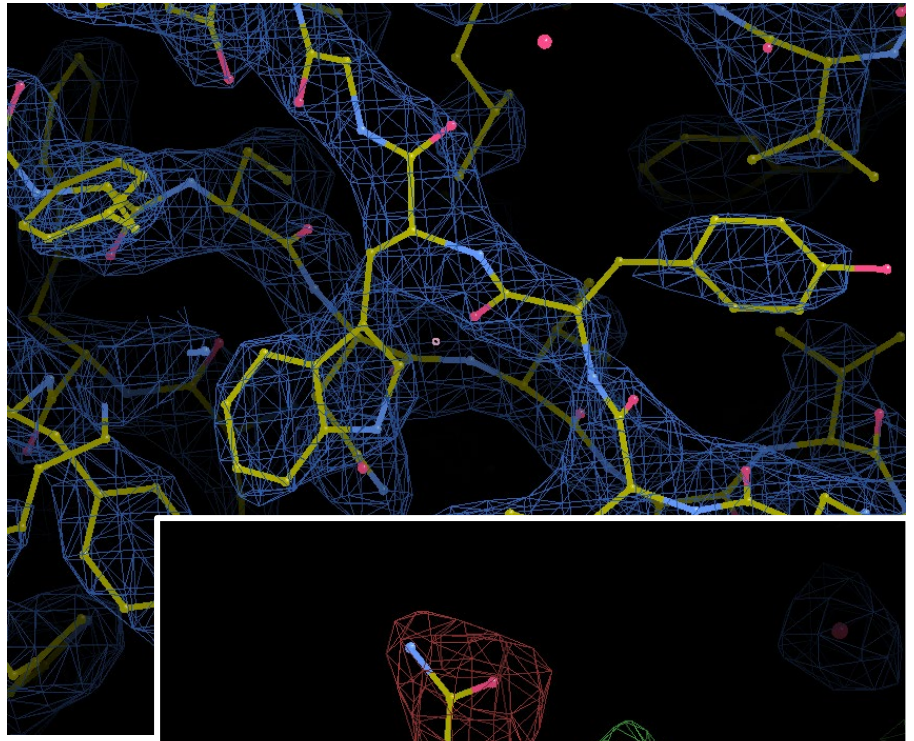


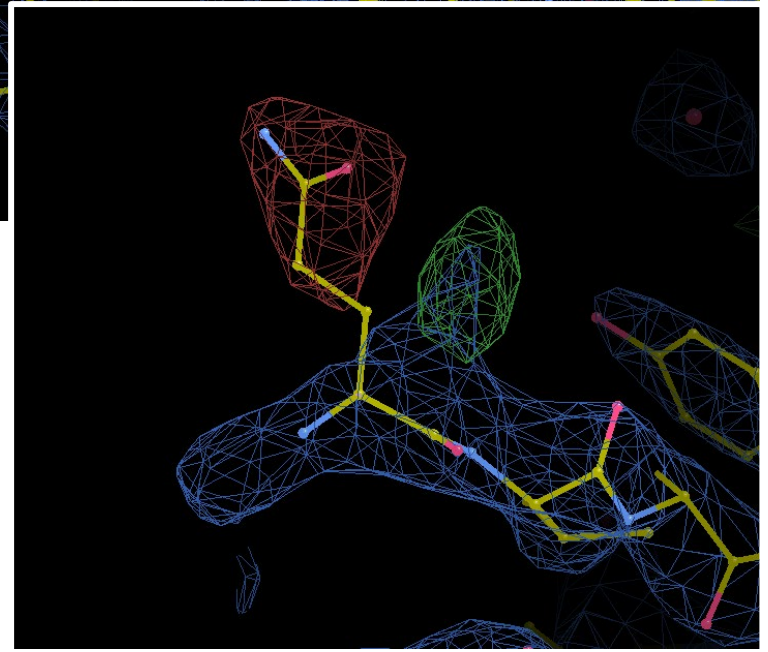
This afternoon's workshop

- We'll use Coot this afternoon to look at maps from a crystal structure

- Blue map (2Fo-Fc) is the electron density
  - This is how we see where the atoms are
  - This includes the protein and solvent

- Red/Green map (Fo-Fc) is the difference map
  - This is where the model and the density disagree
  - Green – the model is missing something
  - Red – something is modelled incorrectly



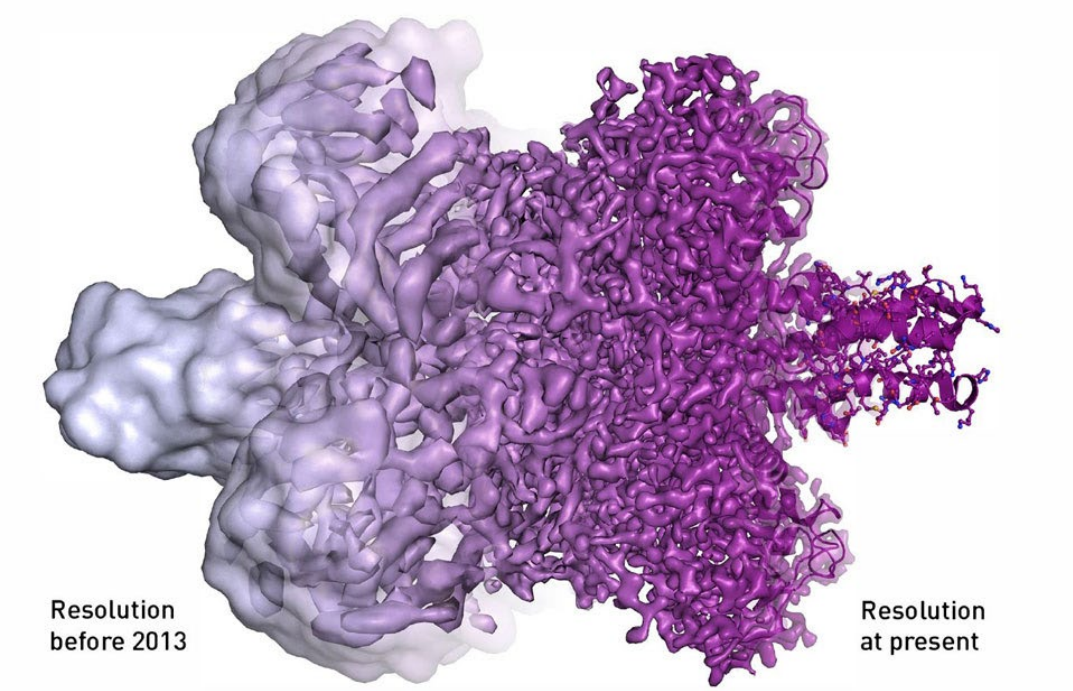


That’s crystallography in a nutshell.

How about cryoEM?

How about Cryo-EM?

- With a purified protein sample >150 kDa you can also try Cryo-EM

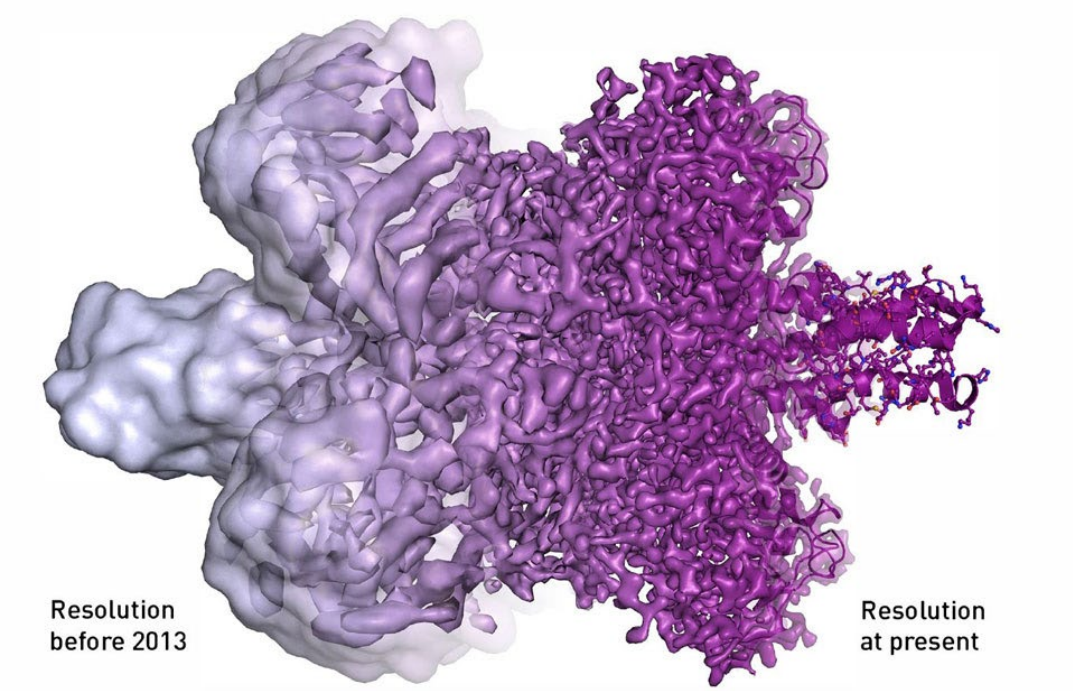


“Resolution Revolution”

From shapeless blobs to atomic resolution

How about Cryo-EM?

- With a purified protein sample >150 kDa you can also try Cryo-EM

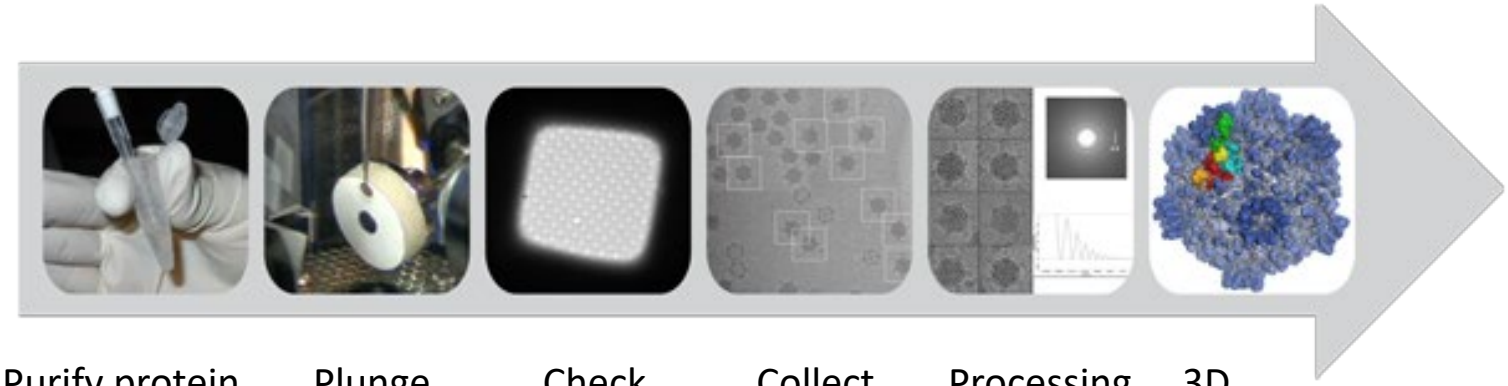


Revolution was driven by advances in technology:

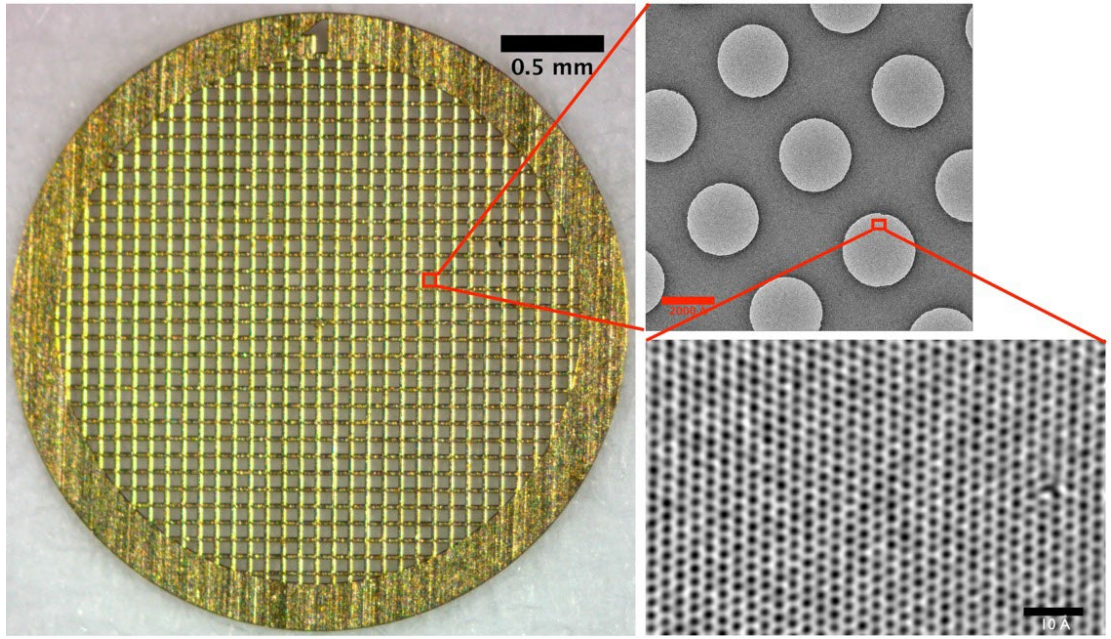
- Direct electron detectors
- Improved microscope stability

How about Cryo-EM?

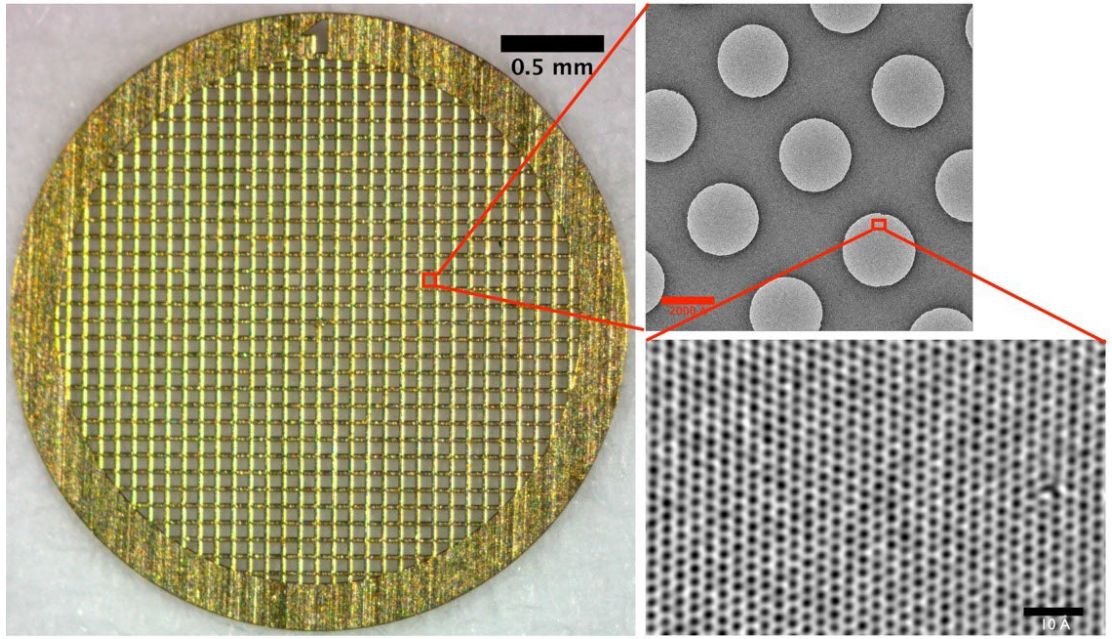
- With a purified protein sample >150 kDa you can also try Cryo-EM
- Different pipeline compared with crystallography:

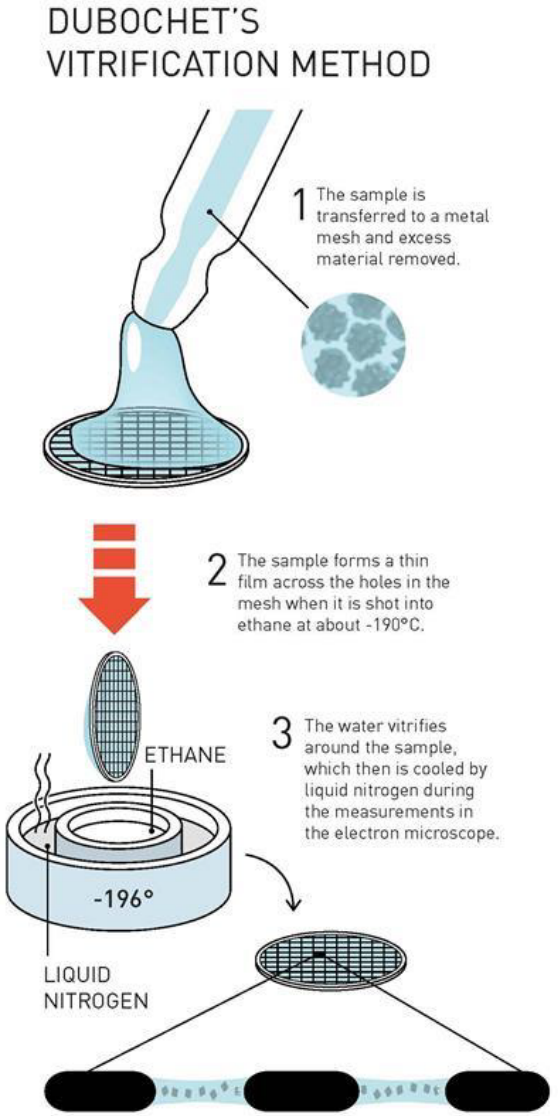


Making grids

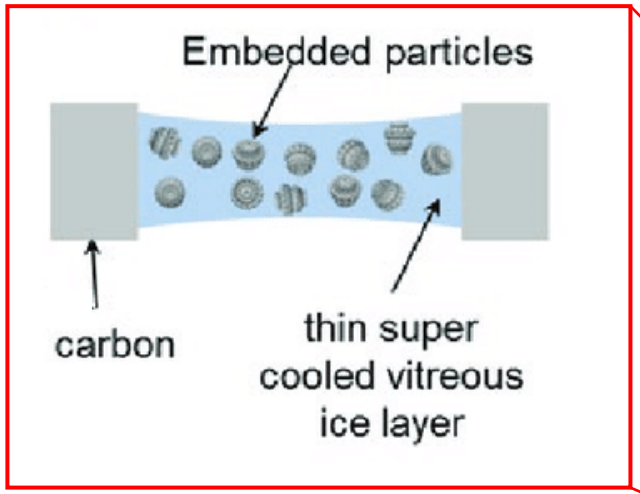


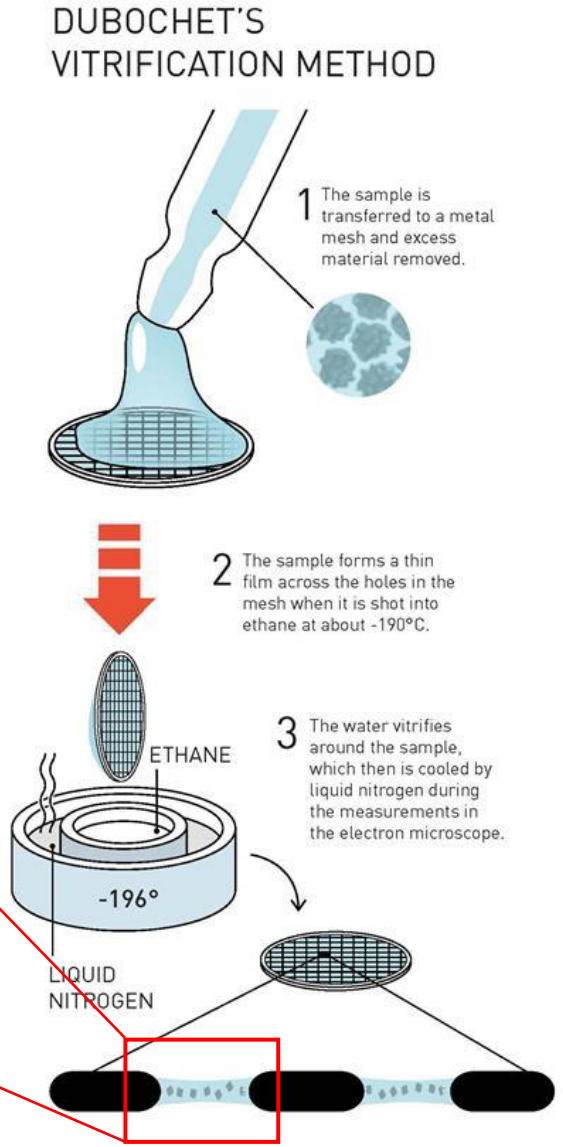
Making grids



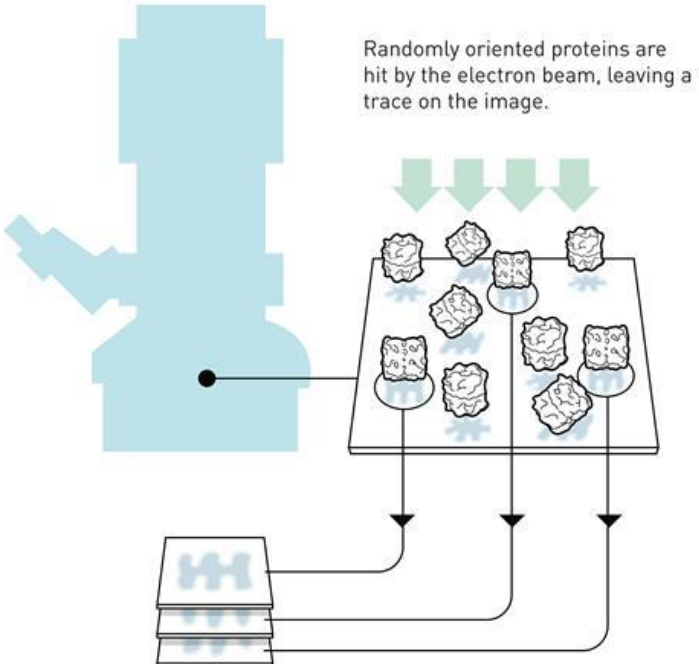


Making grids

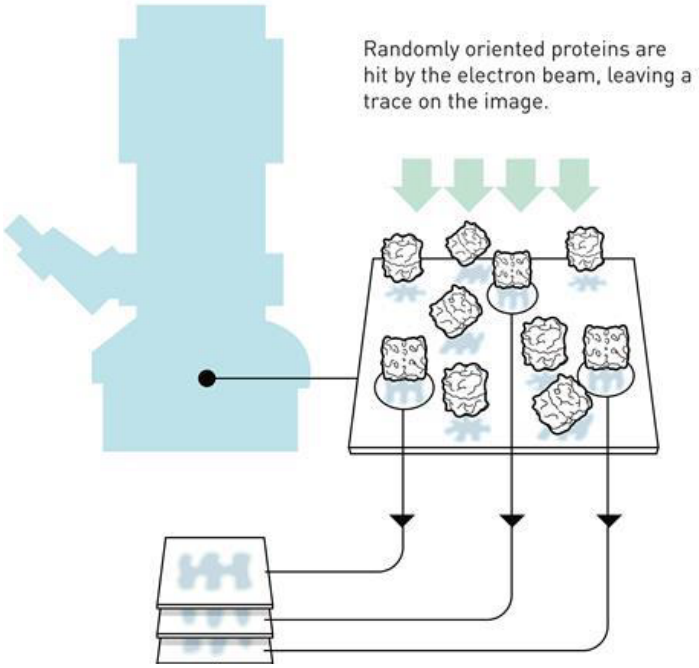




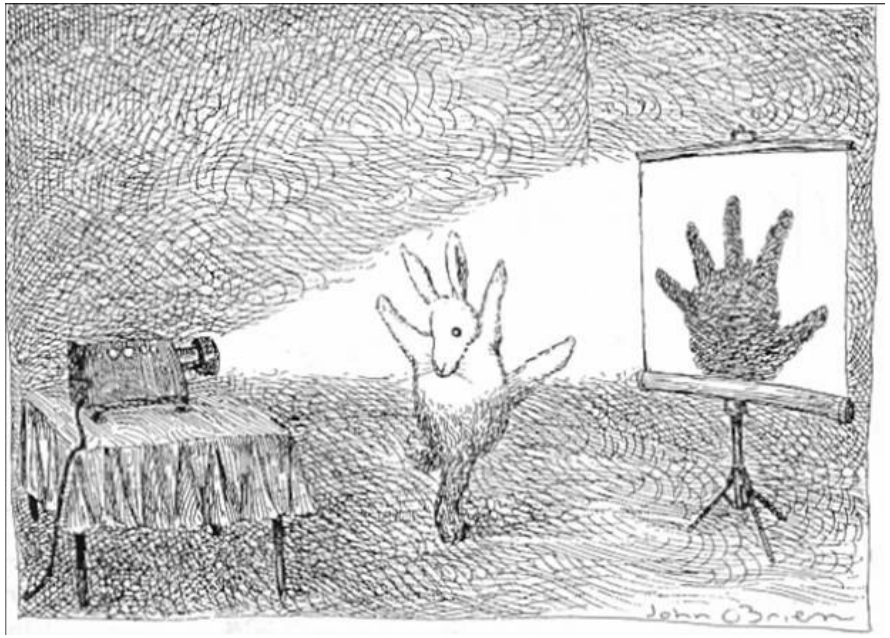
Collecting Images



Collecting Images

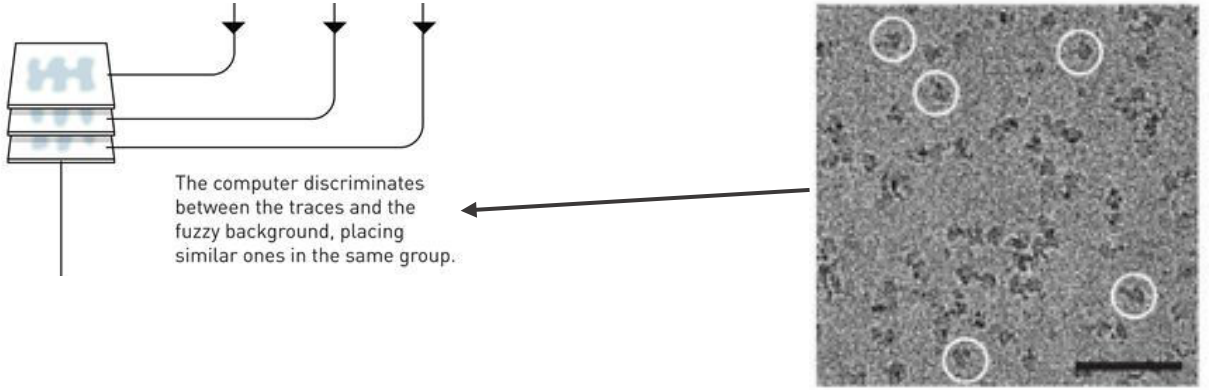


Collecting Images

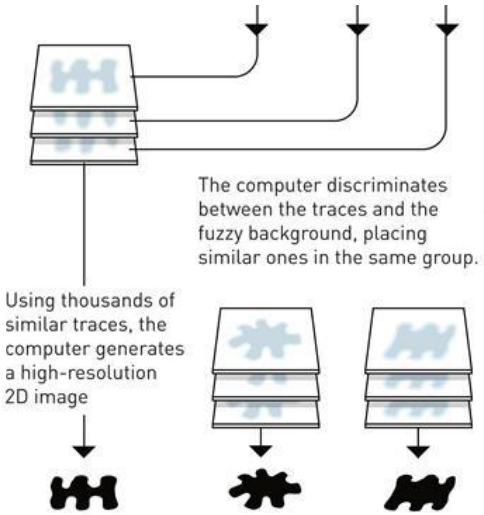


A single 2D projection image is insufficient to determine structure of a 3D object

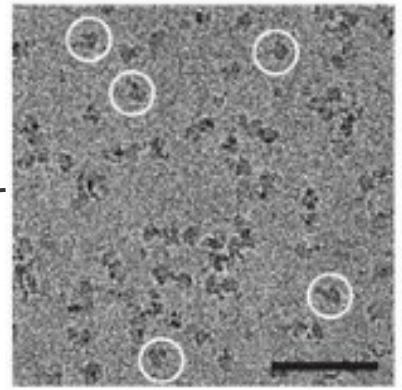
Processing and Data Analysis



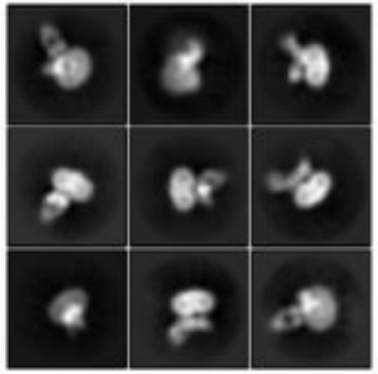
Processing and Data Analysis



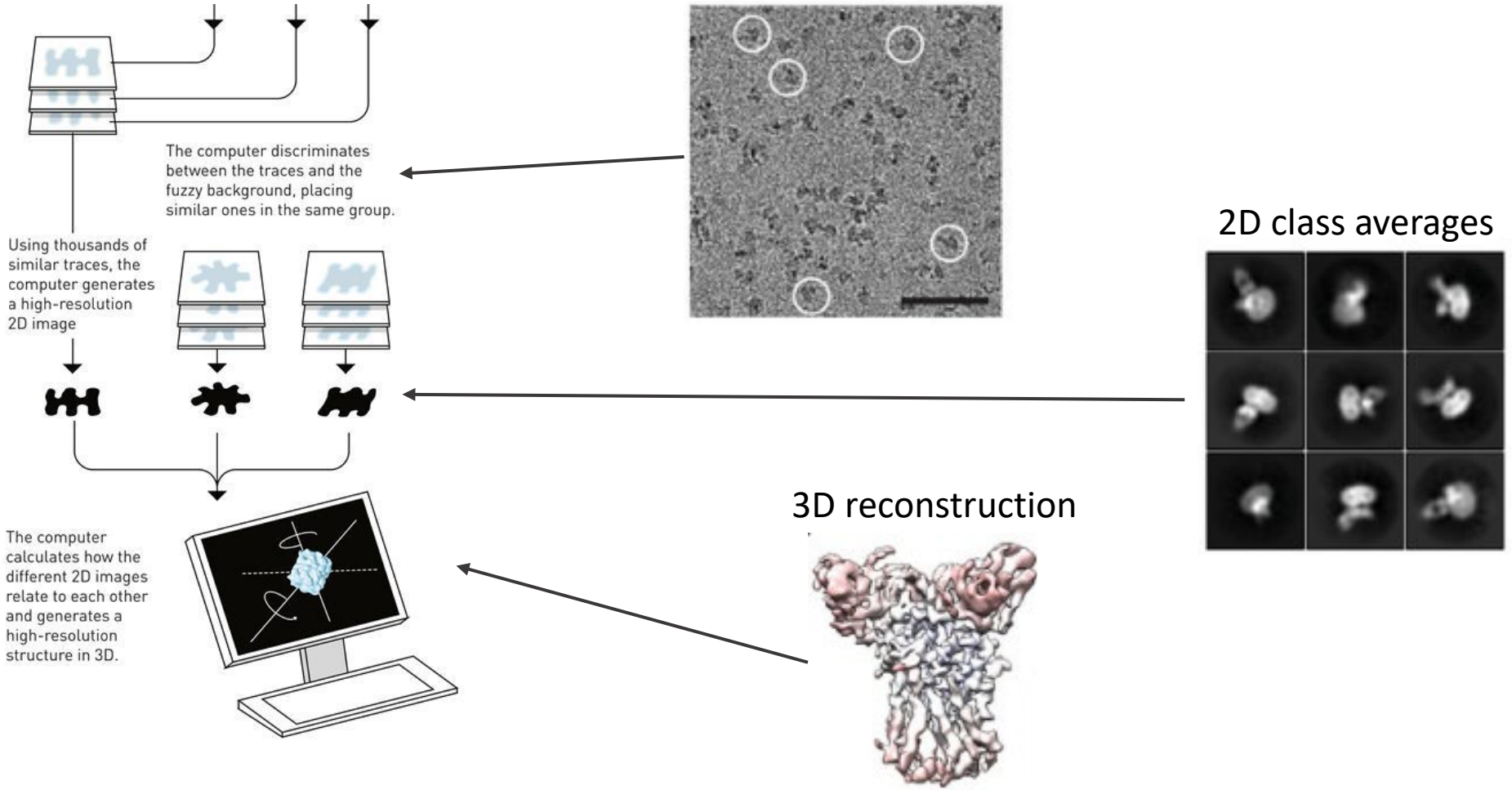
Particle picking



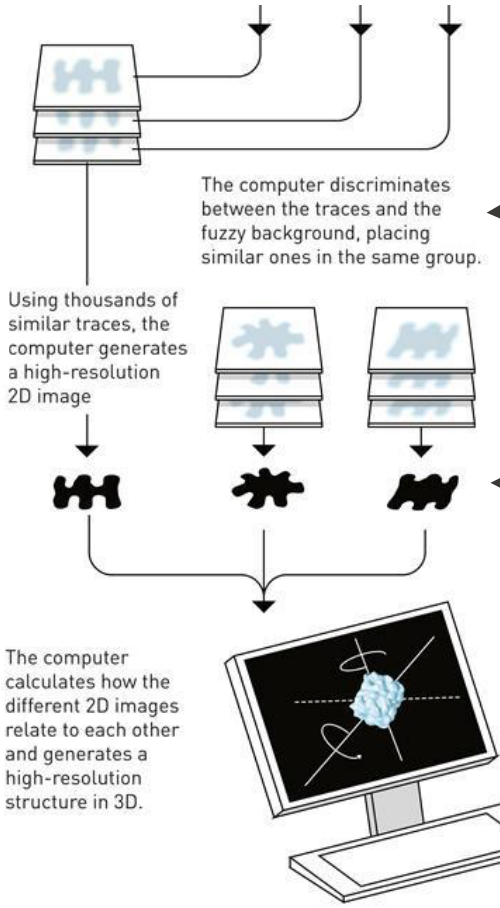
2D class averages



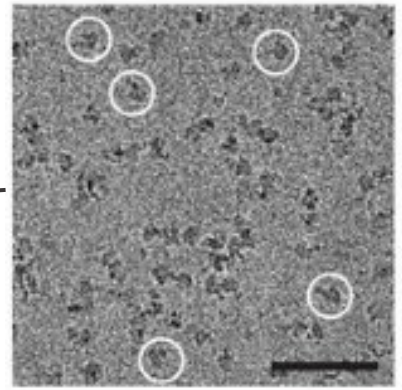
Processing and Data Analysis



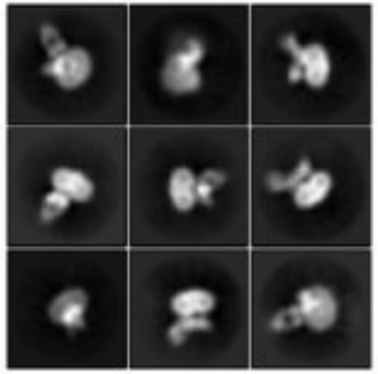
Processing and Data Analysis



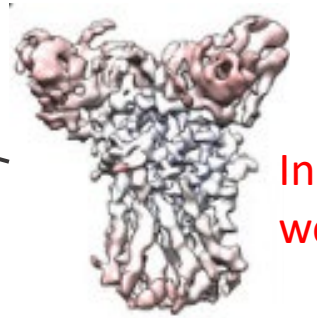
Particle picking



2D class averages



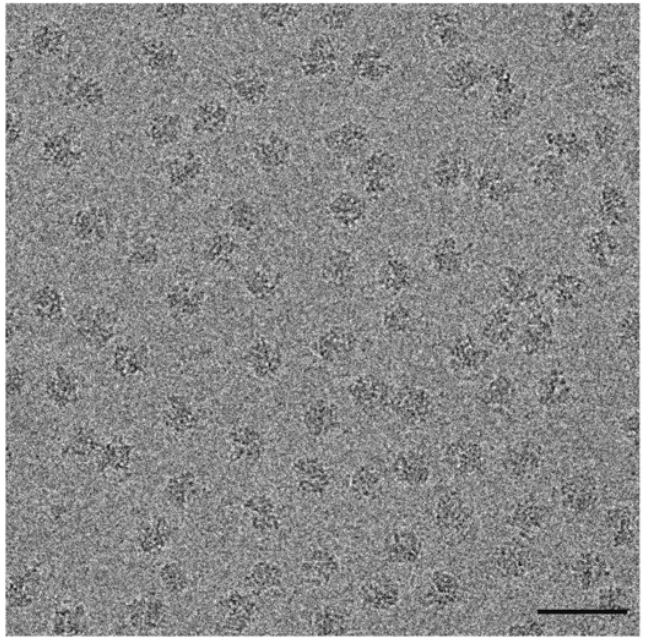
3D reconstruction



In a similar way to crystallography, we build a model into this map

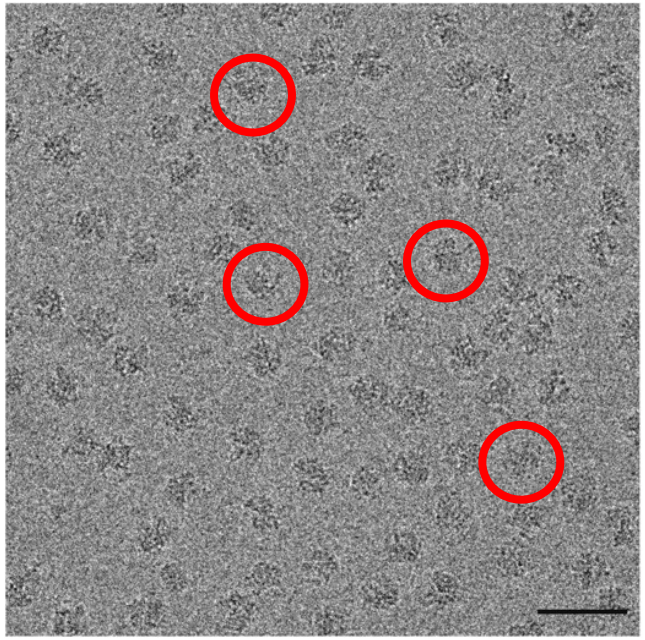
Impure samples – what if you have contaminants?

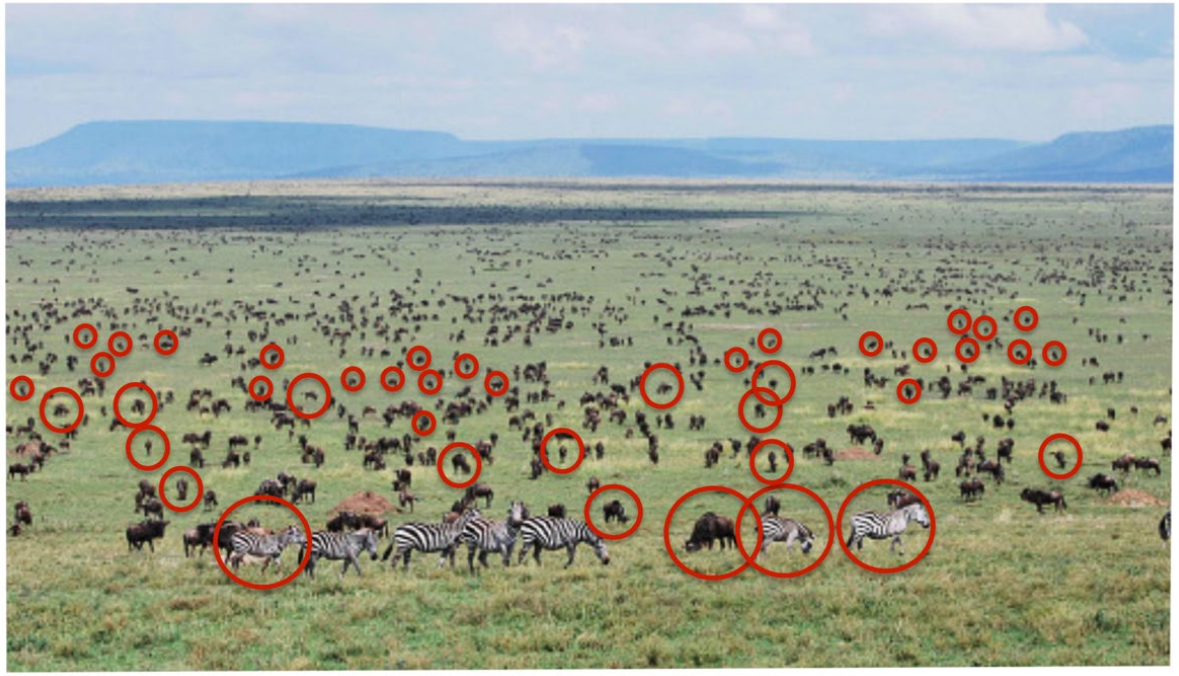
- Consequences of averaging over single particles



Impure samples – what if you have contaminants?

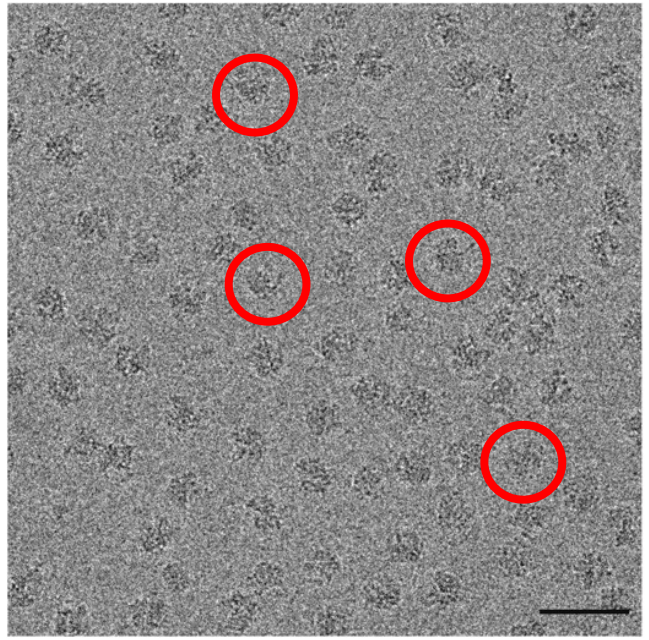
- Consequences of averaging over single particles

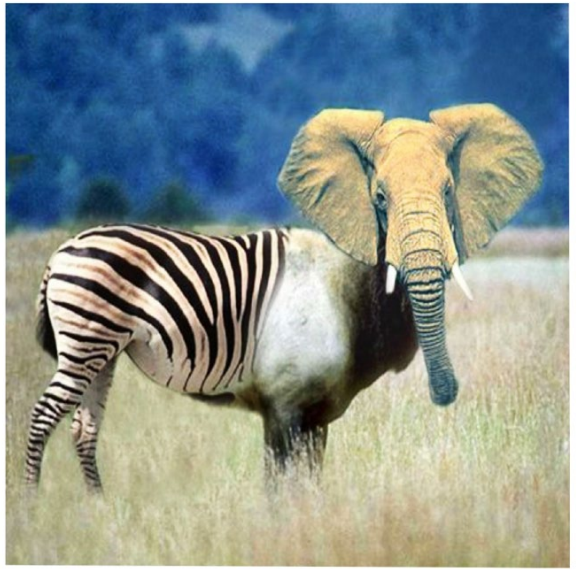




Impure samples – what if you have contaminants?

- Mixed particles

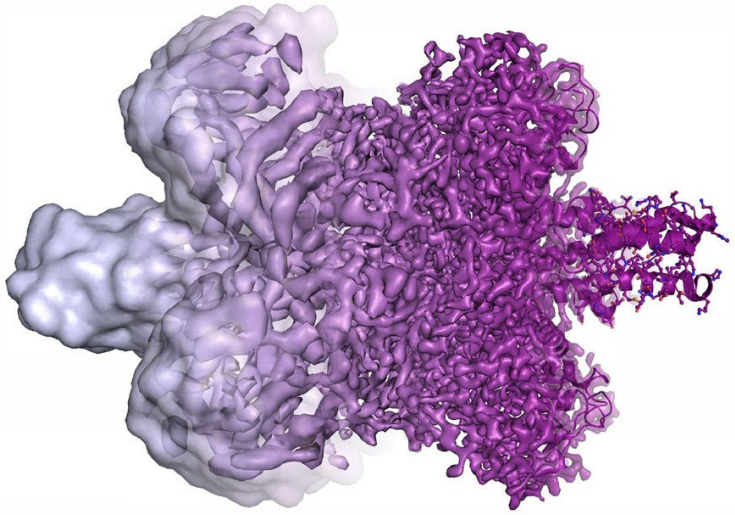


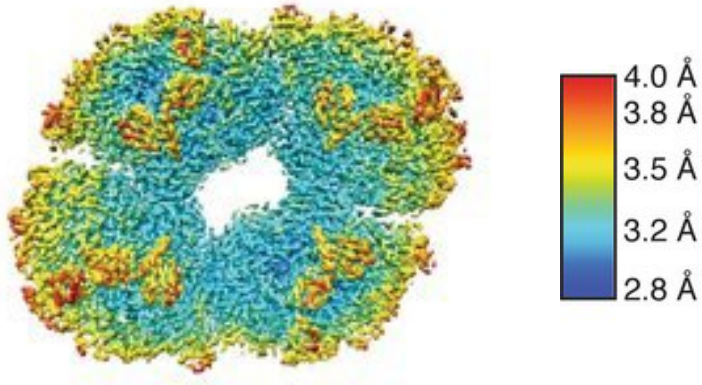


How do we judge the quality of an experimental structure?

Resolution and structure quality

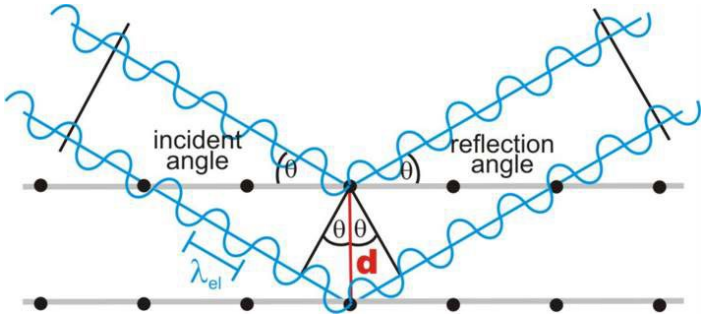
- The overall resolution of a structure is a simple measure of how “good” it is
- Even though there is one “resolution” quoted, some regions are worse than others
- “Resolution” is calculated differently in crystallography and cryoEM

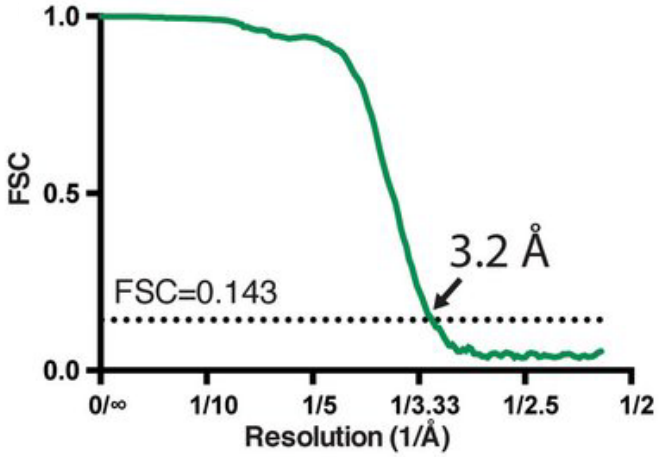




Resolution and structure quality

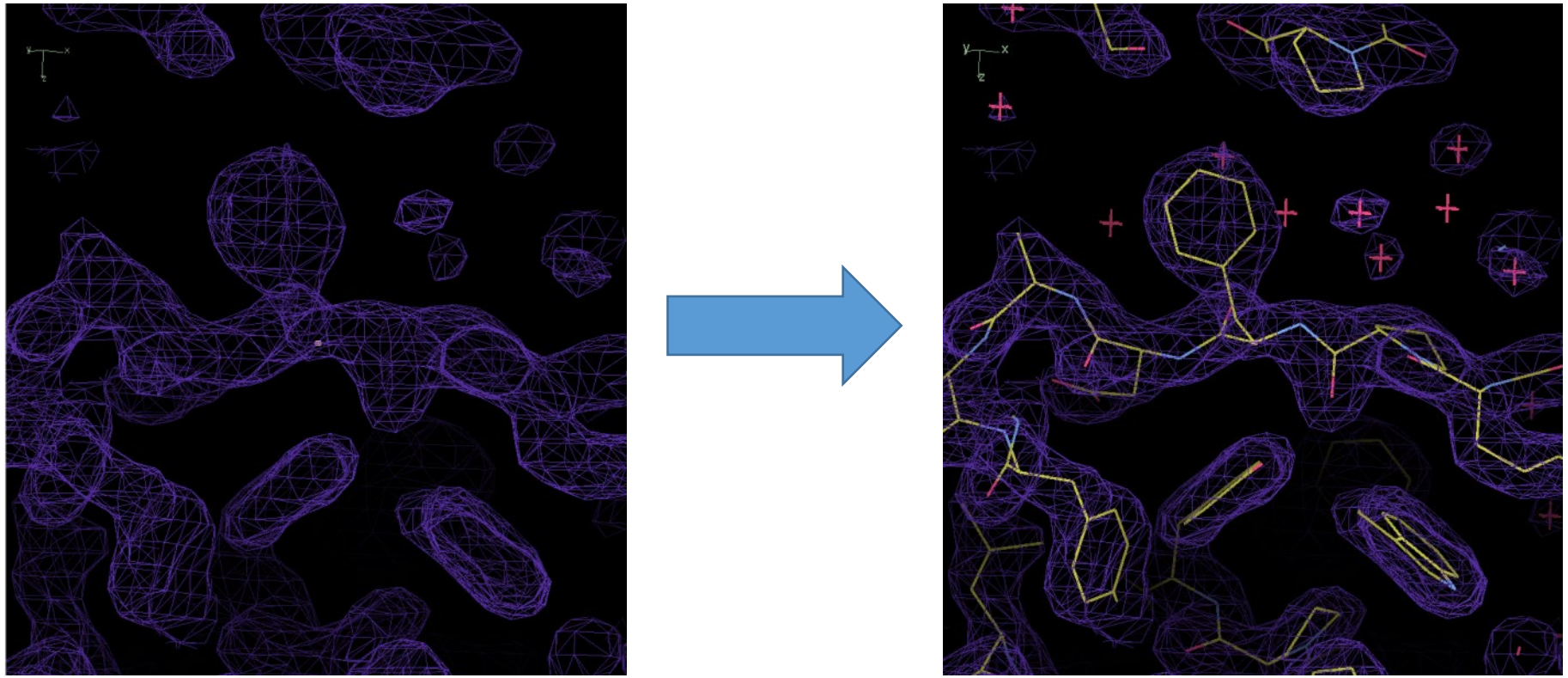
- In crystallography, resolution is determined by Bragg spacing in the crystal
- In cryoEM, resolution is determined by the Fourier Shell Correlation (FSC) for two cryo-EM half-maps
- These two numbers are not equivalent, making what “resolution” means a bit confusing



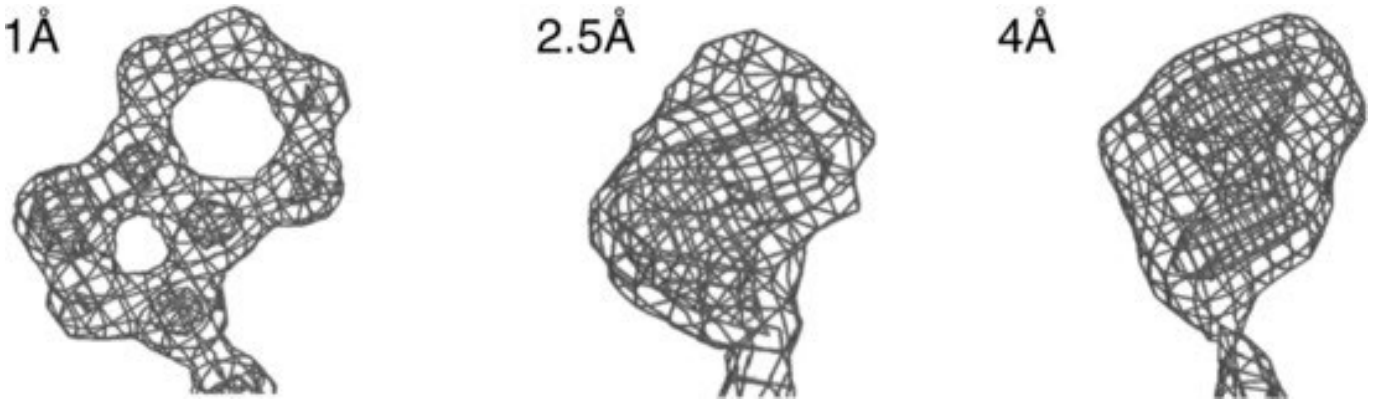


Resolution and Maps

- Resolution determines the quality of the maps to build into

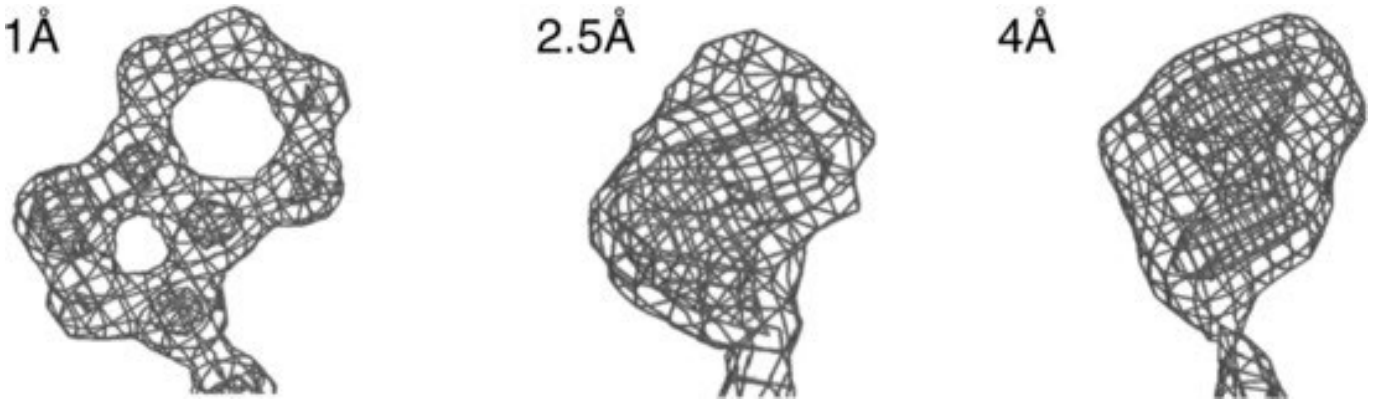


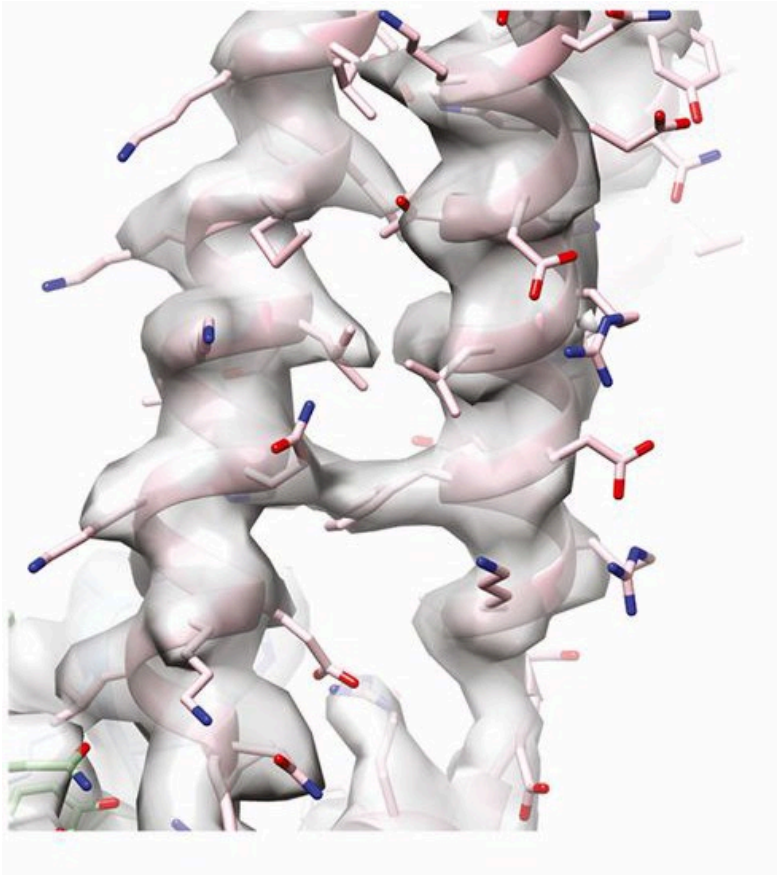
Resolution and Maps



- Clearly much harder to accurately model the tryptophan sidechain here

Resolution and Maps



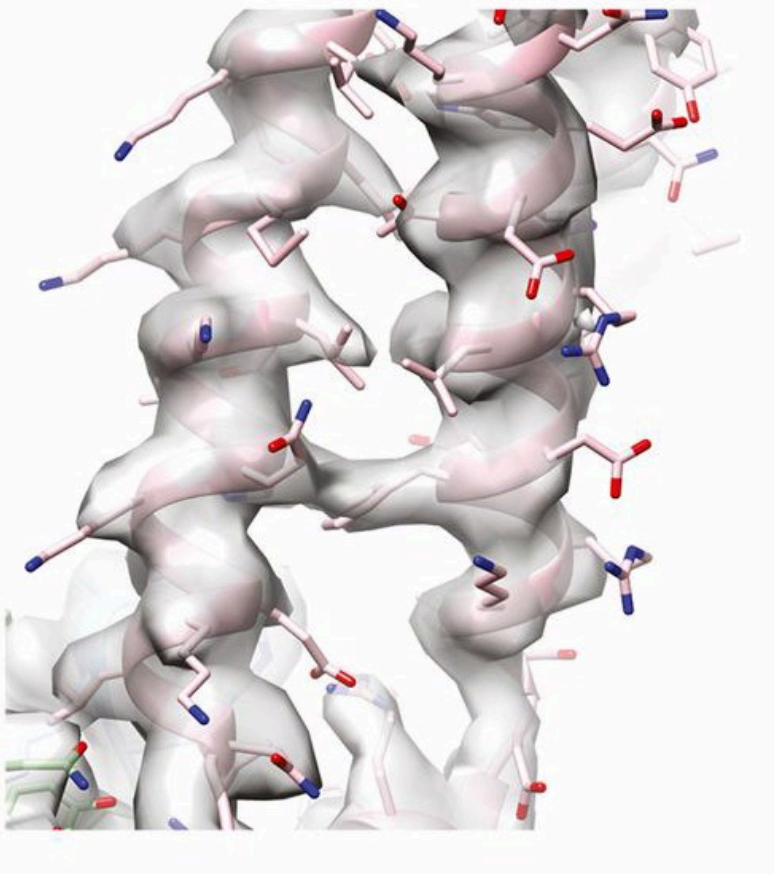


- Clearly much harder to accurately model the tryptophan sidechain here

- But it can be worse!

Maps and Models

- Important to understand the difference between maps and models:
  - Maps are the experimental data
  - Models are the interpretation of that data

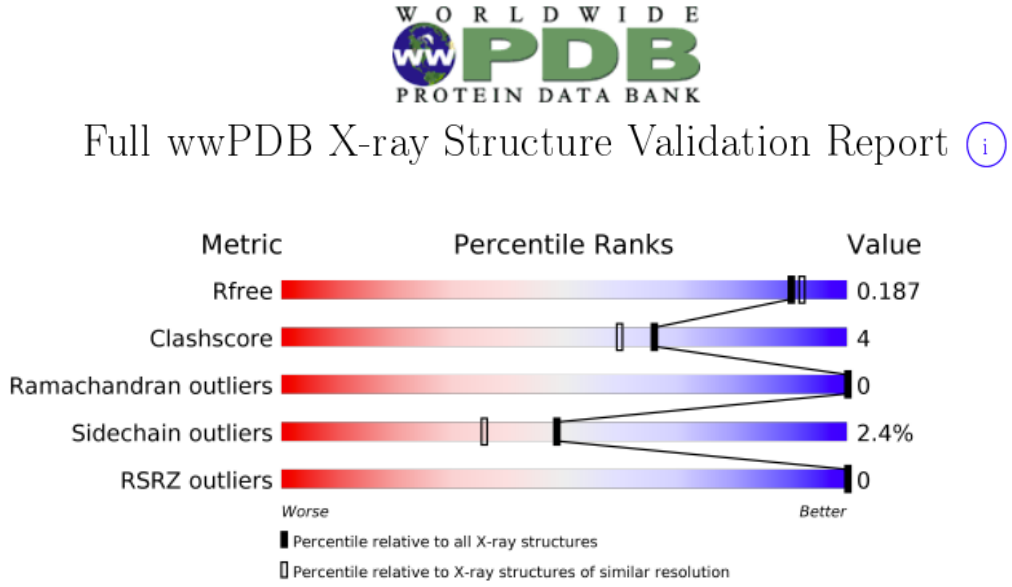


Judging Model Quality

- You can look directly at the maps using WinCoot
  - This is what we're doing later

Judging Model Quality

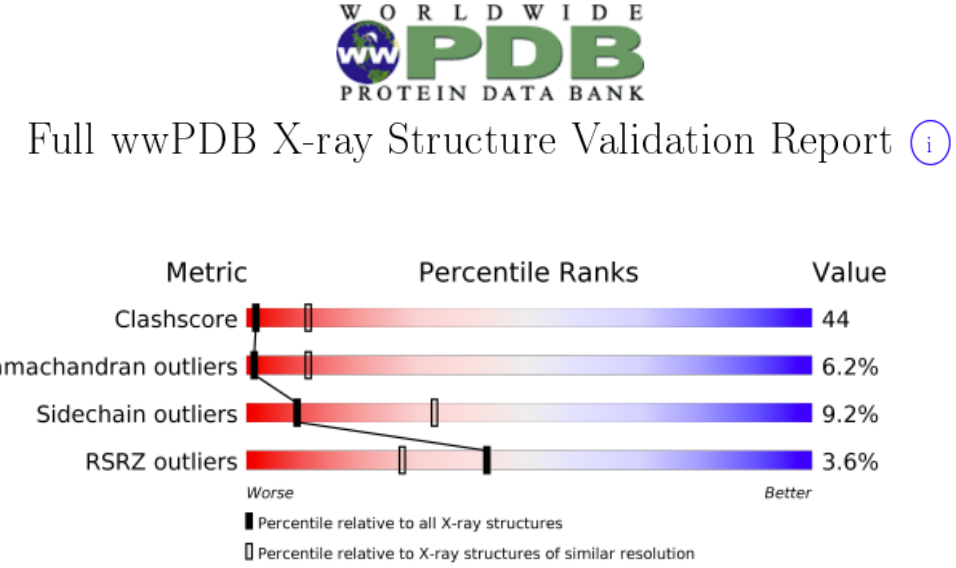
- You can look directly at the maps using WinCoot
  - This is what we're doing later
- Other useful statistics include:
  - PDB validation reports
  - Residuals (Rfree) in crystallography
  - Molprobity clash score
  - Ramachandron plot and outliers



PDB ID: 1CBS
Very good quality, 1.8 Å

Judging Model Quality

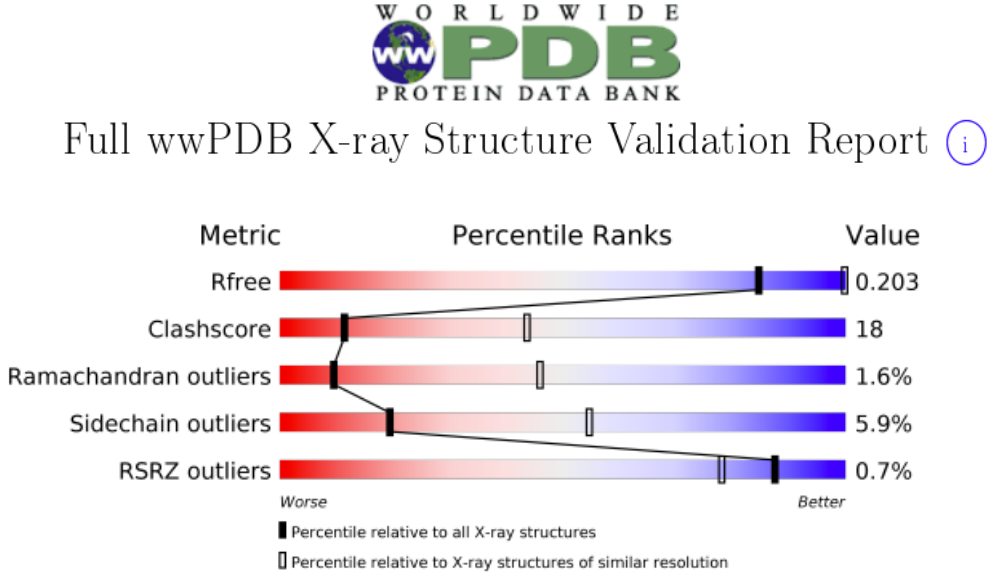
- You can look directly at the maps using WinCoot
  - This is what we're doing later
- Other useful statistics include:
  - PDB validation reports
  - Residuals (Rfree) in crystallography
  - Molprobity clash score
  - Ramachandron plot and outliers



PDB ID: 1FCC
Very bad quality, 3.2 Å

Judging Model Quality

- You can look directly at the maps using WinCoot
  - This is what we're doing later
- Other useful statistics include:
  - PDB validation reports
  - Residuals (Rfree) in crystallography
  - Molprobity clash score
  - Ramachandron plot and outliers



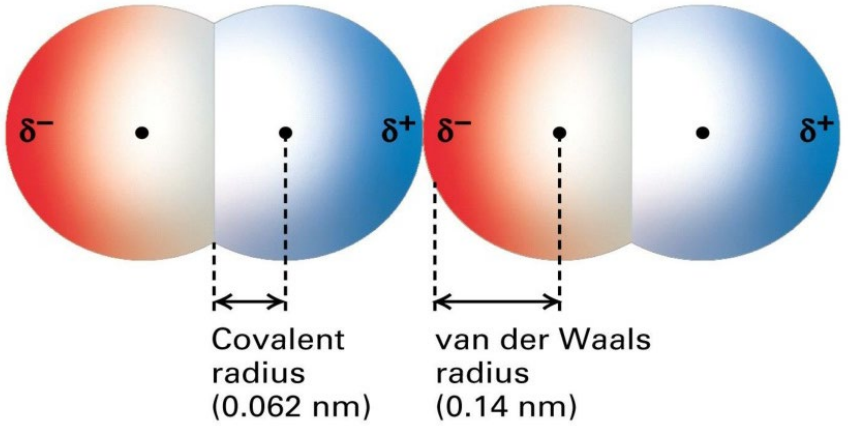
PDB ID: 1EG1
Mixed quality, 3.6 Å

Residuals (R_{free}) in crystallography

- Take a fraction of your original dataset (5%)
- Don’t use it for model building or refinements
- Use it to test how well your model fits these independent data
- Measure of the quality of your model
- Tests if you’re building into “noise” in your dataset
- Quoted R_{free} should go down as the model improves but not be >5% from R_{work}

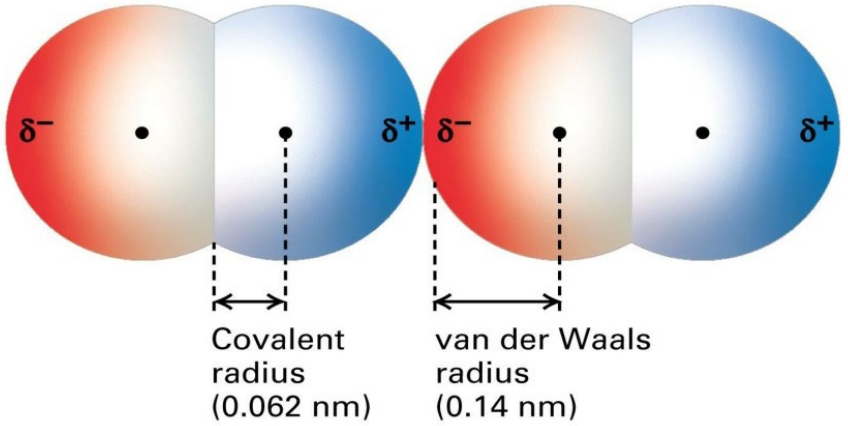
Molprobity Clash Score

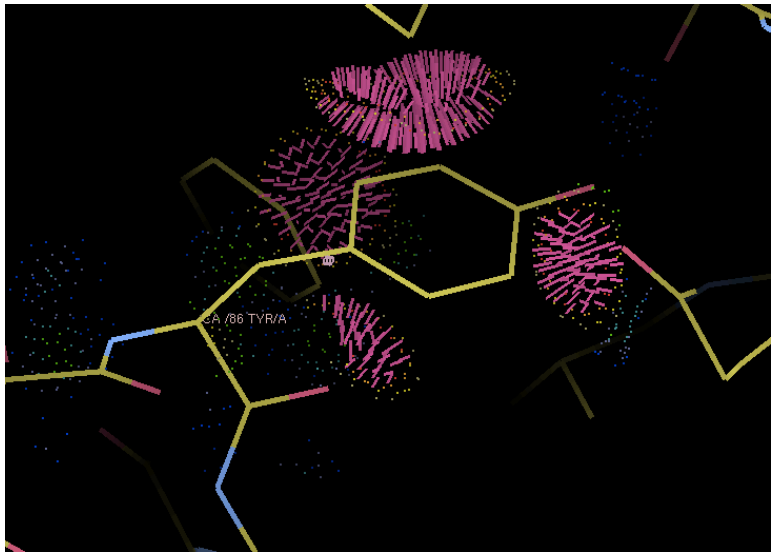
- Electrons distributed around atoms form a van der Waals radius
- If atoms are too close the electron clouds would “clash”



Molprobity Clash Score

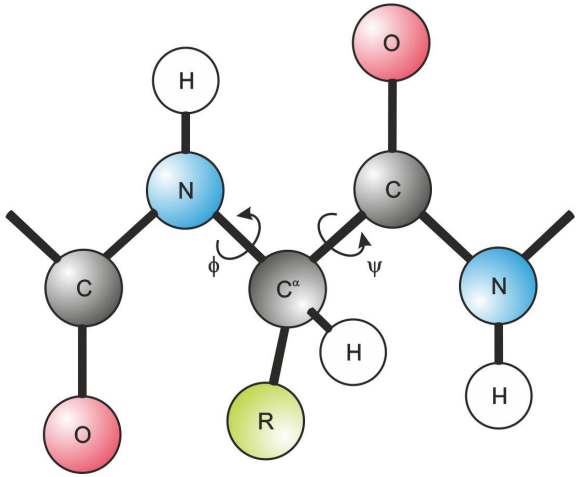
- Electrons distributed around atoms form a van der Waals radius
- If atoms are too close the electron clouds would “clash”
- Can display this in Coot
- Also Molprobity calculates this in its “clash score”





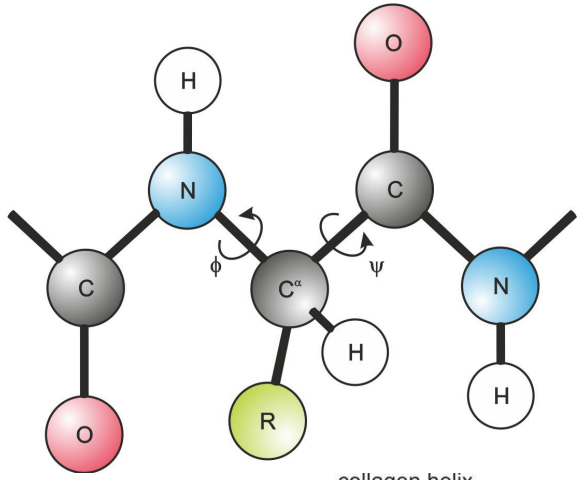
Ramachandran Plots

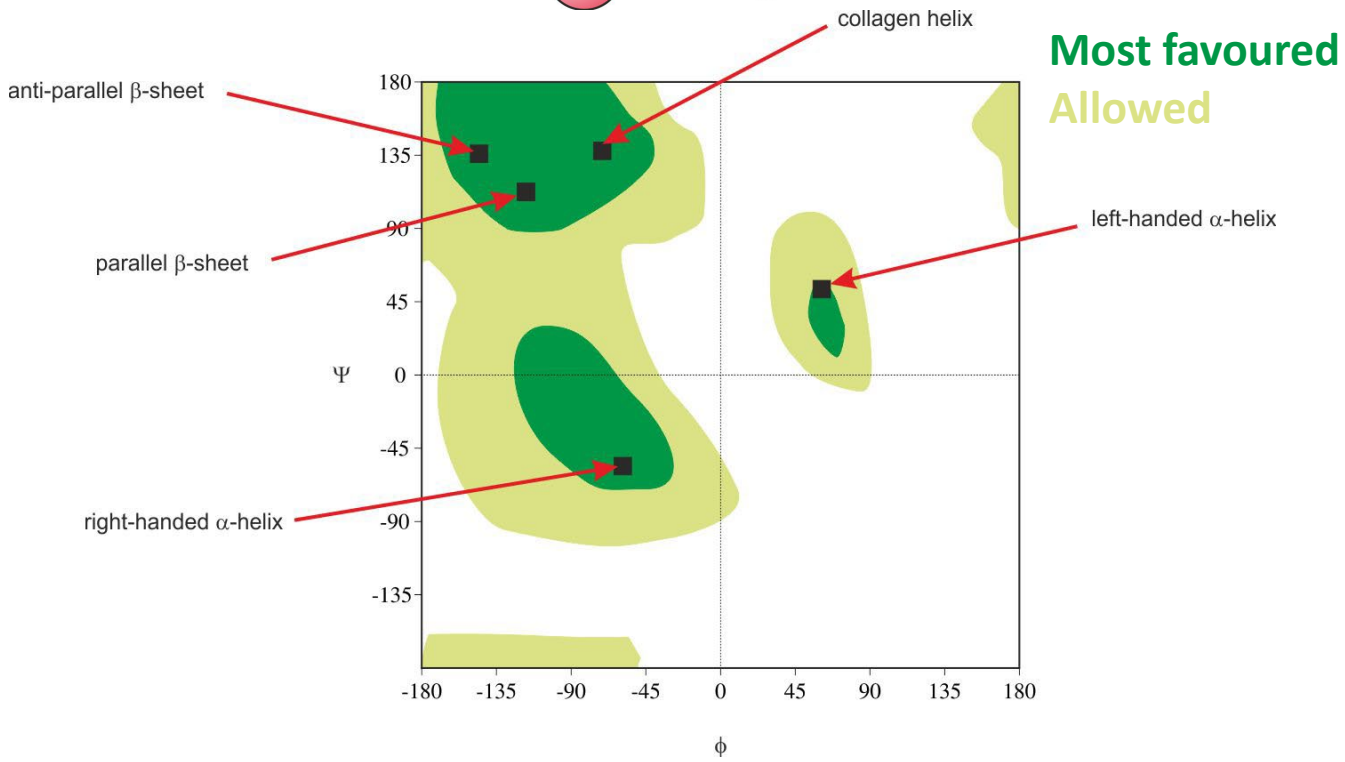
- Allowed geometry based on chemistry of the peptide backbone



Ramachandran Plots

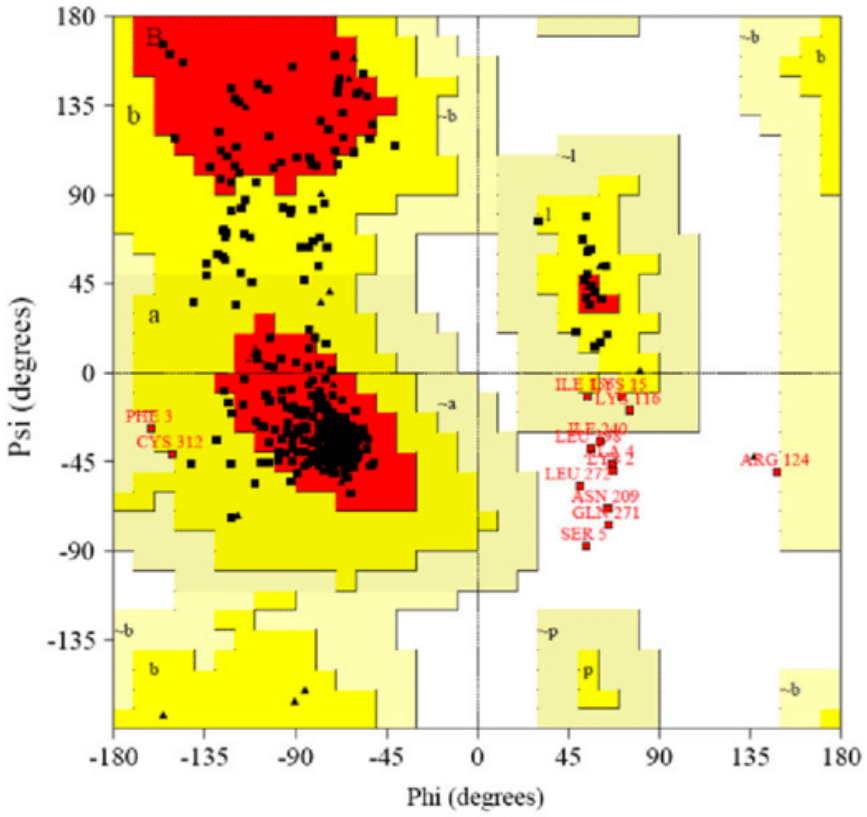
- Allowed geometry based on chemistry of the peptide backbone
- Plot of the backbone phi (ϕ) and psi (ψ) torsion angles





Ramachandran Plots

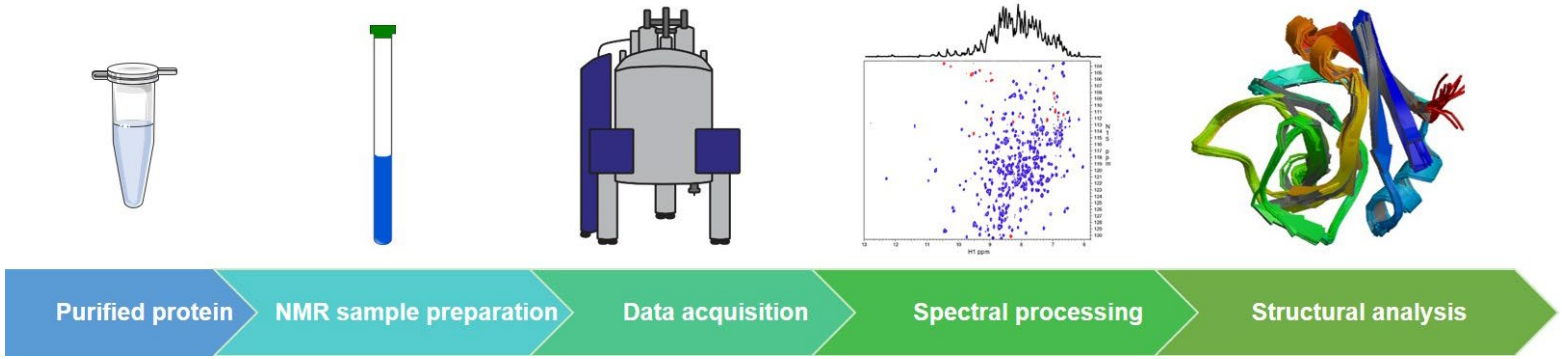
- Allowed geometry based on chemistry of the peptide backbone
- Plot of the backbone phi (ϕ) and psi (ψ) torsion angles
- Residues in disallowed regions are likely modelled wrongly



Other techniques – solution structures

Other techniques – solution structures

- NMR: nuclear magnetic resonance

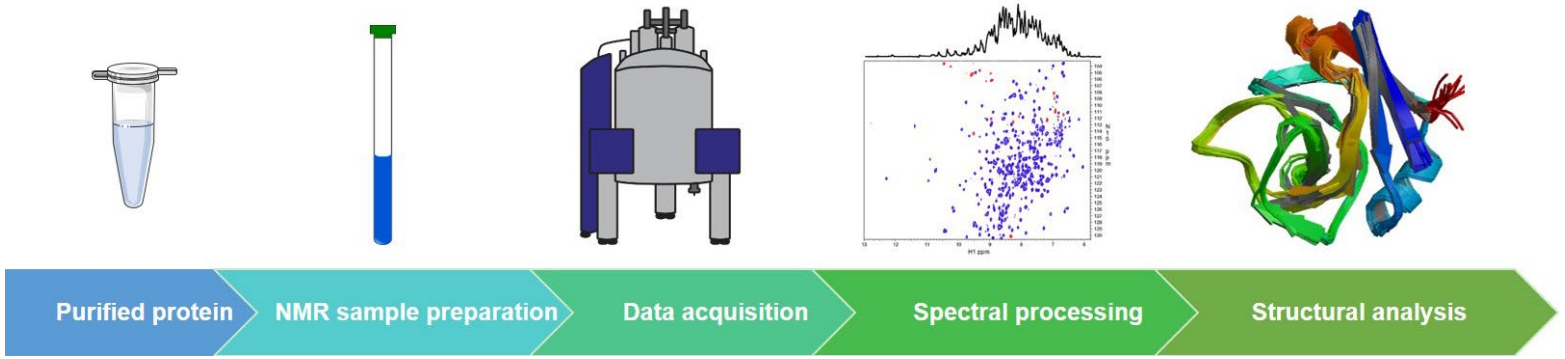


Limited to small proteins

Good for dynamics

Other techniques – solution structures

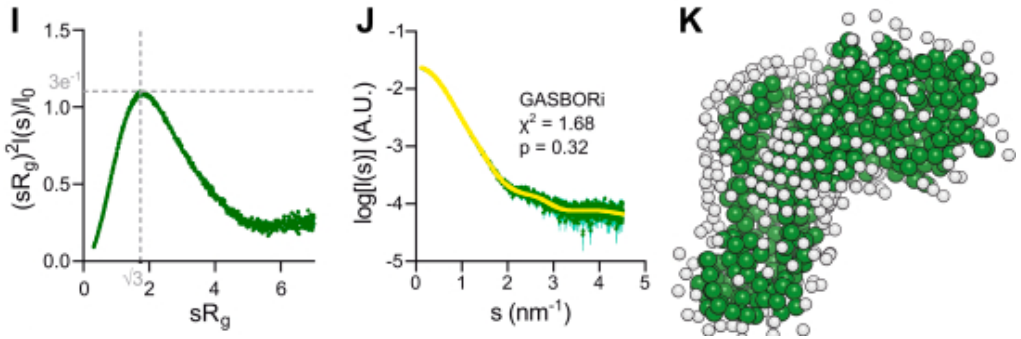
- NMR: nuclear magnetic resonance



Limited to small proteins

Good for dynamics

- SAXS: small-angle X-ray scattering



Low resolution

No size limit

Pros and Cons of Different Methods

- These are all complementary but there are some important differences

NMR

Pros
Captures dynamics
In solution structure

Cons
Proteins need to be small
Typically low pH buffers

X-ray Crystallography

Pros
Can do small and big proteins
Relatively cheap

Cons
Has to crystallise!
Typically captures one conformation
Can have crystal packing artefacts

Cryo-EM

Pros
Small amounts of sample
No “phase problem”

Cons
Can’t do small proteins (<150kDa)
Expensive
Protein must tolerate freezing

Experimental structure determination

- Today we learnt about how to experimentally determine a structure
  - Techniques for determining structures
  - How to collect and process structural data
  - How to build a new structure
  - How to judge the quality of a structure

- Tomorrow we will learn how to
  - Predict a structure using AlphaFold
  - How to evaluate the quality of this prediction
  - What AlphaFold can (and can't) do