Protein Crystallography

Tara Christie

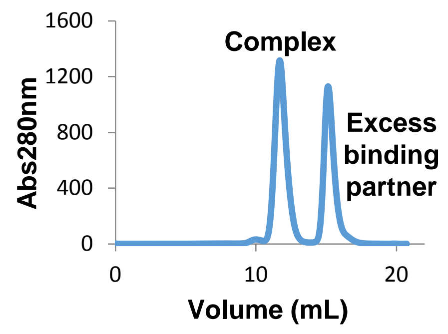
1st Asia Oceania Synchrotron School

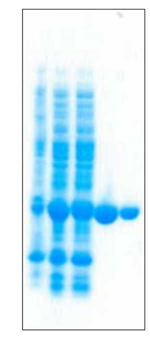
Overview

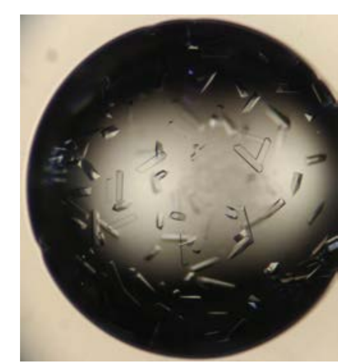
- Protein crystallography- what and why?
- Sample preparation
- Crystallisation
- Collecting data
- Data processing

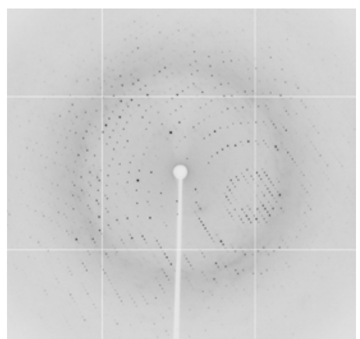
and refinement

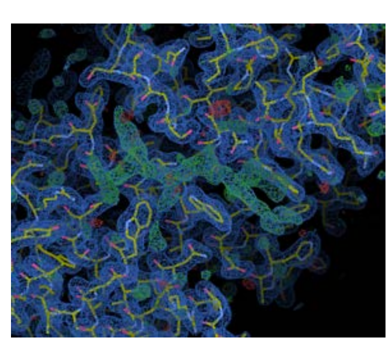
- Example

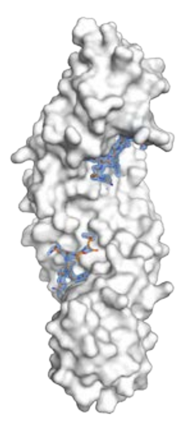






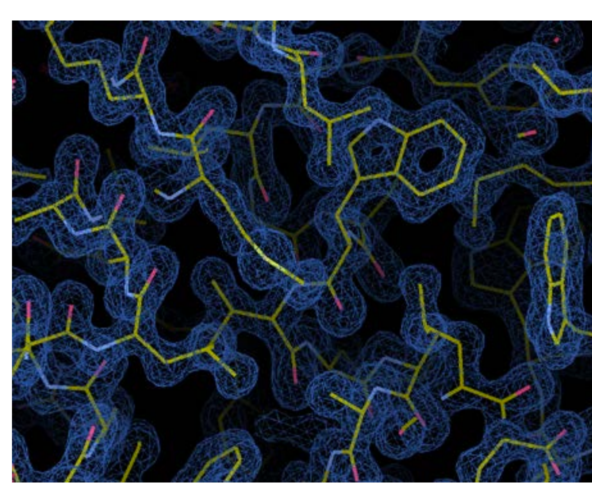


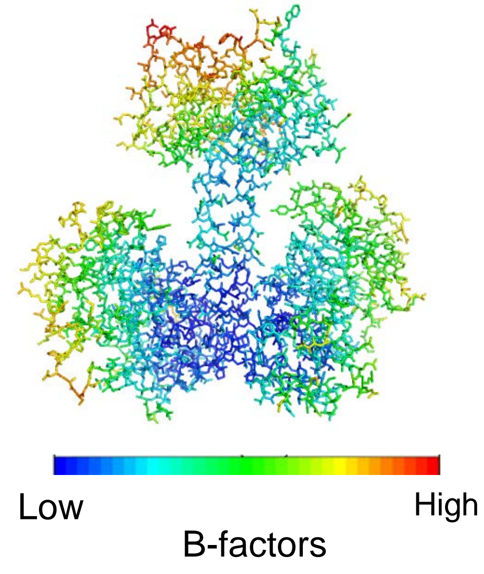




What is protein crystallography?

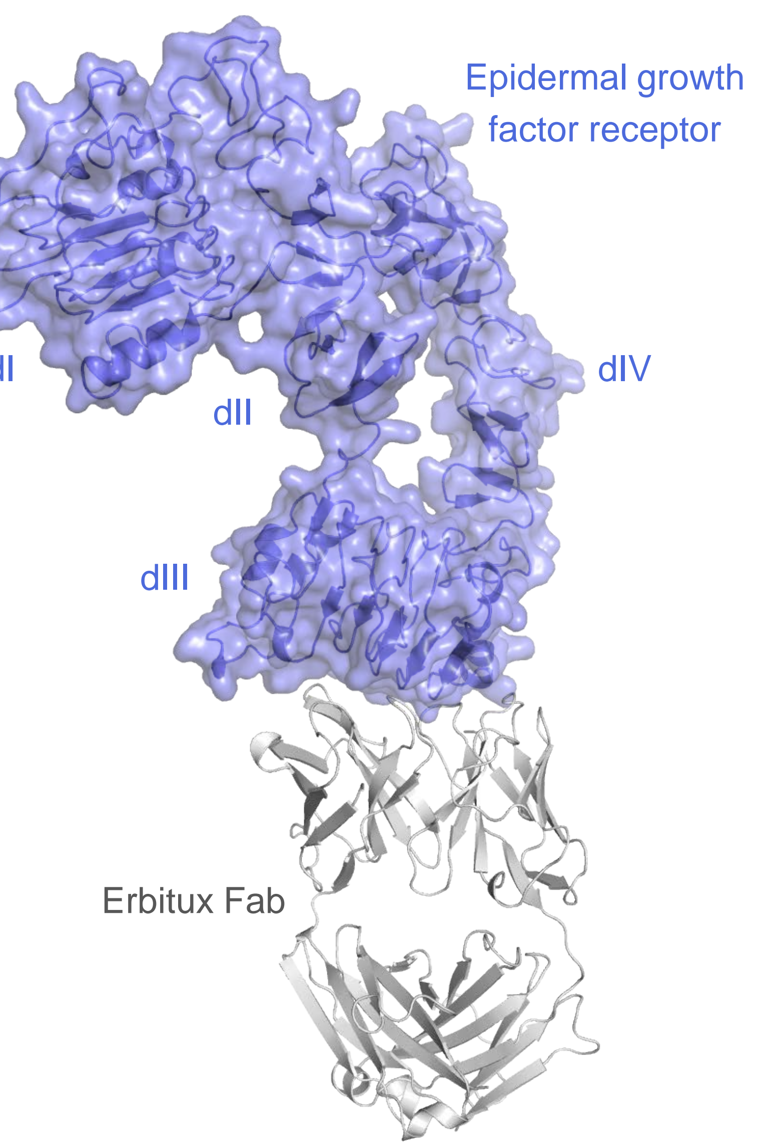
- Determination of the atomic structure of a macromolecule
  - 3D atomic positions (x,y,z)
  - Limited information on flexibility/ dynamics (poor/ absent electron density, B-factors)





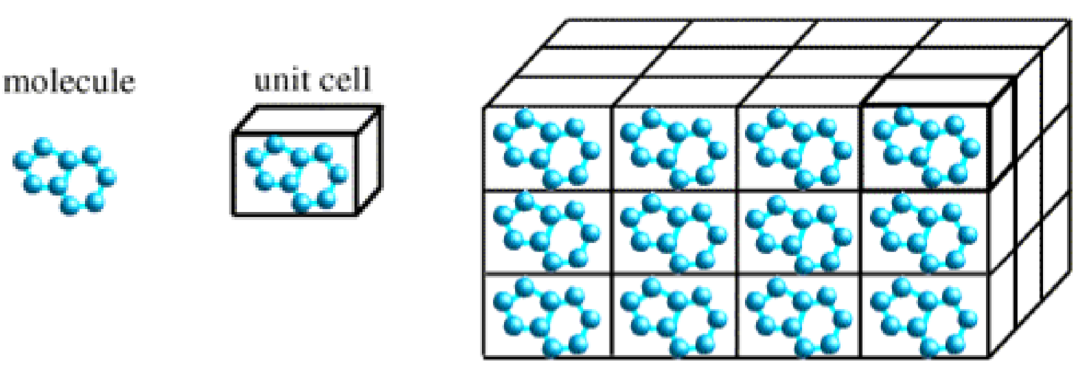
Why should we study protein structure?

- Molecular structure is related to protein function
  - Interaction with small molecules/ nucleic acids
  - Interaction with binding partners
  - Understanding catalytic mechanism/ mode of action
  - Drug design



Why crystals?

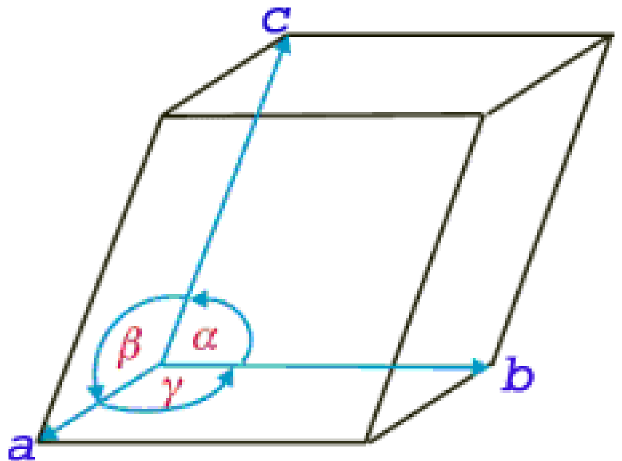
- Diffraction of x-rays with single protein weak
- X-rays are ionising so high doses would destroy the sample
- Crystals are a repeating arrangement of objects in a lattice
- Amplify the signal from x-rays interacting with electrons of atoms/ molecule,



The building block of a crystal is the unit cell, which is translationally repeated in three dimensions to form the crystal lattice

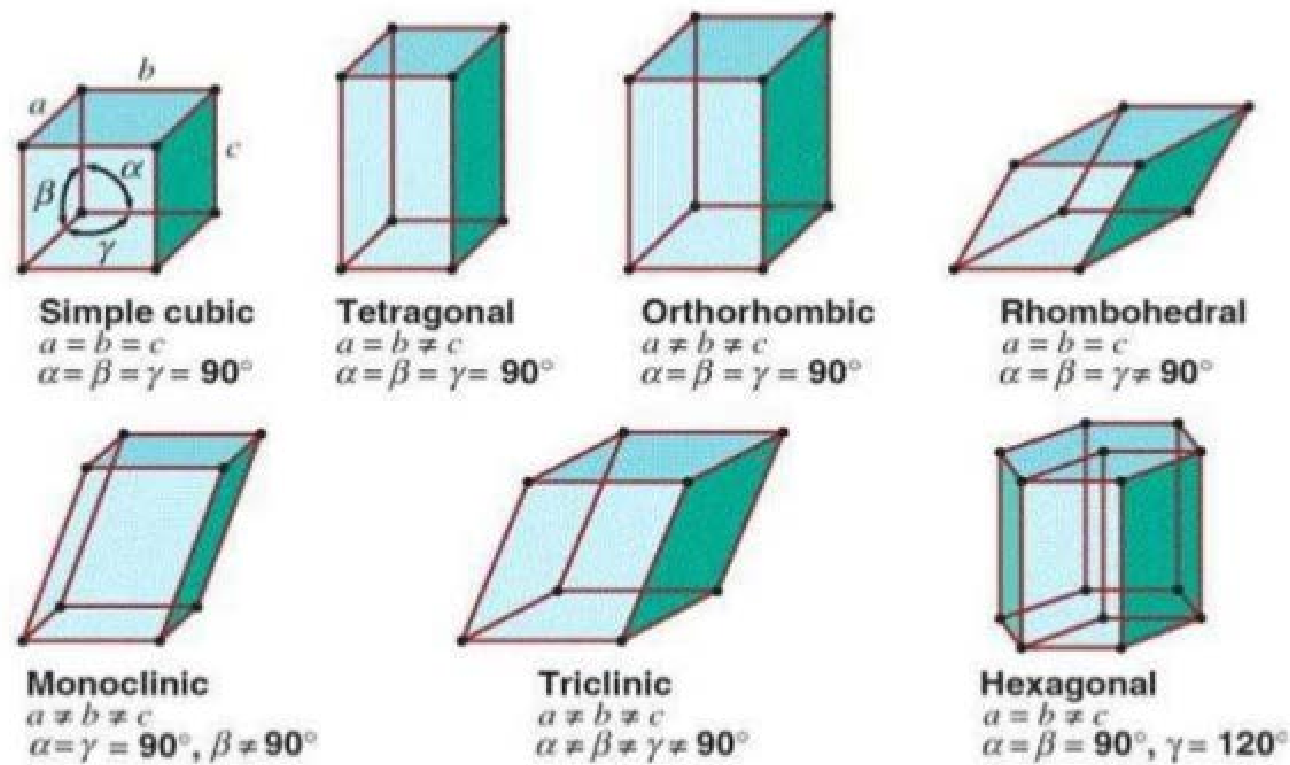
The unit cell

- A unit cell can be described by the length of its three edges (a, b, c) and the angles between them (**α**, **β**, **γ**).
- Parallelepiped- there are 4 edges to a face, and 6 faces in a unit cell.
- All unit cells within a crystal should be identical



Crystal systems

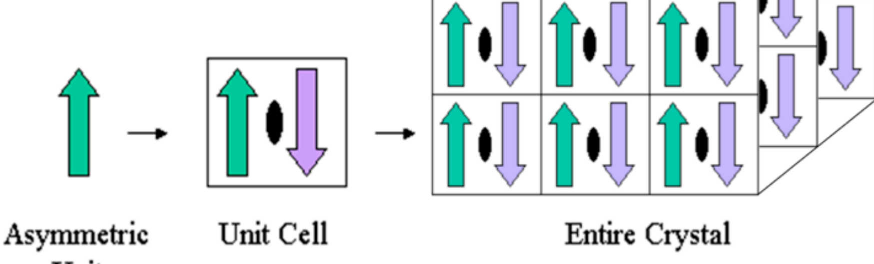
- The crystal lattice is comprised of objects that can be arranged to fill three dimensional space completely.



Space groups

- The unit cell can be comprised of more than one protein molecule related by a symmetry operation.
- Only 2-, 3-, 4- and 6- fold rotational symmetry operations are allowed.
- Screw axis- combination of rotation and translation
- Mirror planes are not allowed in protein xtals.

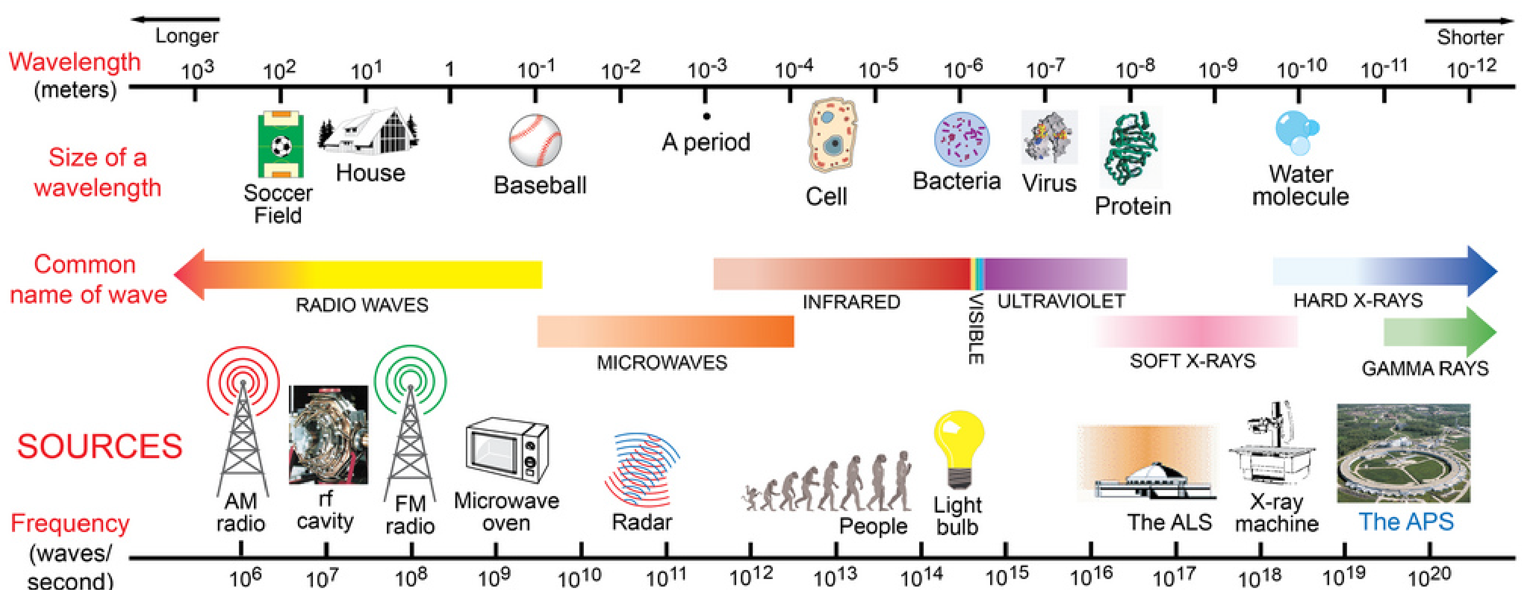
The unit cell has two fold symmetry (rotated 180^0 about an axis shown as a black oval).



RCSB PDB

Why x-rays?

- What we resolve (see) is limited by the wavelength of light used
- To resolve atomic scale (bonds, atoms) ie. ~0.5-4 Å, we need to use x-rays



Electromagnetic spectrum

© Advanced photon source

Synchrotron radiation

- Advantages compared to a laboratory x-ray generator
  - More intense radiation
  - Time- minutes (seconds!) instead of hours/ days
  - Tunable wavelengths- identification of metal(s) bound to protein, sulfur/ native-SAD phasing
  - Beam size- collect data on best section of crystal, reduce background

Australian Synchrotron

Protein preparation

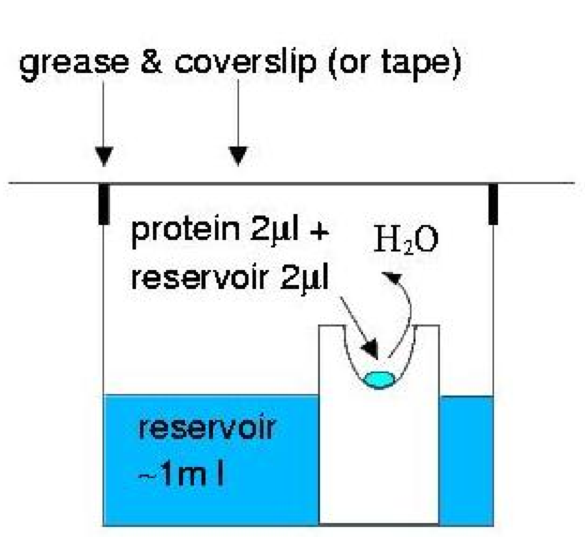
- Clone suitable construct (Full length/ domains/ N- or C-terminal truncations etc.) into a suitable vector with affinity tag (eg. His-tag)
- Express protein of interest in E.coli/ yeast/ SF9/ HEK cells. Need mg amounts of protein. Post-translational modifications?
- Purify protein using affinity matrix, (ion exchange, heparin, etc.), with a final gel filtration step
  - Buffer exchange
  - Monodisperse, non-aggregating
- Purity- run on gel
- Concentrate enough to reach supersaturation

Crystallisation

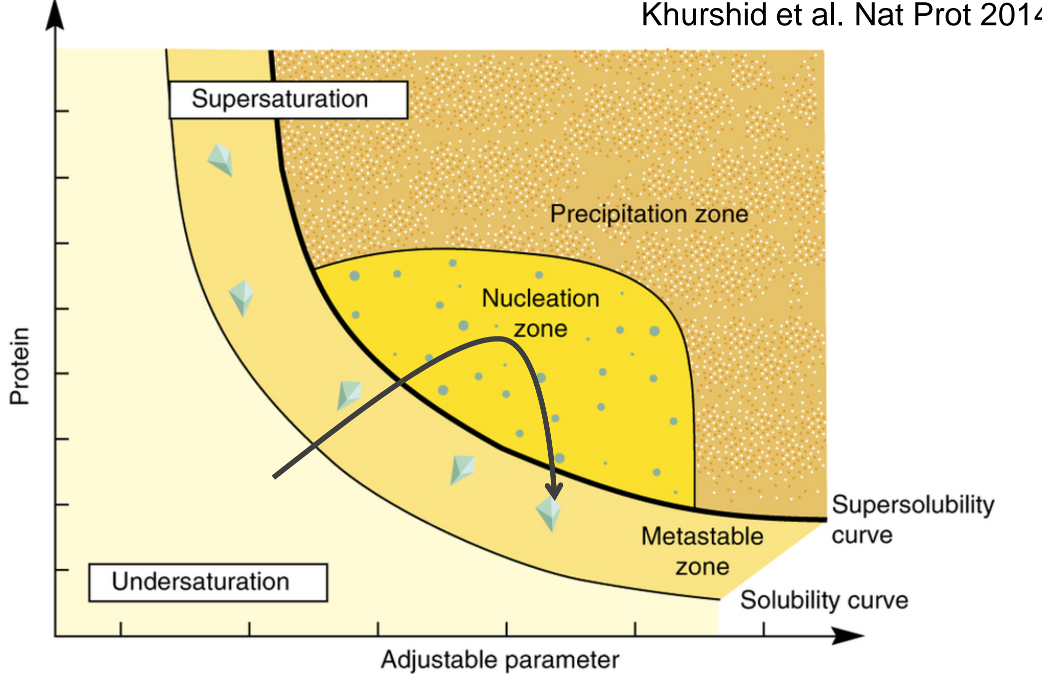
- Screen ~100's-1000's of conditions
  - Commercially available screens (Hampton, Molecular dimensions, Jena, Rigaku)
  - Temperature
  - with additives (small molecules), different protein concentration, different protein: reservoir ratios, drop size etc.
  - Different experiment types, vapour diffusion most common

Vapour diffusion

- The protein drop solution has lower reagent concentration than the reservoir solution. Water equilibration throuh the vapor phase raises both precipitant and protein concentrations in the drop.



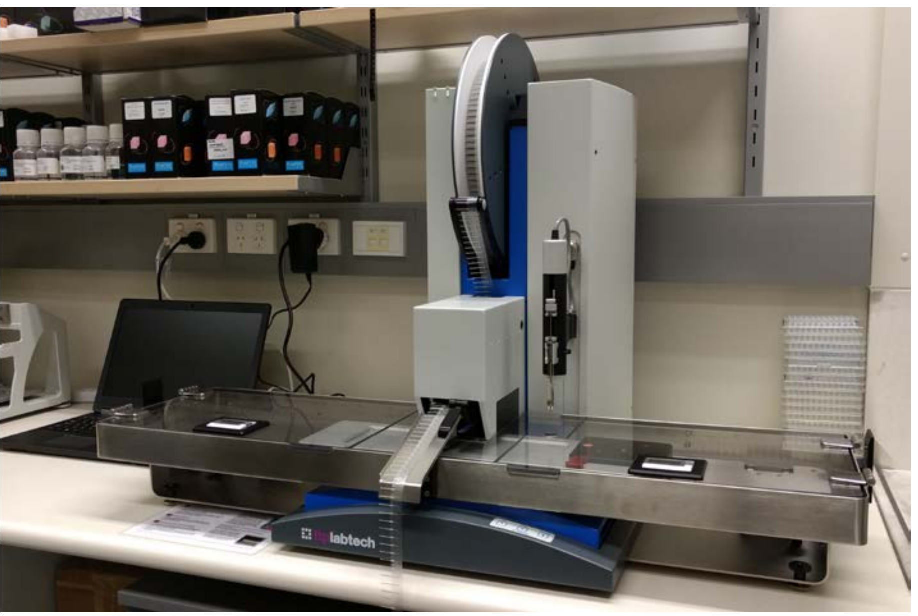
Khurshid et al. Nat Prot 2014



Preicpitant/ additive/ pH/ temperature etc.

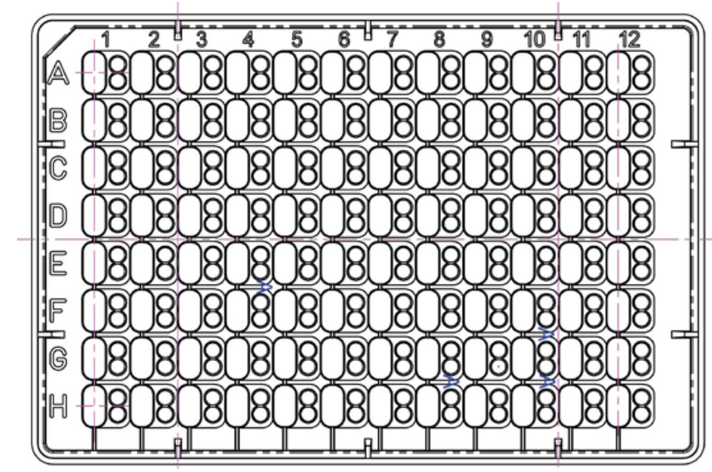
Crystallisation

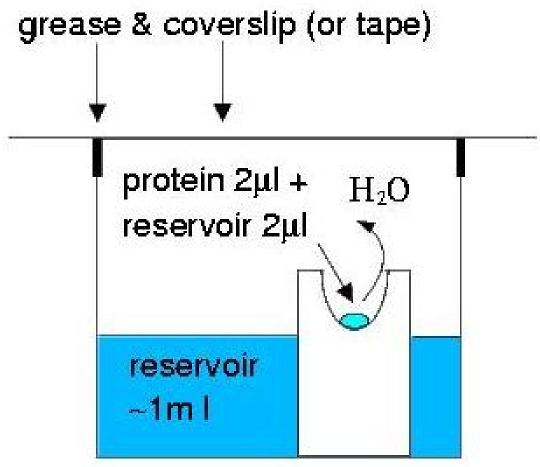
- Use a robot
  - 50-100 nL protein per well
  - 96 wells in ~ 2 mins
- Use 2-16 screens per protein
- Temperature 20ºC



Mosquito, TTP LabTech

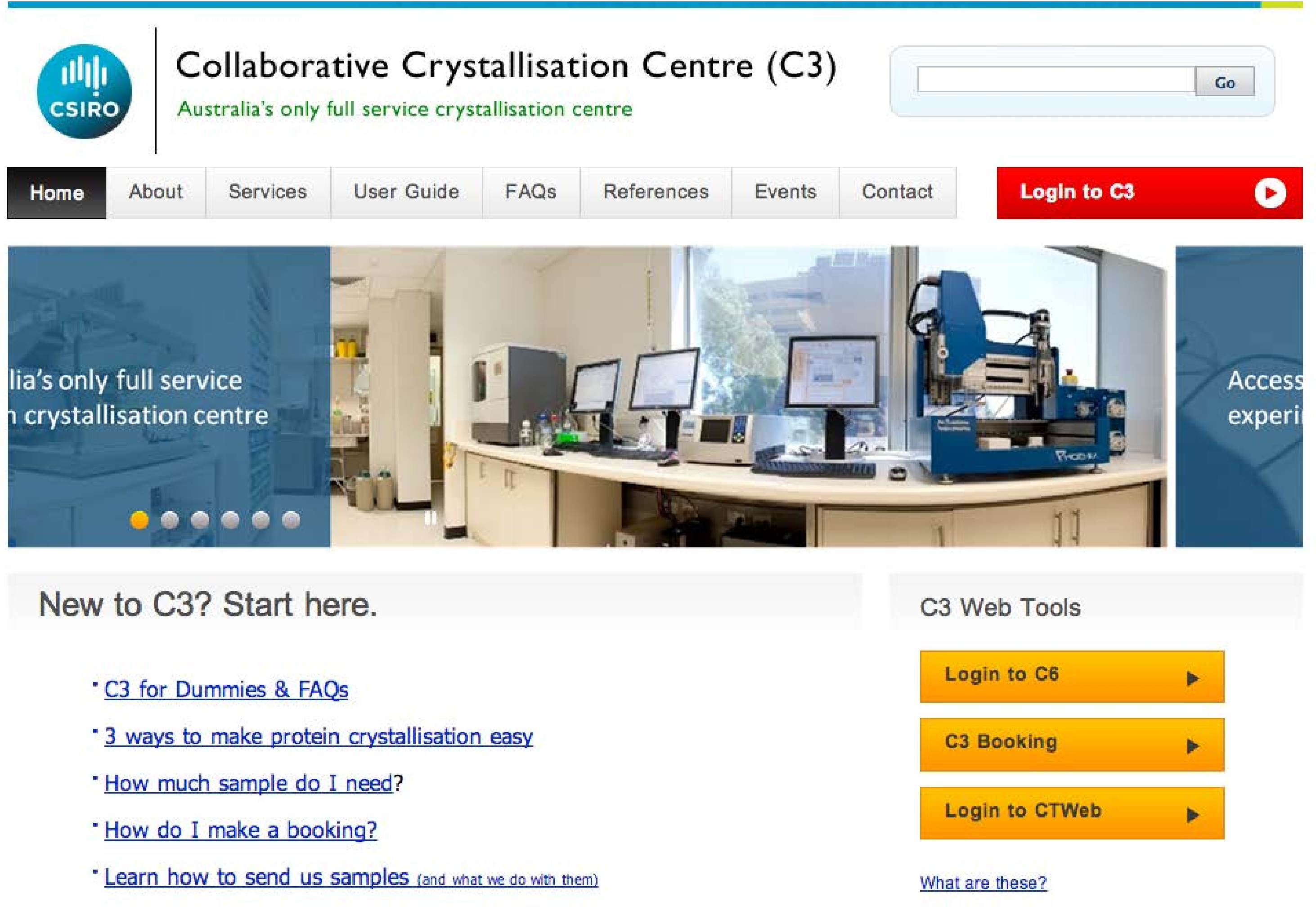
MRC plate





Airlie McCoy, University of Cambridge

Collaborative Crystallisation Centre (C3)

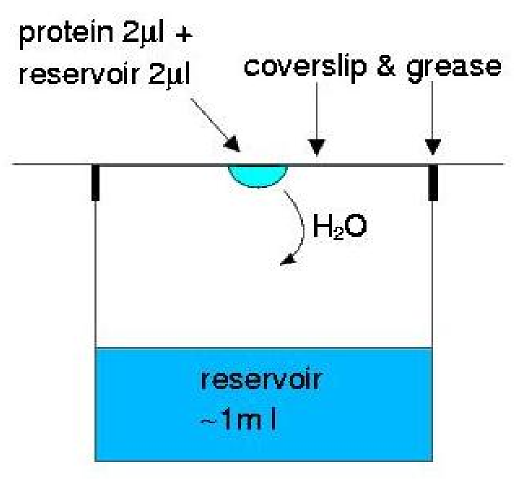


Janet Newman, CSIRO

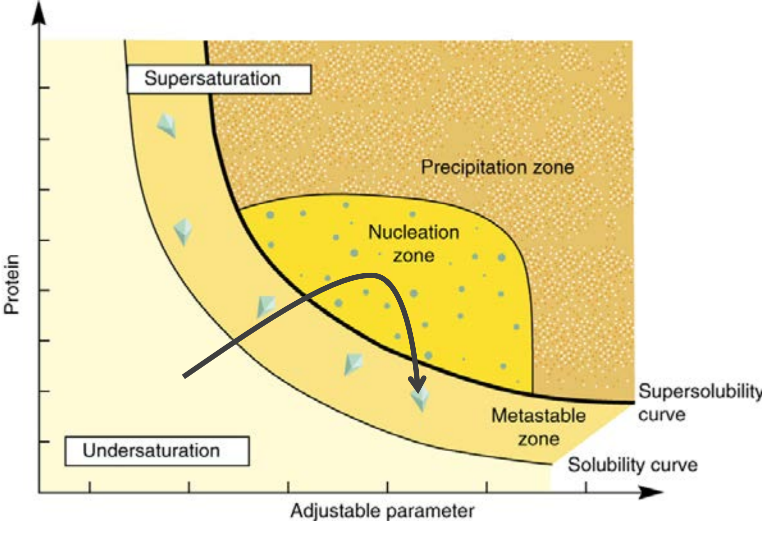
Crystal optimisation

- Fine screening of components
- Additive screen (small molecules, Hampton Research)
- Change kinetics (drop size, temperature, ratio)
- Add small amounts of cryo?

VDX plate
(Hampton research)



Airlie McCoy
University of Cambridge



Khurshid et al. Nat Prot 2014

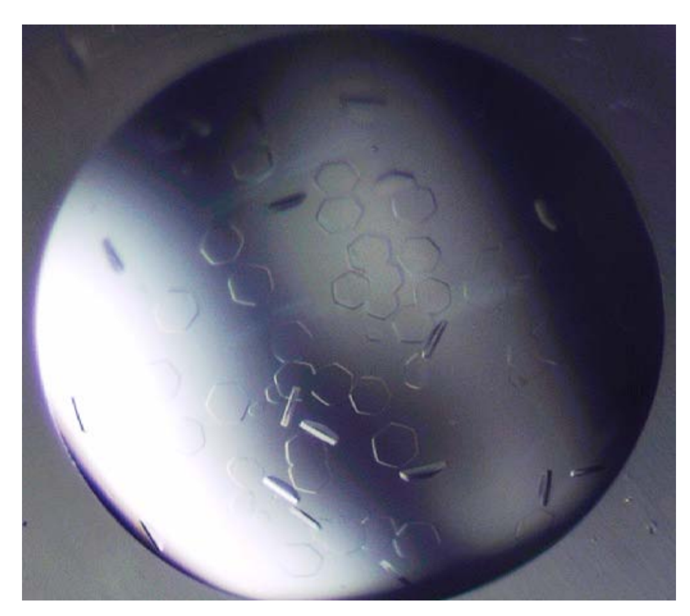
Crystal optimisation

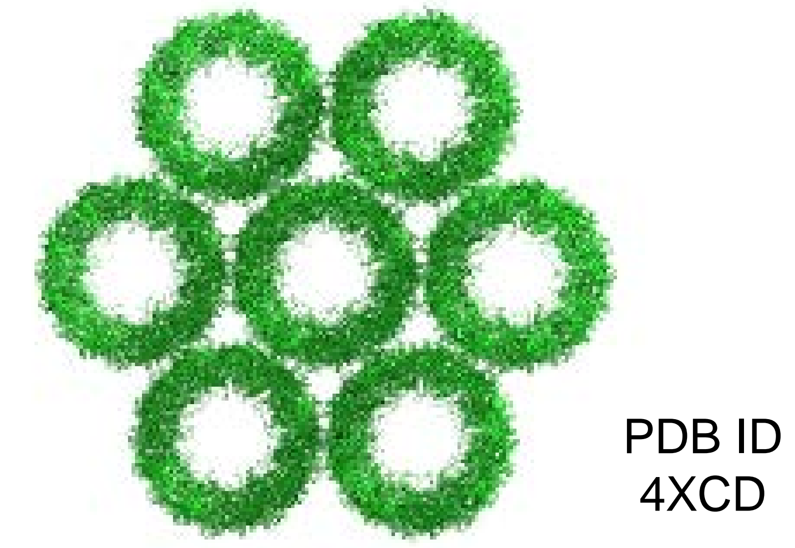
- Seeding
  - Crush up a crystal from a previous experiment and transfer to a new drop by streak seeding (whisker, horse hair, needle)
  - Can do multiple rounds of seeding

Crystal optimisation by iterative seeding

Crystal symmetry

- Appearance indicative of internal symmetry





PDB ID
4XCD

Hexagonal crystal system

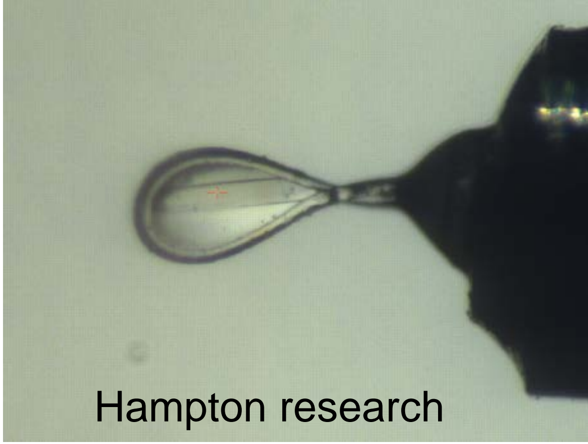
$a = b \neq c$; $\alpha = \beta = 90°$, $\gamma = 120°$

Alastair Stewart, VCCRI

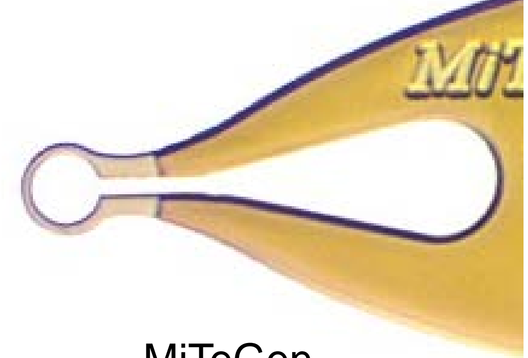
Crystal mounting

- Loop crystal- Cryoprotect (but should also test at room temperature)
  - Reservoir + glycerol (eg. 20% glycerol, ethylene glycol). Time?

https://www.youtube.com/watch?v=i2G1fYtjXt8



Hampton research



MiTeGen

Crystal freezing

- Plunge into liquid nitrogen
- Transfer crystal into puck
- Send frozen pucks to synchrotron
- Freeze xtal at convenient time

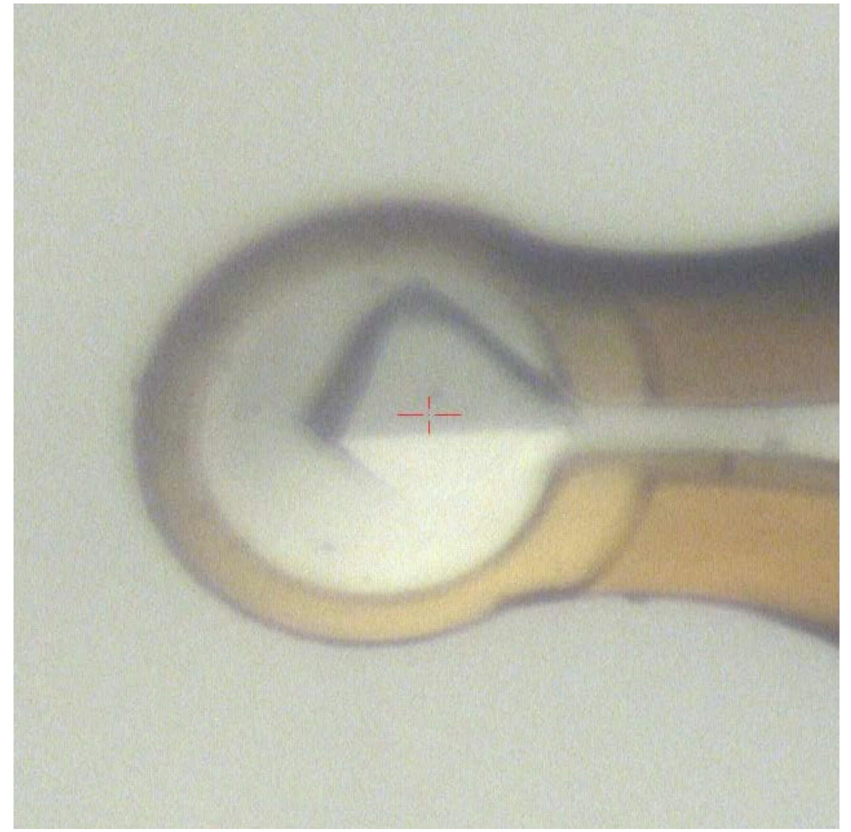
Data collection

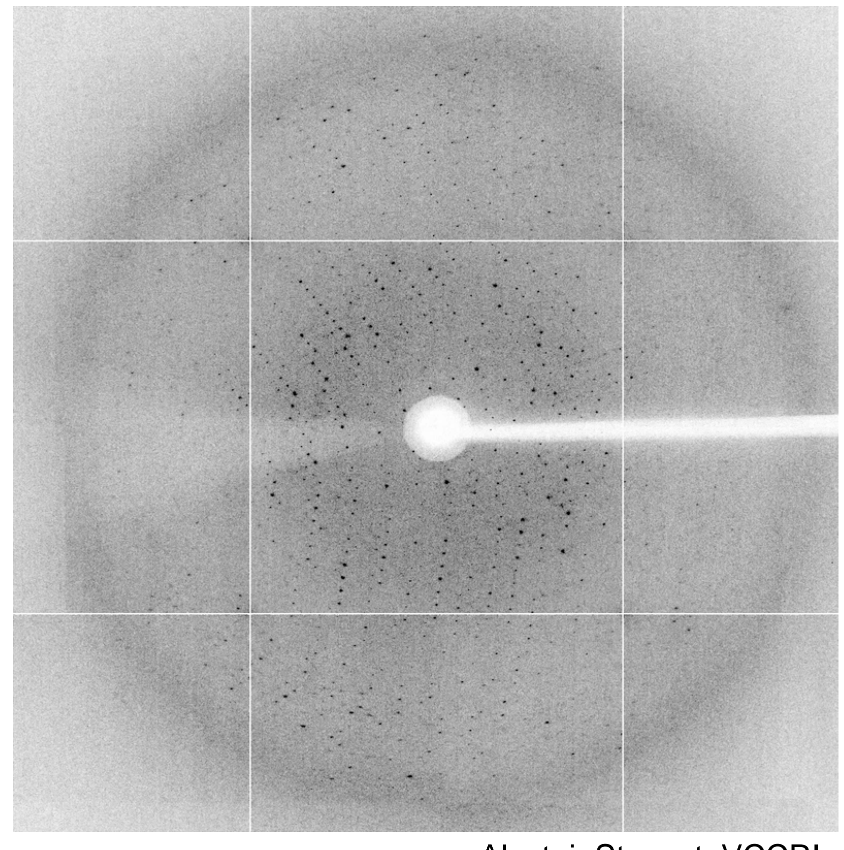
- Robot can be driven on site (hutch) or remotely
- Mount crystal on goniometer
- Center on crystal using Blu-Ice
- Expose crystal to x-rays (find a good part?)
- Diffraction patterns recorded on detector
  - 16M Eiger detector MX2
    - Continuous read out, photon counting
    - Small pixel size, 360^0 in 36s
  - ADSC detector on MX1

EIGER X product pages...

Data Collection

When we expose a crystal to x-rays, some Bragg planes will be in the correct orientation (diffraction condition) and we will see spots (reflections) for them. By turning the crystal, different sets of planes will then be in the correct orientation and we will see new diffraction spots.

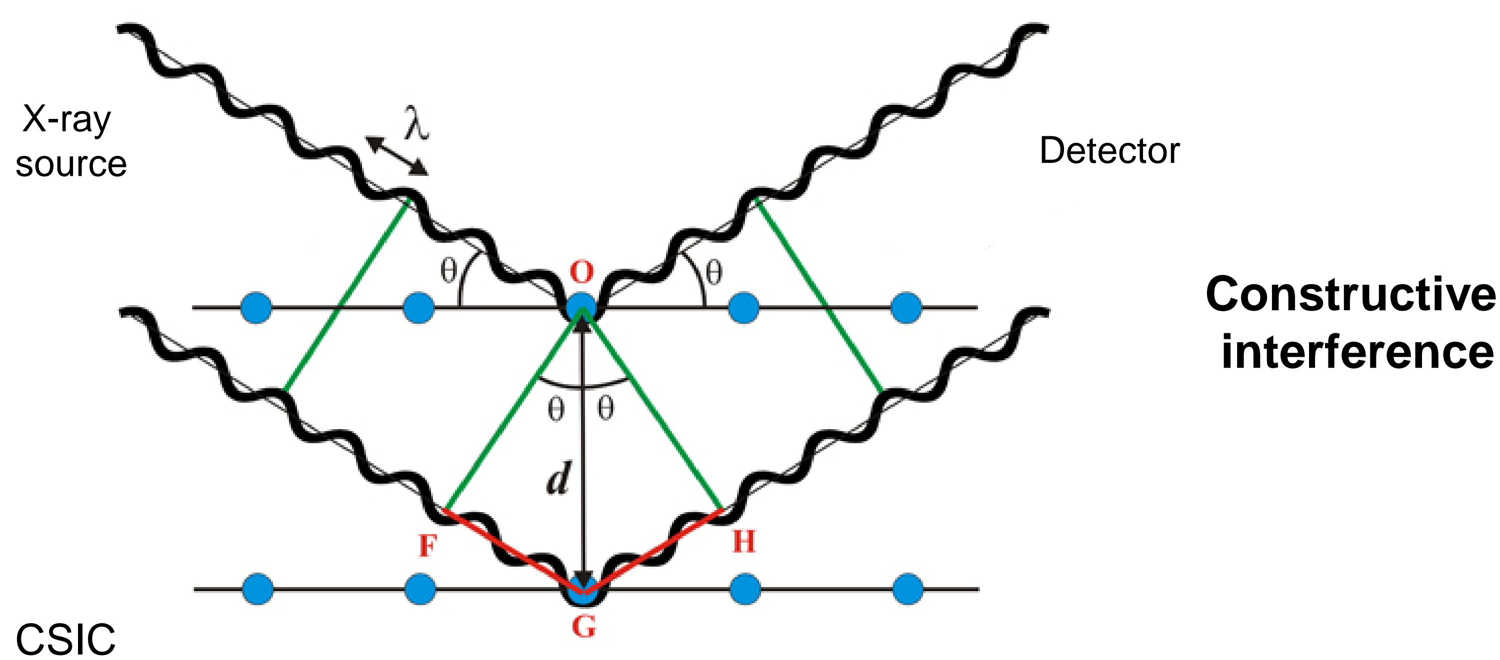




Alastair Stewart, VCCRI

Bragg's Law $n\lambda = 2d \sin \theta$

Scattering is 'visible' when the waves are in phase and constructively interfere. Ie. the path difference between waves scattered by adjacent lattice planes is an integer number of the x-ray wavelength.

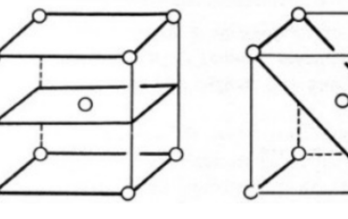
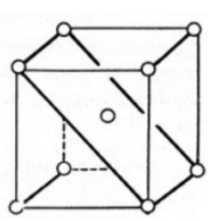
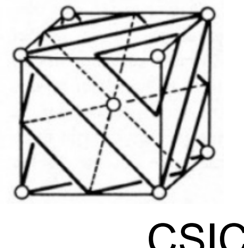


Detector

Constructive interference

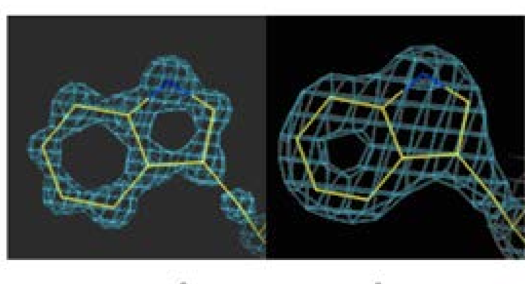
Data collection

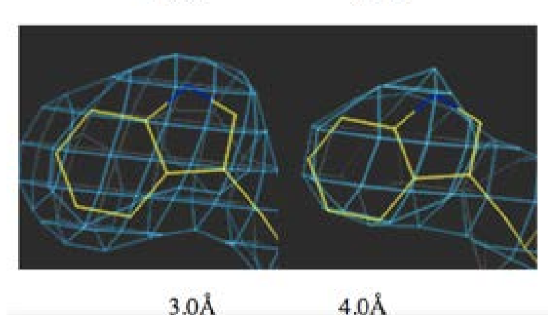
- Collect dataset via oscillation method
  - Crystal is rotated and different sets of reflections are recorded

CSIC

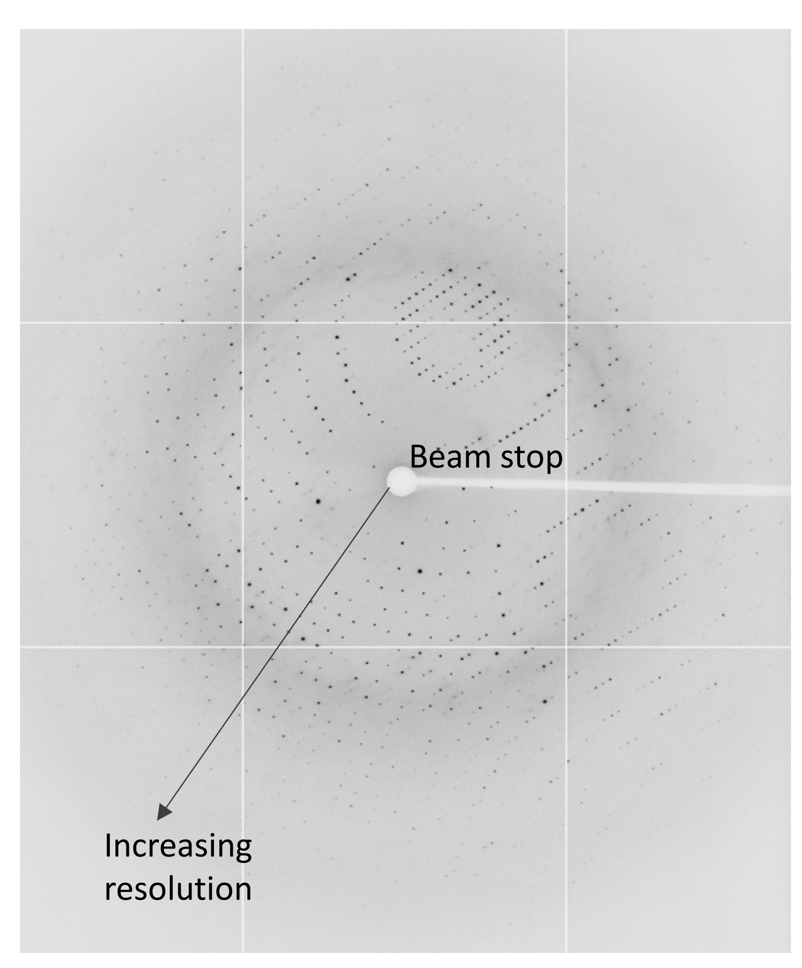
- Considerations:
  - Maximise resolution, be careful of radiation damage
  - High multiplicity
  - Spot separation, fine slicing
  - Use whole detector area





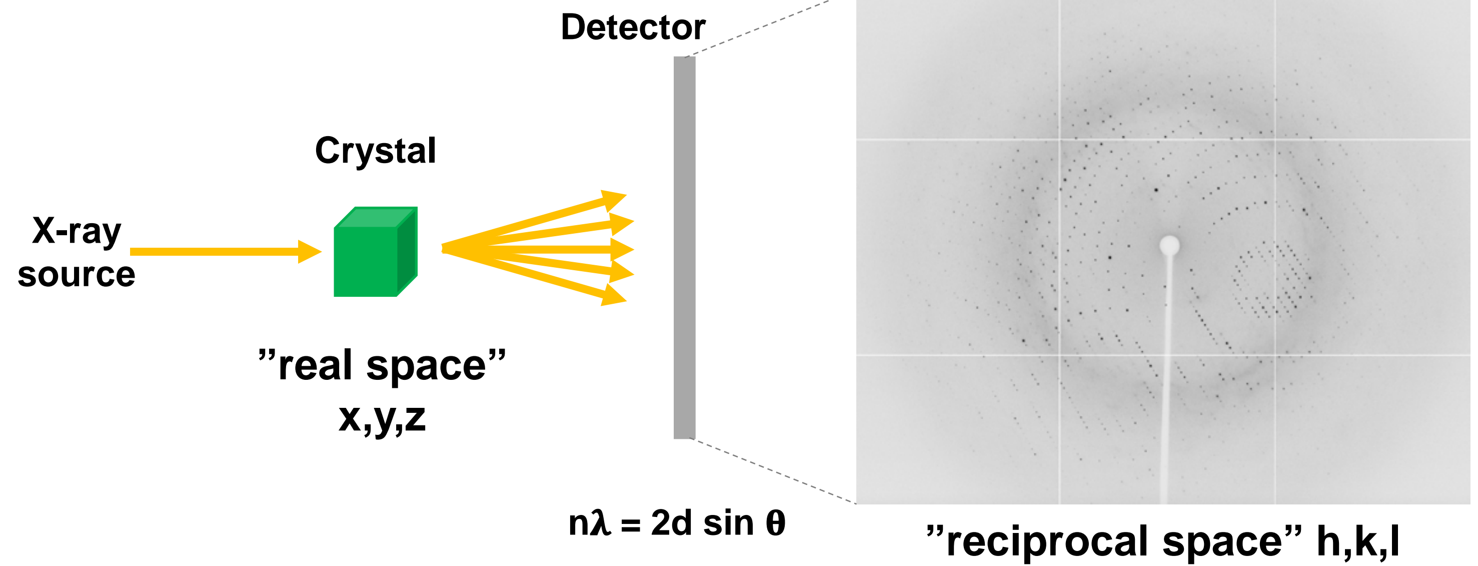
Diffraction from a protein crystal

- Position of reflections related to crystal lattice, unit cell dimensions, orientation of your crystal
- Intensity of reflections related to crystal components (electron density of protein)
- Number of reflections related to size of unit cell (a larger cell has more reflections at a given resolution)
- High resolution reflections are far from the center of the diffraction image



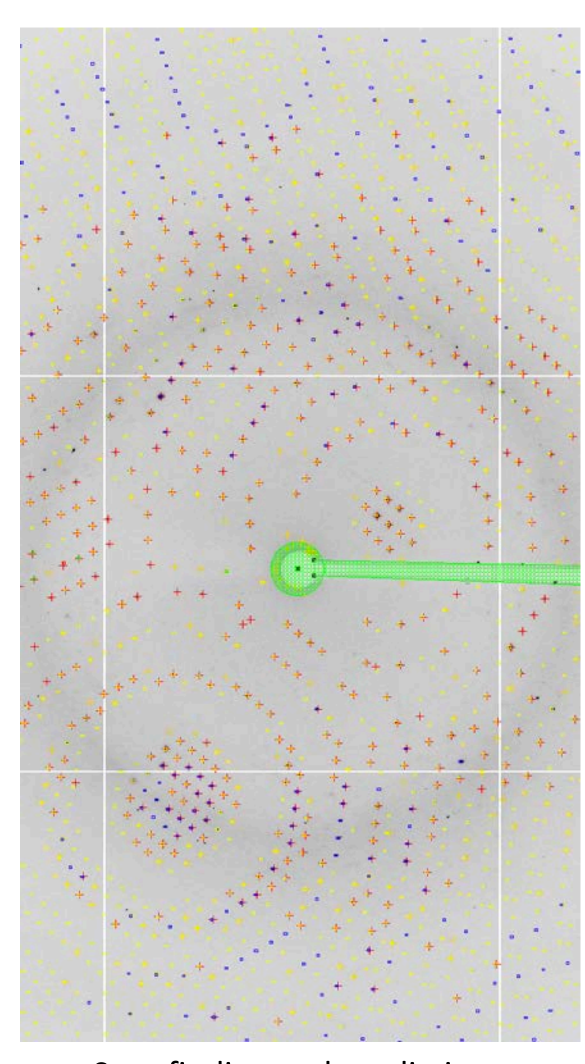
Reciprocal vs real space

- The diffraction pattern recorded is related to the crystal lattice that gives rise to it, but this relationship is reciprocal
  - Large crystal axes cause short distances between reflections, while small crystal axes cause large distances between diffraction spots.



Data processing

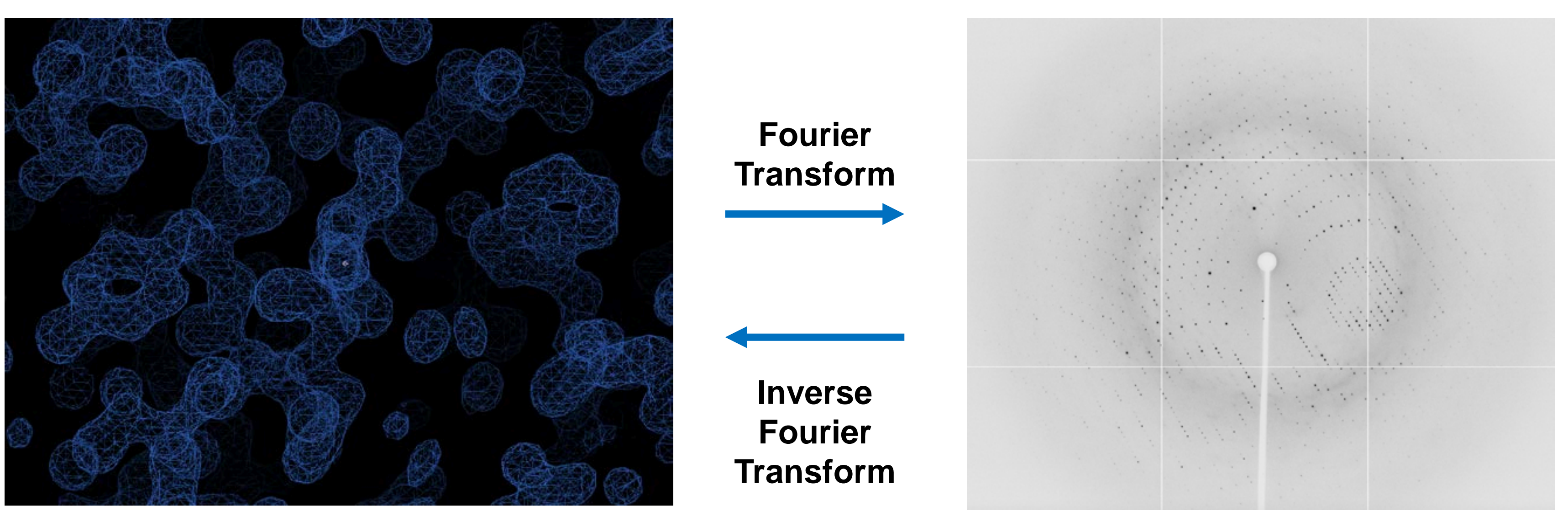
- XDS, Mosflm, HKL2000 and others
  - Determine unit cell dimensions, crystal symmetry, orientation of the crystal, estimation of mosaicity
  - Measure the intensity of each spot (integration)
- Aimless- scale and merge data
  - Scaling- tries to minimise differences between symmetry related and duplicate measurements of a reflection
  - Estimates structure factor amplitude |F|
  - Assess data quality



Spot finding and prediction in mosflm

Diffraction from a protein crystal

- Electron density in the crystal (real space) diffract x-rays to form a diffraction pattern (reciprocal space).
- They are related by a Fourier transform resulting from Bragg diffraction

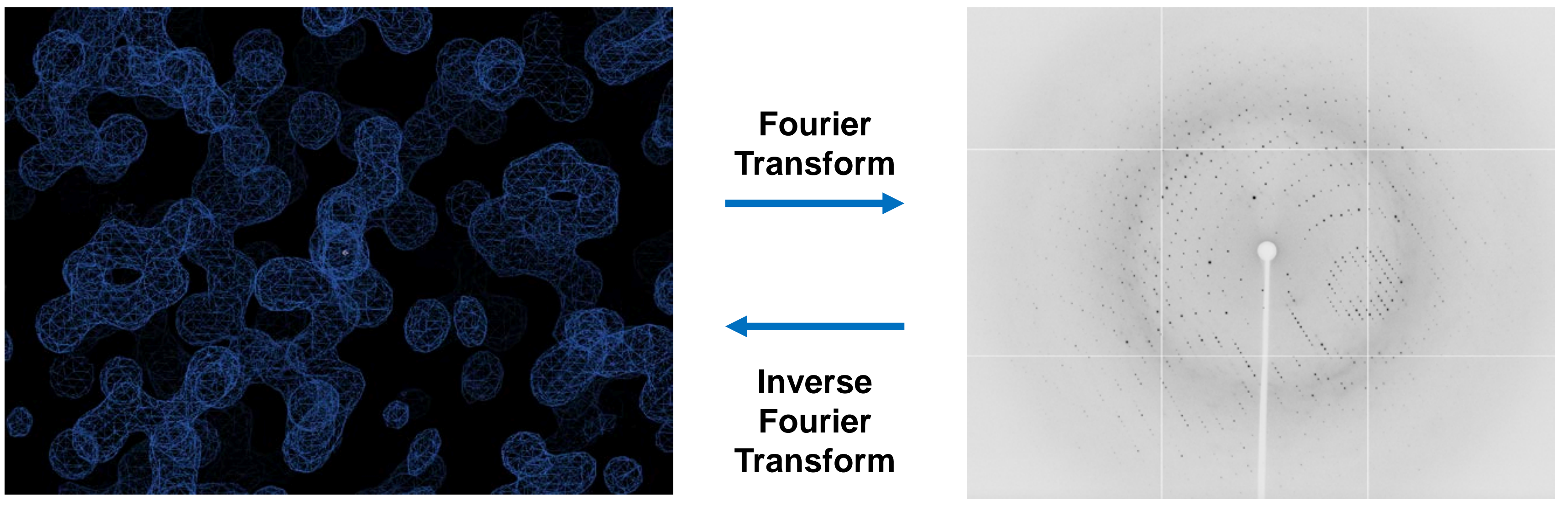


”real space” electron density **”reciprocal space” diffraction pattern**

Phase problem

- What we measure from our diffraction image is the intensity (I) of the reflection, which is related to the structure factor amplitude $|F_{hkl}|$

$$I = |F_{hkl}|^2$$



”real space” electron density

”reciprocal space” diffraction pattern

Phase problem

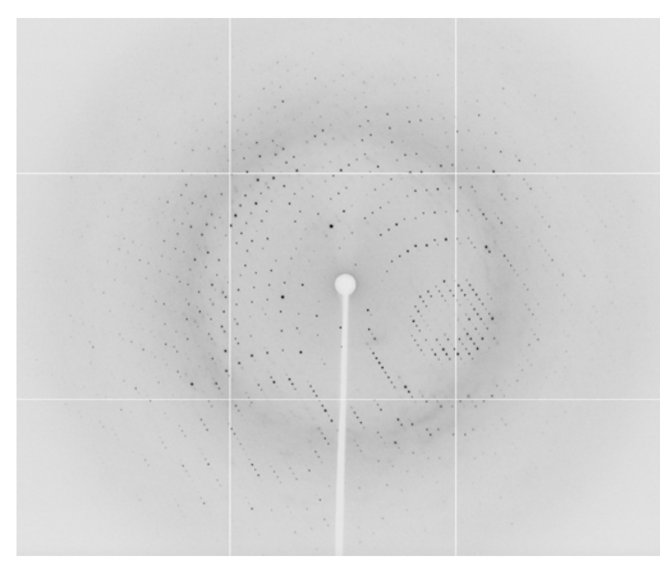
- However, each reflection corresponds to a wave with an amplitude **and a phase (**α_{hkl}**)**, ie. the structure factor $\mathbf{F_{hkl}}$, *which is the Fourier transform of the electron density in the crystal*

$$F_{hkl} = |F_{hkl}|\exp(i\alpha_{hkl})$$

Phases are lost!

Fourier Transform

Inverse Fourier Transform



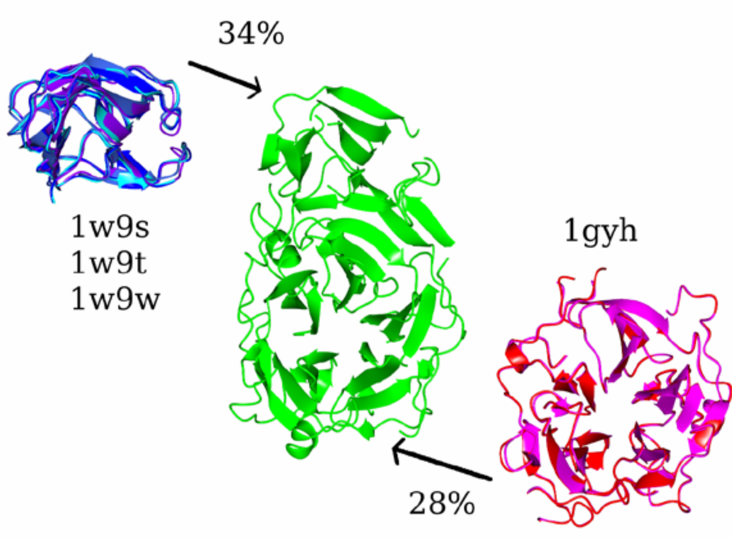
”real space” electron density

”reciprocal space” diffraction pattern

Molecular replacement

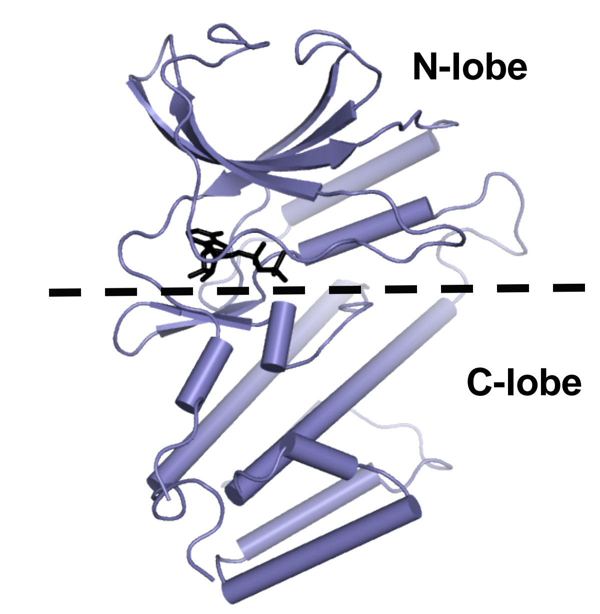
- If known related model available- molecular replacement
  - Provides initial estimates of phases
  - Rotates and translates the provided model within the asymmetric unit of the target crystal
  - Predicts diffraction at a given orientation
  - Position (solution) is found

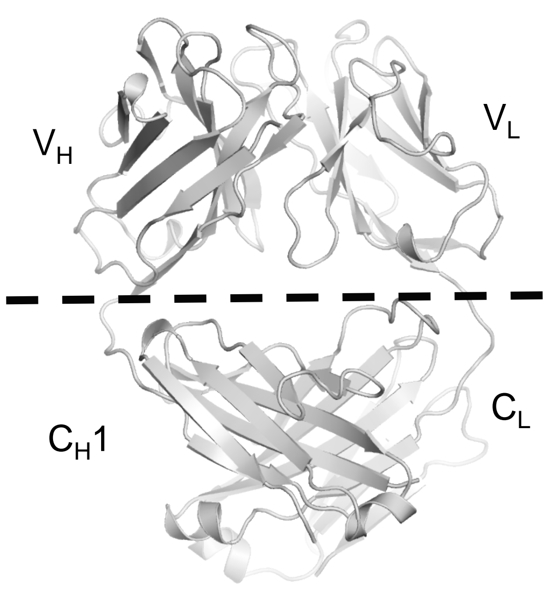
  by the best agreement of measured and calculated structure factors



Molecular Replacement

- Phaser (CCP4, Phenix), Molrep and others
- ~30 % sequence identity
- Remove regions that differ (loops) and truncate side chains (Chainsaw CCP4)
- Divide structure into different domains (hinge)
- Provides an initial model for building and refinement



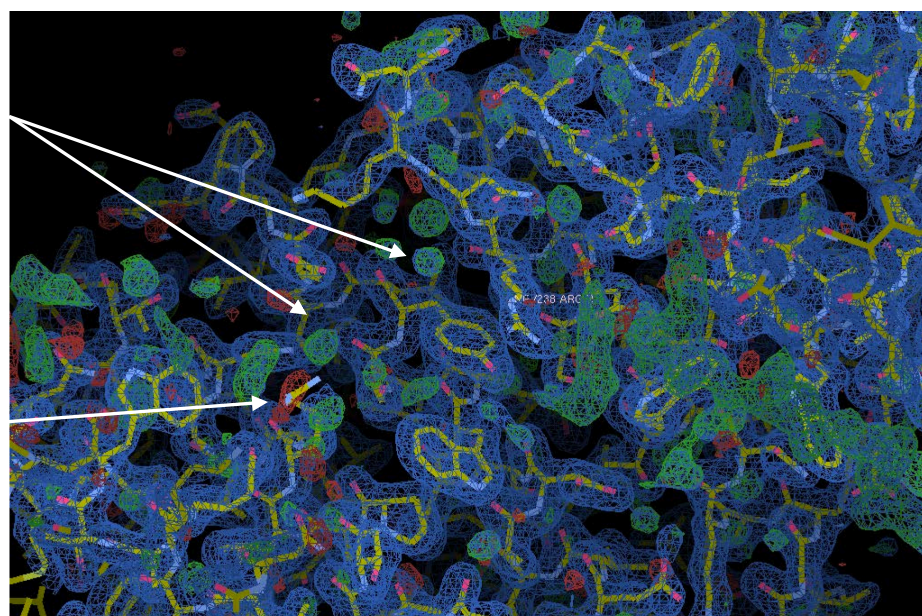


Model building and refinement

- Add loops, side chains, waters, etc. in real space (COOT)
- Refine against data (Refmac, Phenix)
- Difference density map- green indicates atoms missing, red indicates atoms should be removed

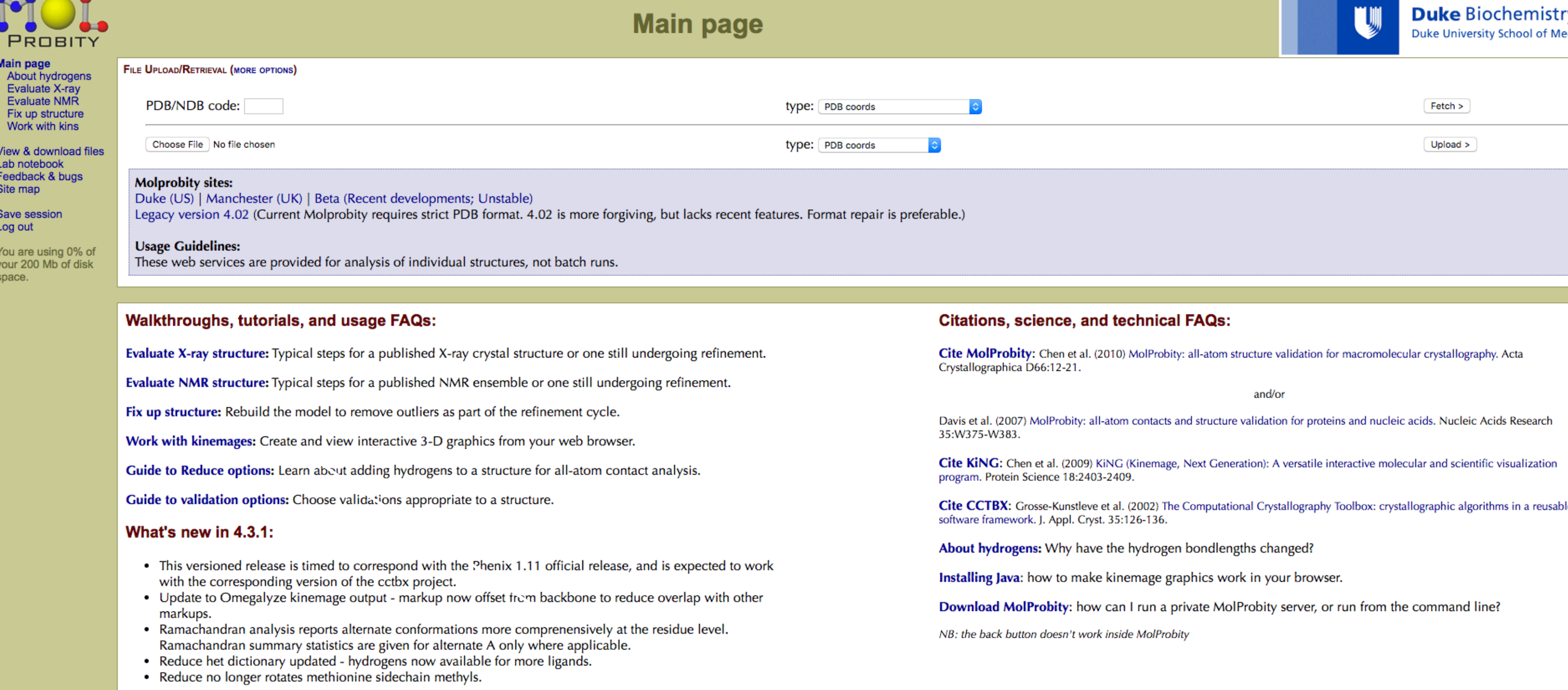
Missing waters

Incorrect side chain rotamer

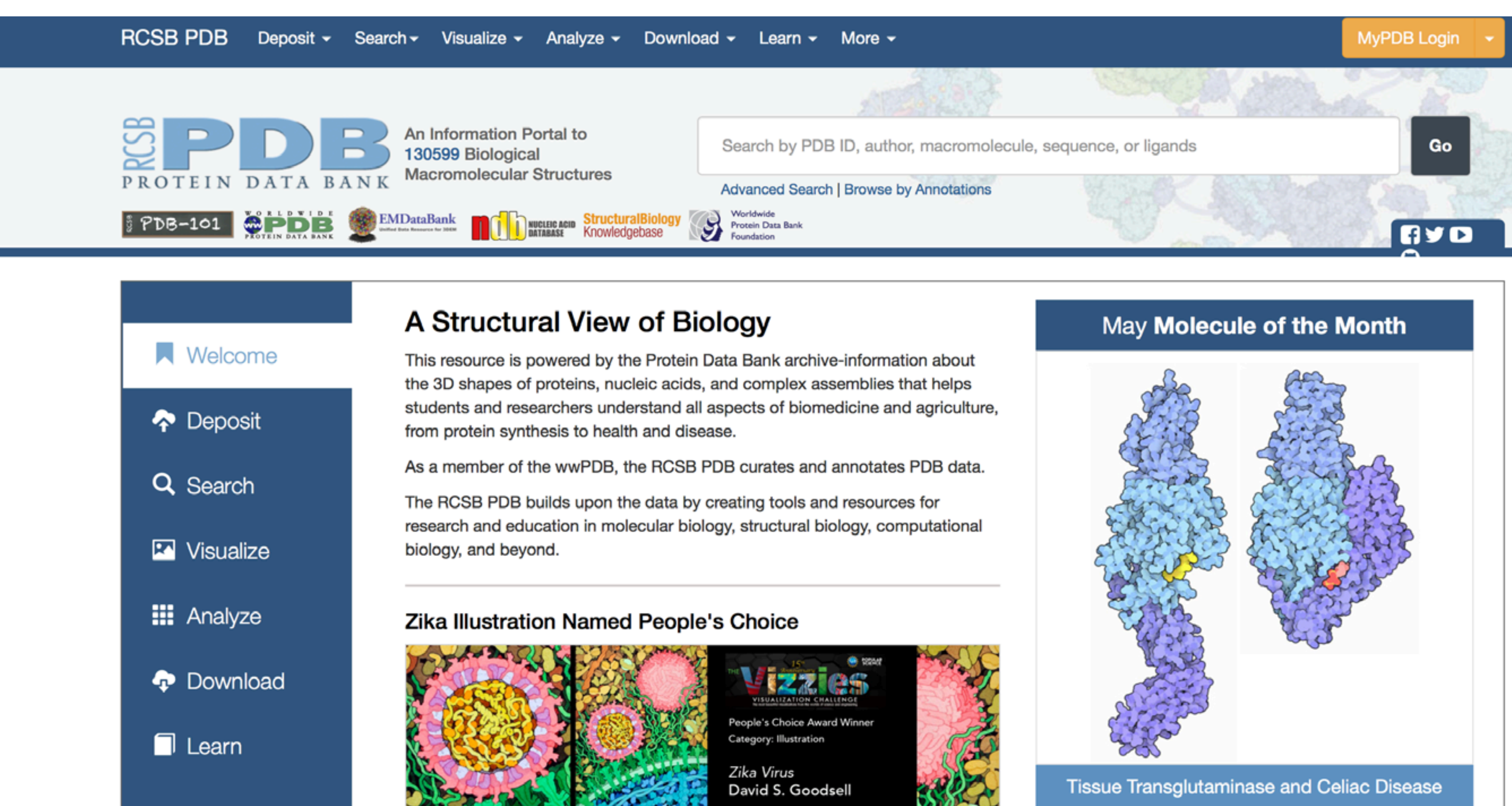


Peptide needs to be built

Validation- MolProbity server

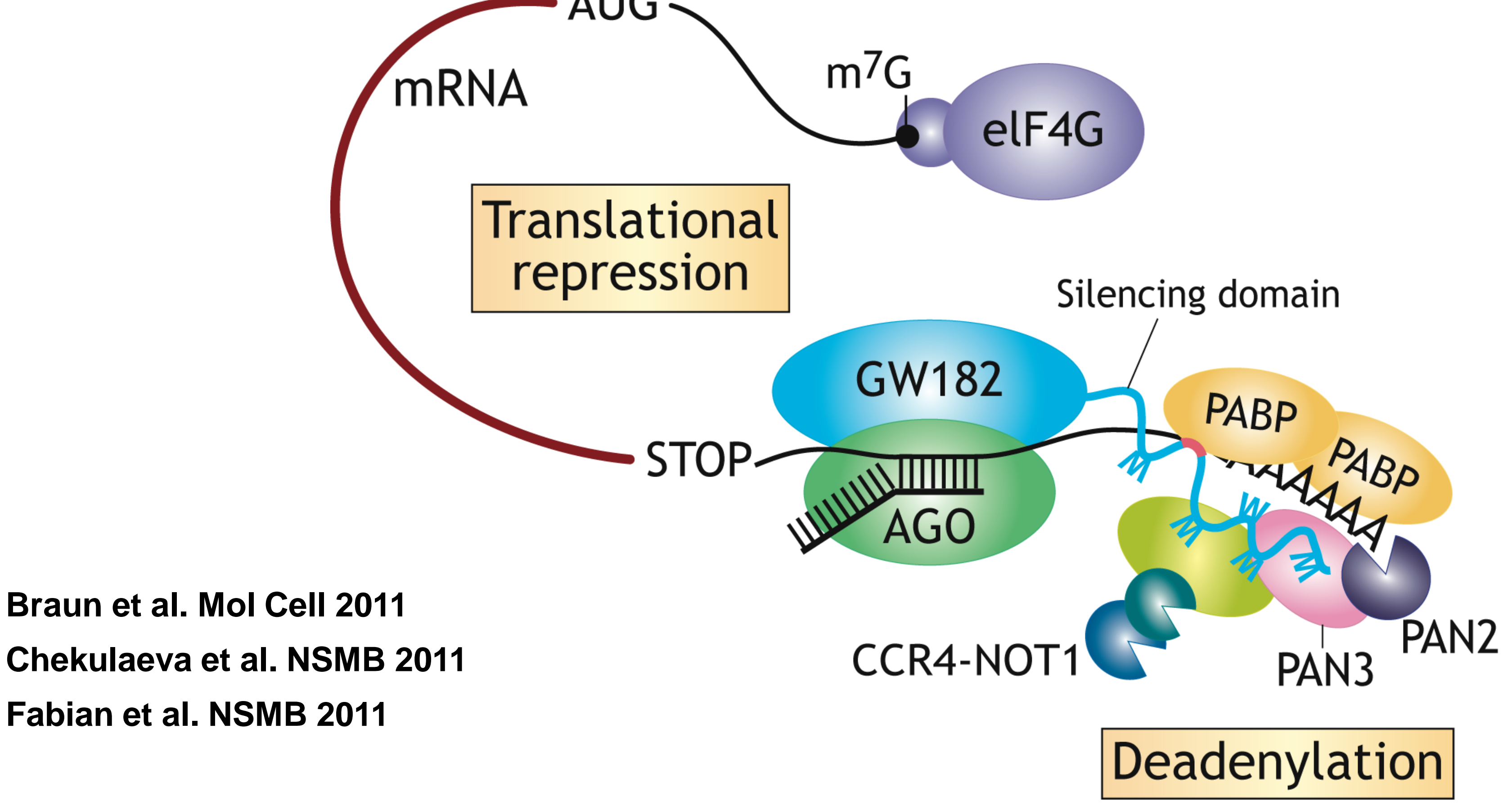


Deposition- PDB



Protein structure-function relationships

- The PAN2-PAN3 deadenylase complex is directly recruited to miRNA targets



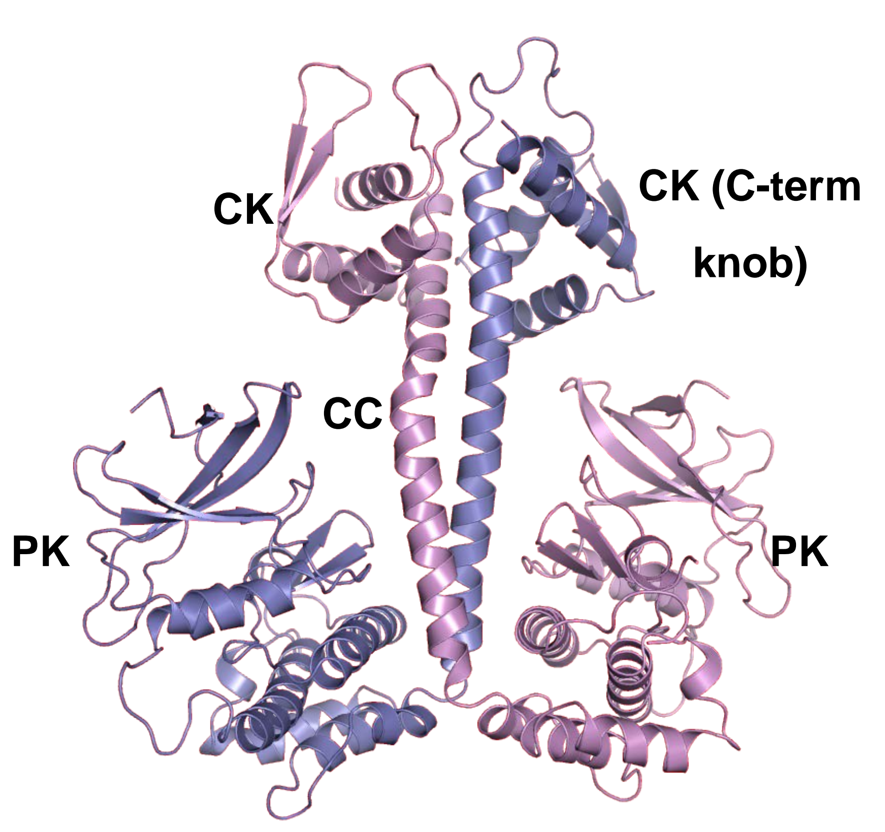
Braun et al. Mol Cell 2011

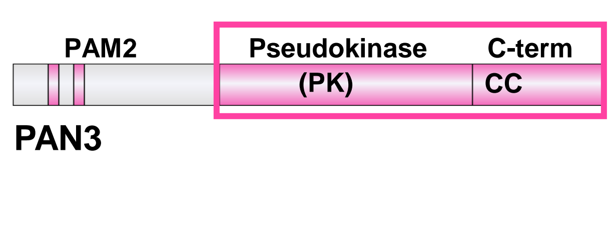
Chekulaeva et al. NSMB 2011

Fabian et al. NSMB 2011

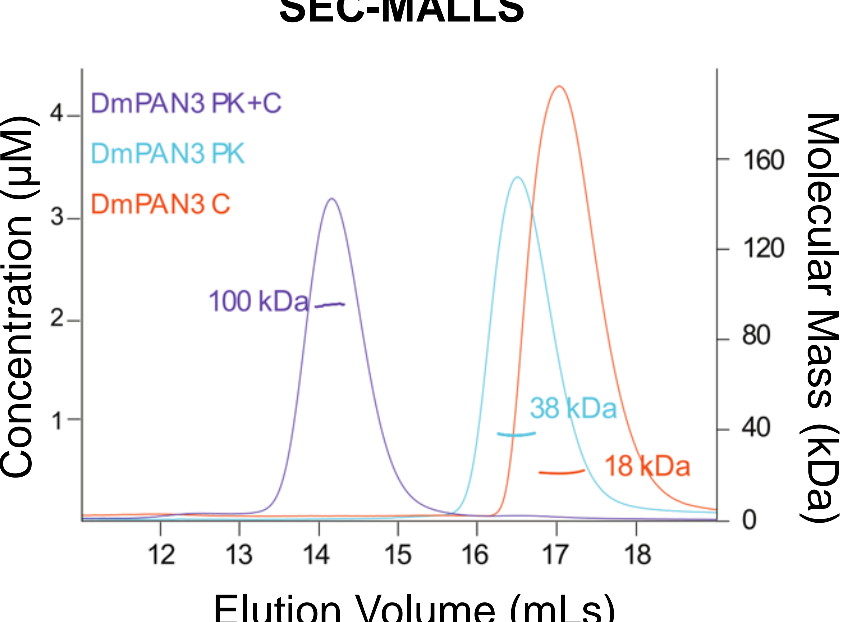
Structure of the PAN3 pseudokinase

- PAN3 forms dimers in crystal lattice and in solution

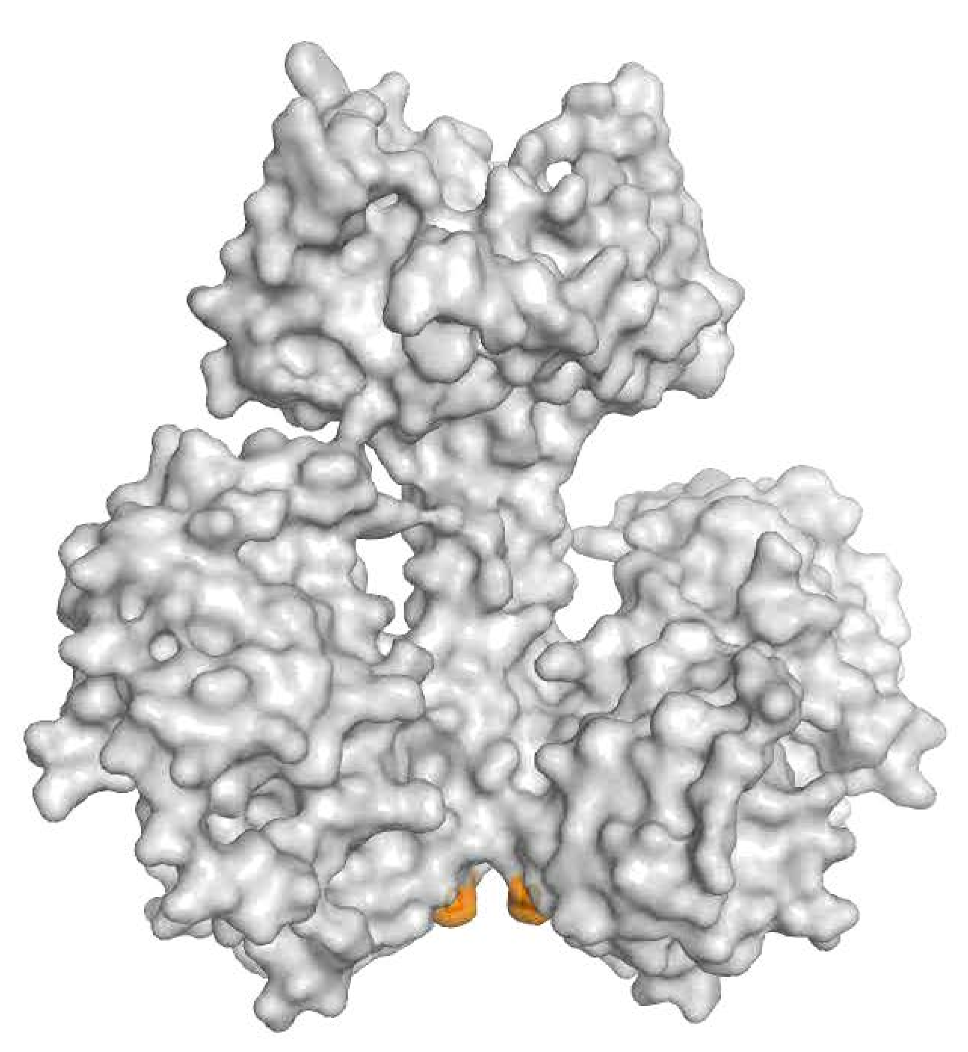




SEC-MALLS

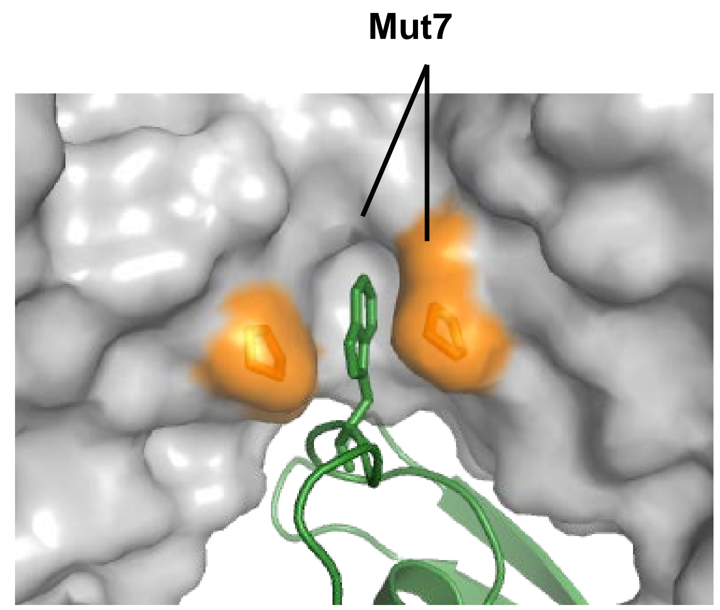


PAN3 dimers harbour a Trp binding pocket

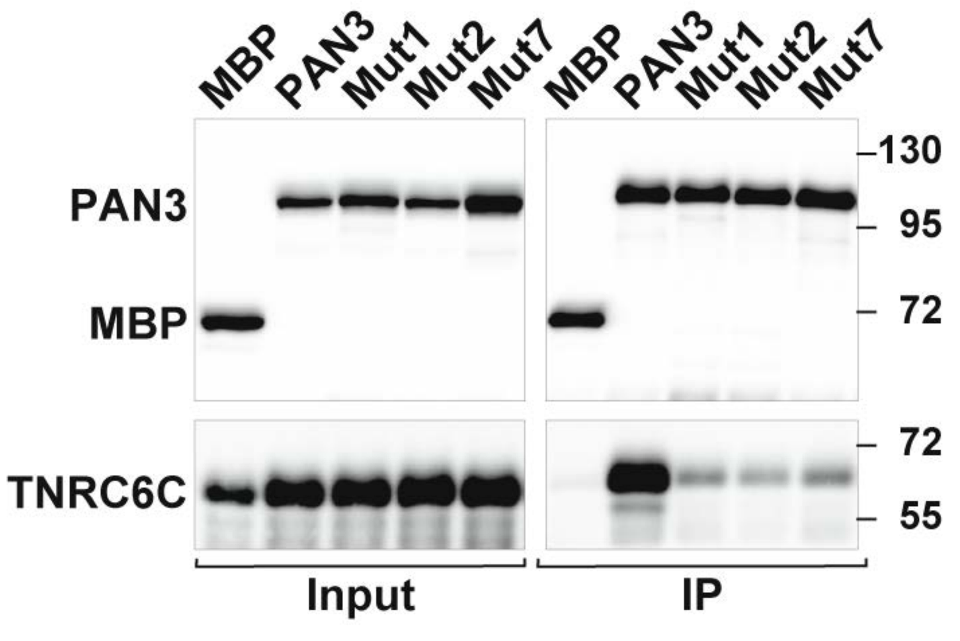


Functional validation- PAN3 example

- Identification of a GW182 binding pocket at the PAN3 dimerisation interface



Dm PAN3 symmetry mate

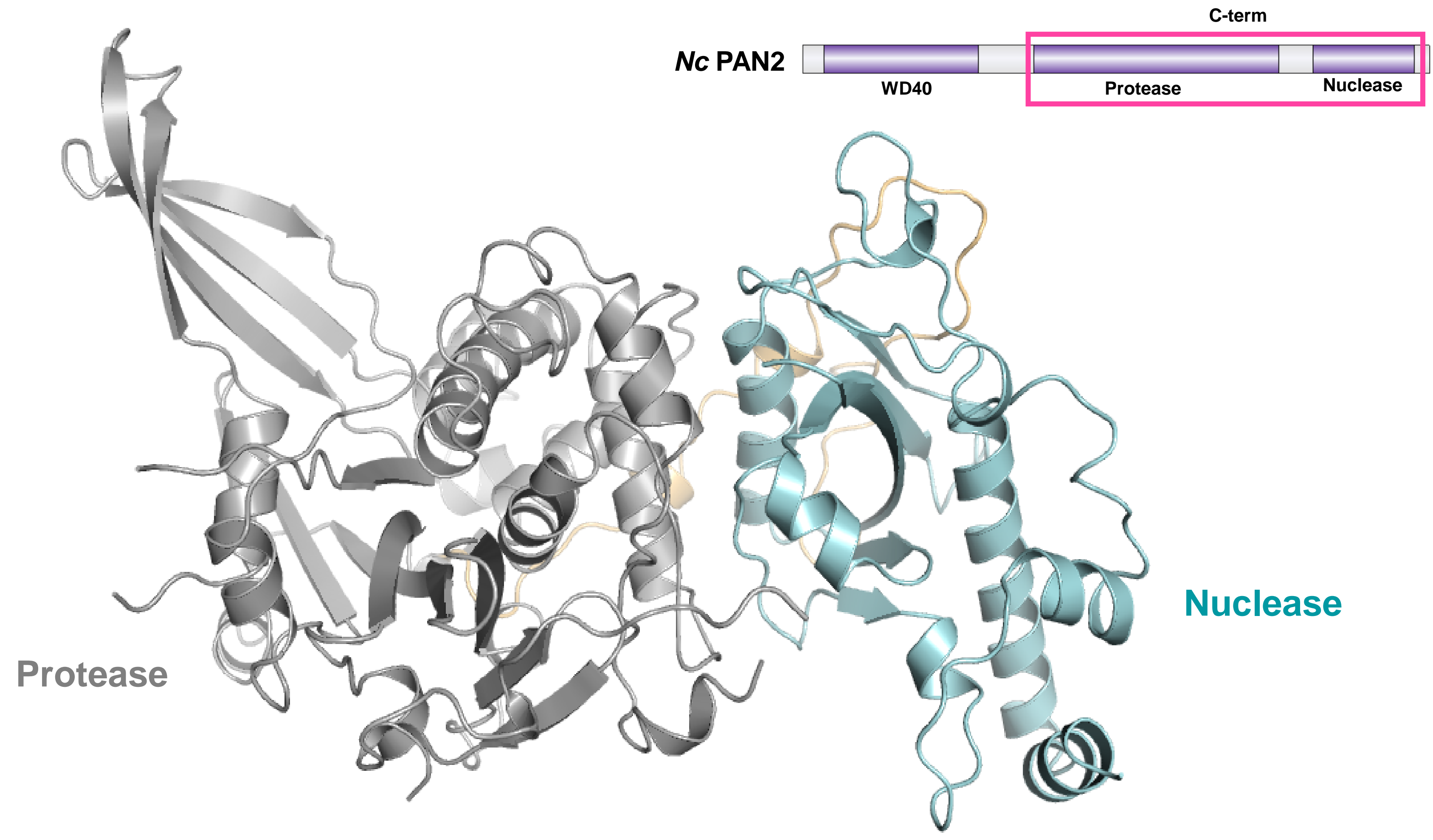


Mut7- W pocket mutant

Mut1, Mut2- dimerization mutant

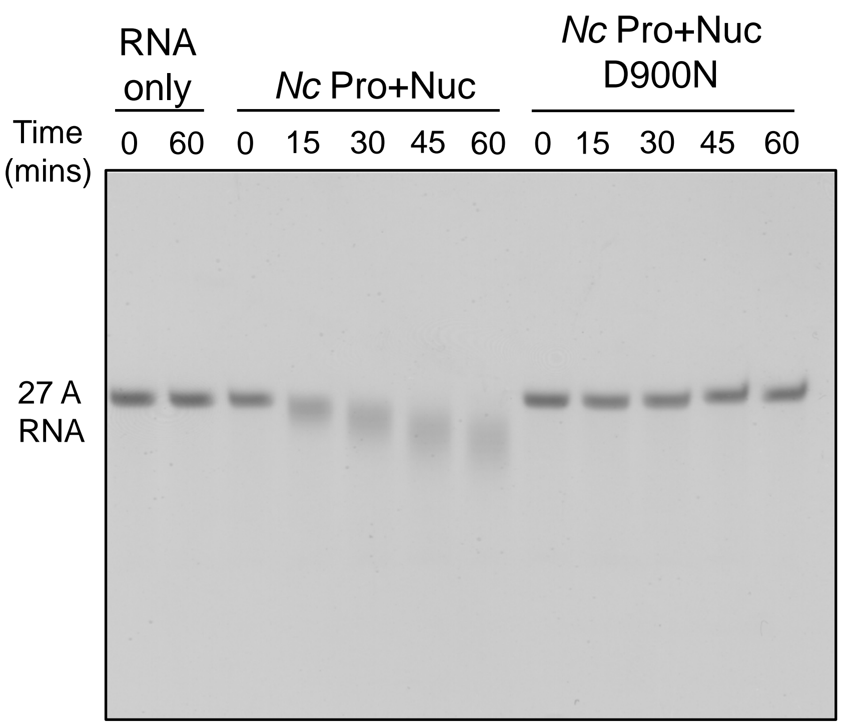
Structure of PAN2 protease-nuclease

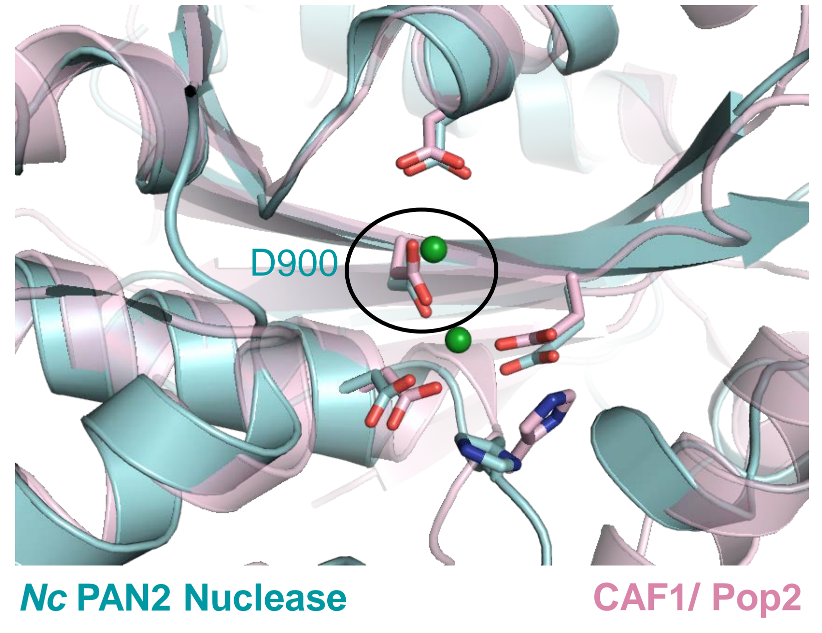
- PAN2 is the catalytic subunit of the deadenylase complex



Functional validation- PAN2 example

- PAN2 is a polyA RNA nuclease

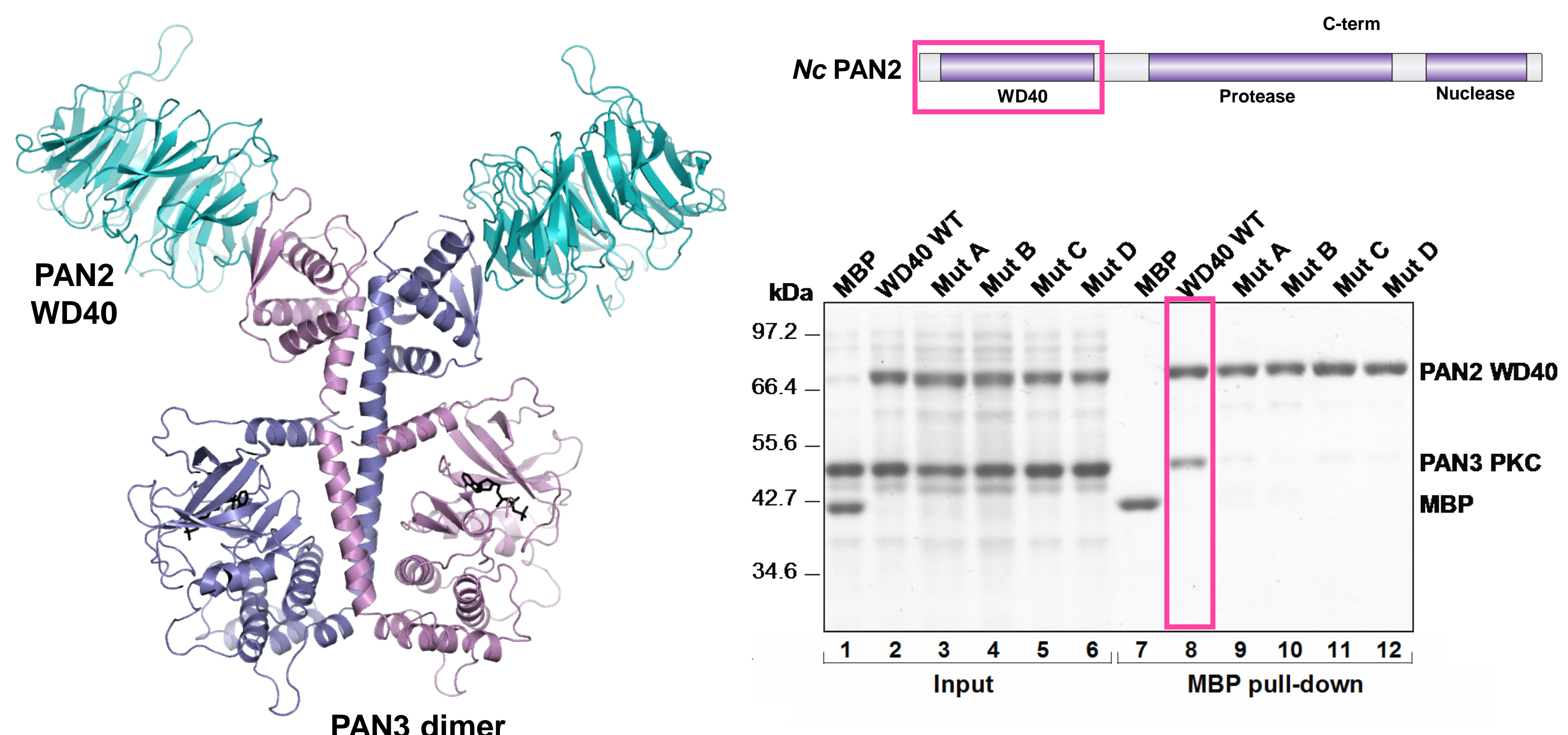




***Nc* PAN2 Nuclease** **CAF1/ Pop2**

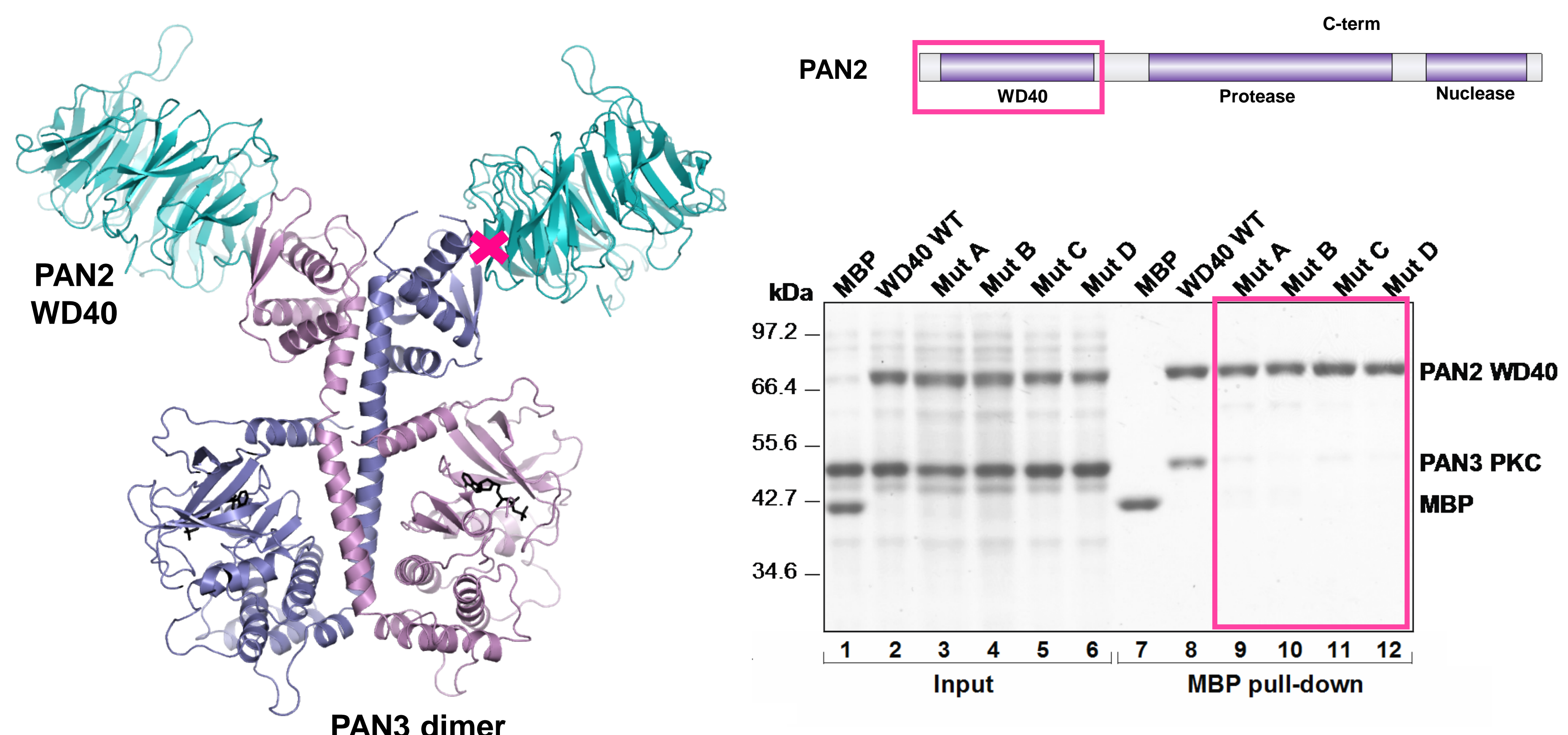
Functional validation- PAN2-PAN3 complex

- The PAN2 WD40 domain interacts with PAN3



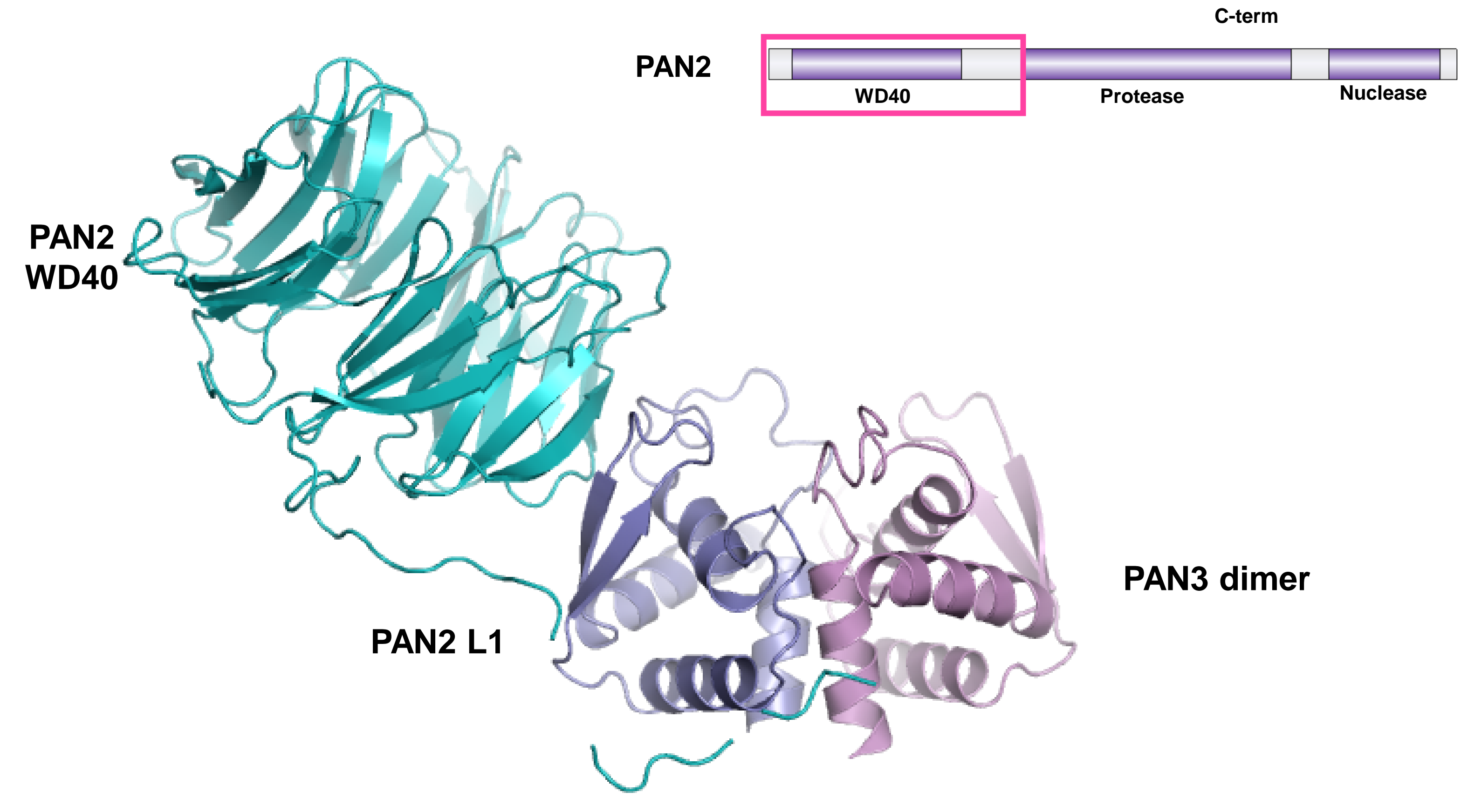
Functional validation- PAN2-PAN3 complex

- The PAN2 WD40 domain interacts with PAN3



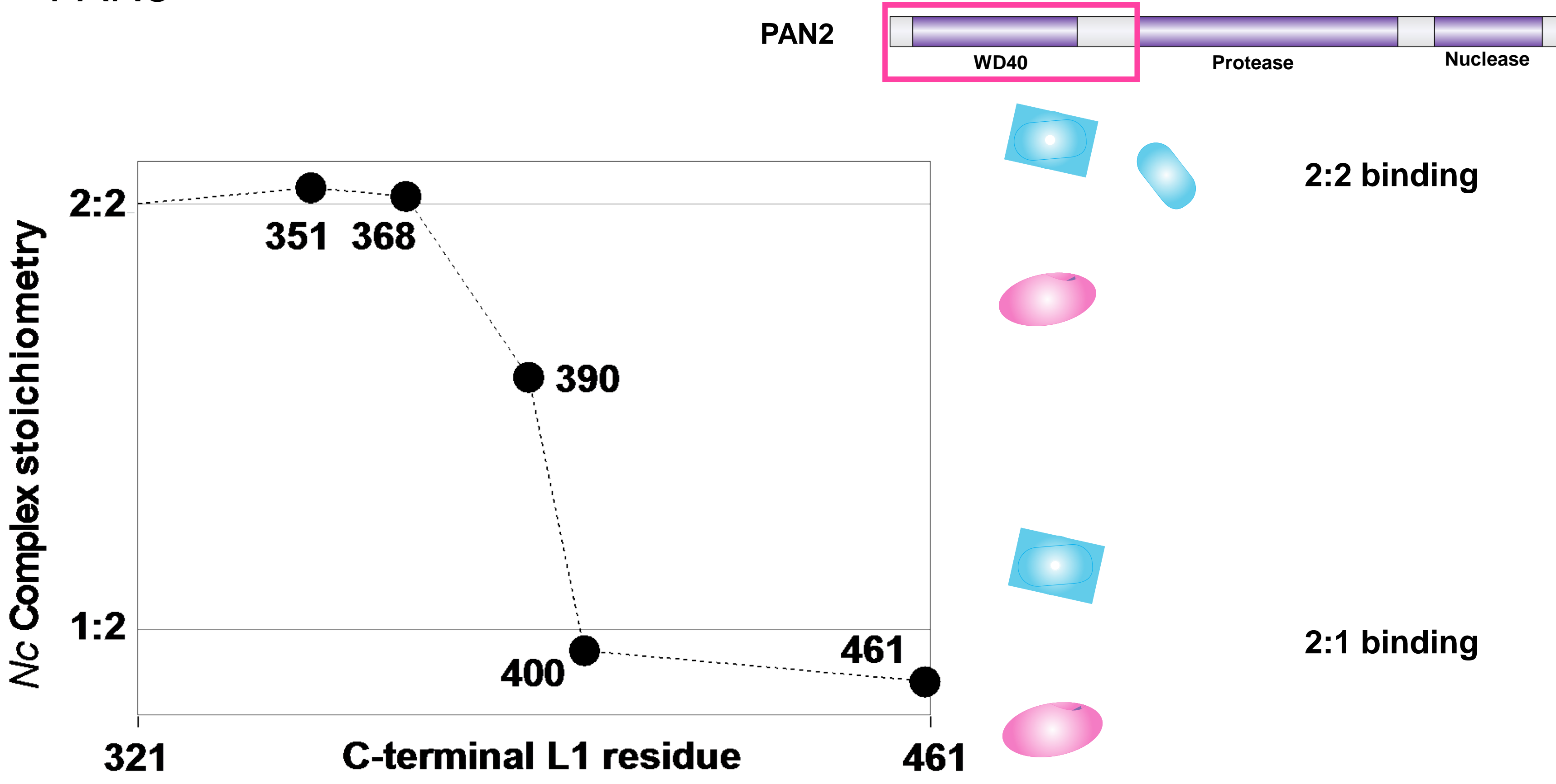
Structure of the PAN2-PAN3 complex

- The PAN2 L1 region also interacts with PAN3

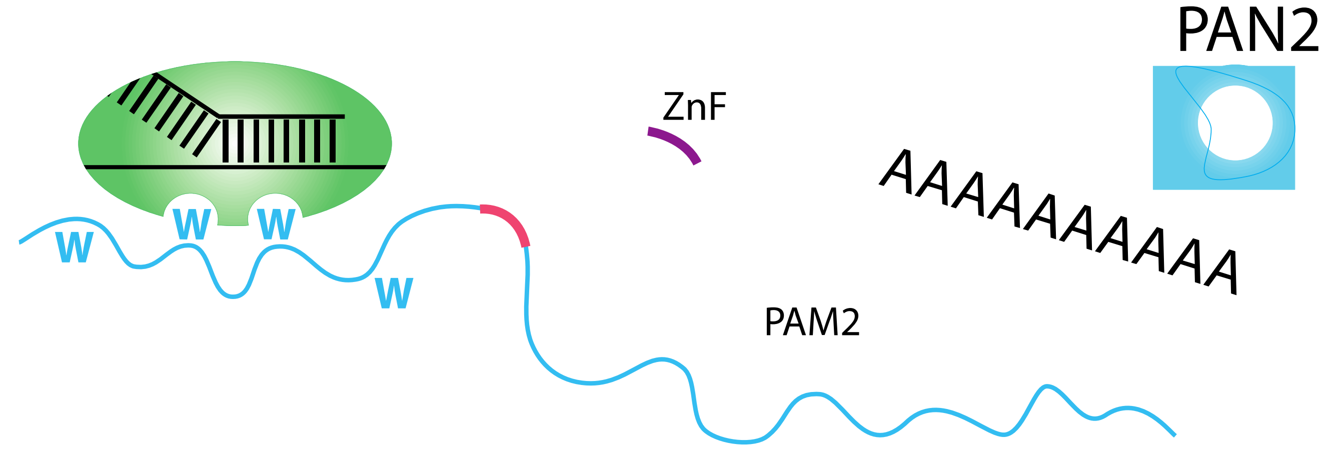


Functional validation- PAN2-PAN3 complex

- The L1 region of PAN2 determines the stoichiometry of binding to PAN3

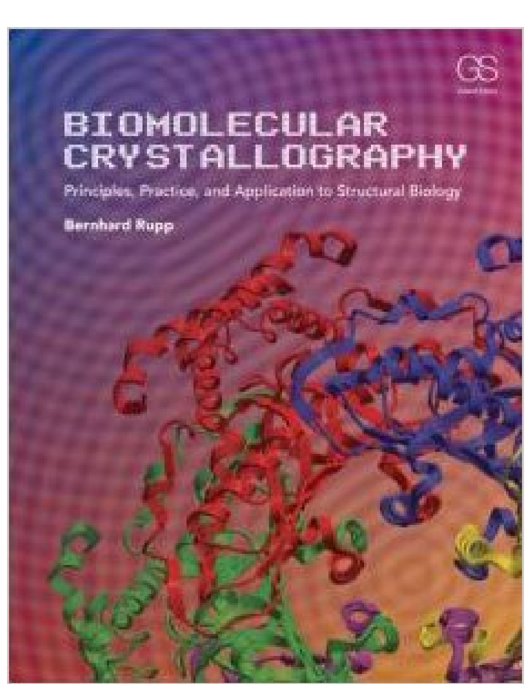


Model of PAN2-PAN3 recruitment to miRNA targets



Useful links

- Bernhard Rupp’s webpage:
  http://www.ruppweb.org/Xray/101index.html
- Bernhard Rupp’s book: Biomolecular Crystallography
- Interactive e-Crystallography course:
  http://escher.epfl.ch/eCrystallography/
- CCP4bb, phenixbb



Acknowledgements

- Daniela Stock
- Alastair Stewart

- Elisa Izaurralde
- Oliver Weichenrieder
- Andreas Boland
- Eric Huntzinger
- Stefanie Jonas
- Daniel Peter