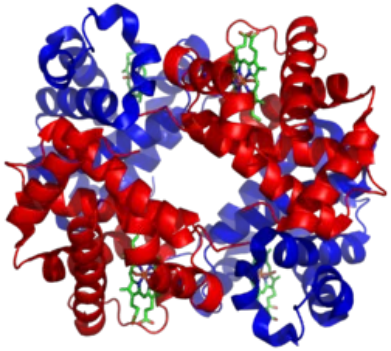
What is structuralbiology?

Structural biology is the study of the molecular structure of biological macromolecules, particularly proteins and nucleic acids, and how alterations in their structures affect their function.

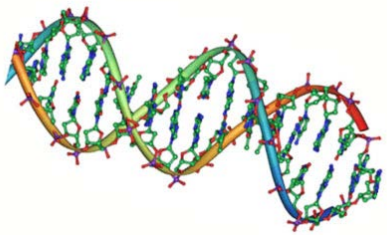
Examples of Nobel laureates applying structural biology:

The structure of hemoglobin



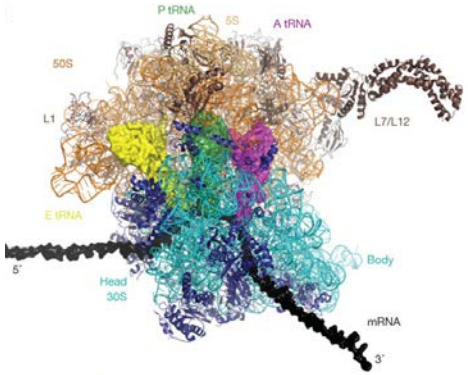
1962-Chemistry
Max F. Perutz
John C. Kendrew

The double helix of DNA



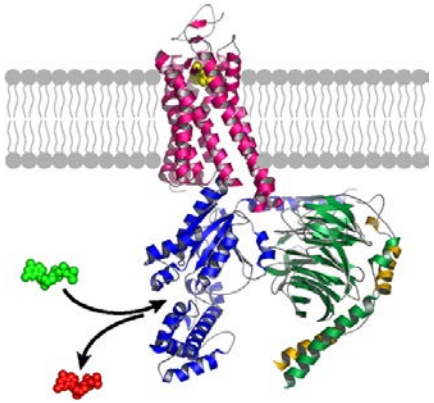
1962-Medicine
Francis HC Crick
James D Watson
Maurice HF Wilkins

The structure of the ribosome



2009-Chemistry
Venkatraman Ramakrishnan
Thomas A. Steitz
Ada E. Yonath

The structure of G protein coupled receptors

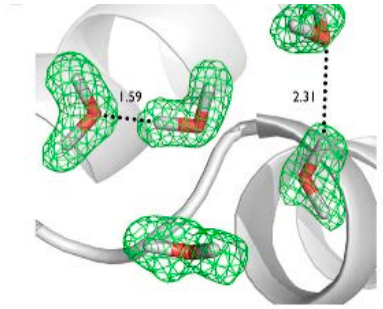


2012- Chemistry
Brian Kobilka
Robert Lefkowitz

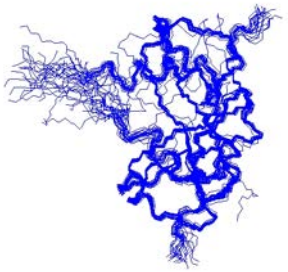
How do the techniques complement each other?

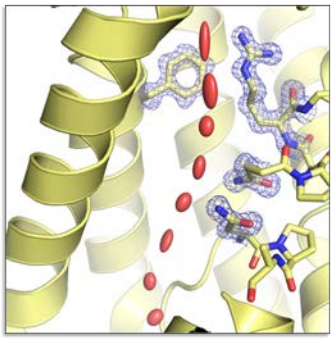
Very high resolution — High resolution — moderate resolution

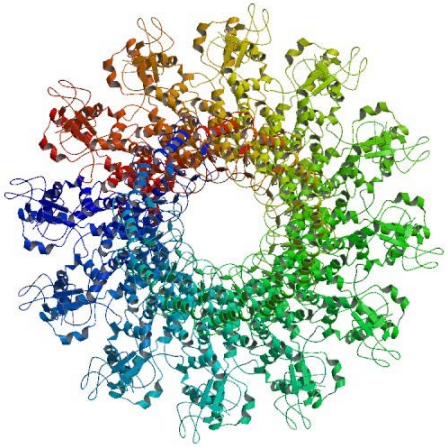
Atoms/Protons — **Hydrogen bonds/interactions** — **Large proteins/complexes**



From Zoe Fischer

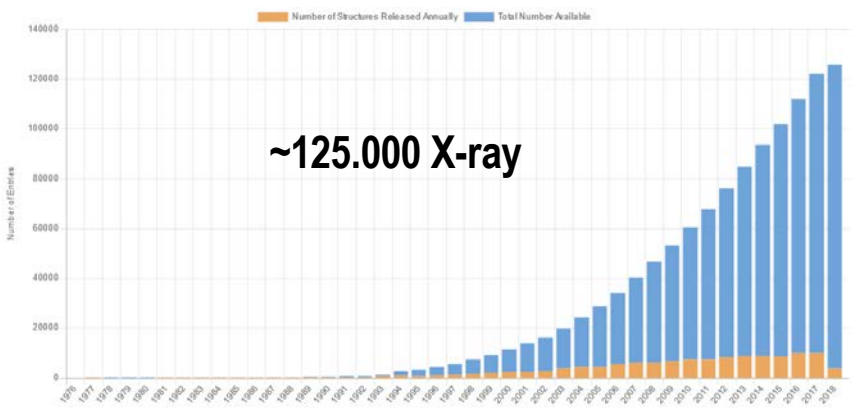


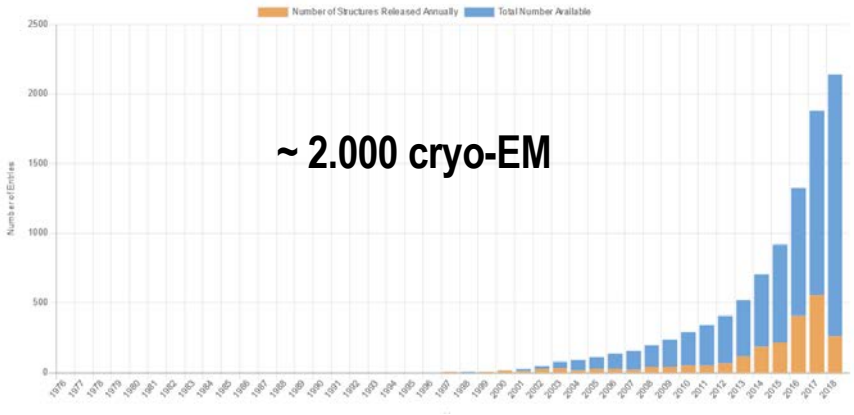


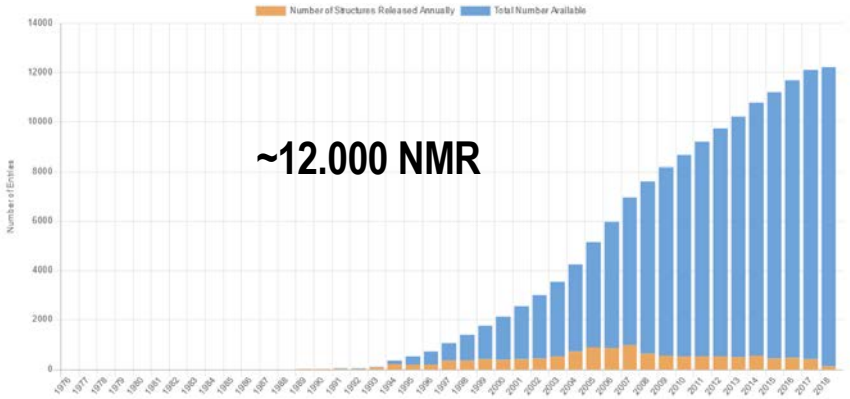


Neutroncrystallography — NMR — X-ray — Cryo-EM

X-ray crystallography (X-ray), NMR, cryo-electron microscopy (cryo-EM) and neutron crystallography structures in Protein Data Bank (PDB)







Approx. 100 structures determined by neutron crystallography

Why X-ray crystallography ?

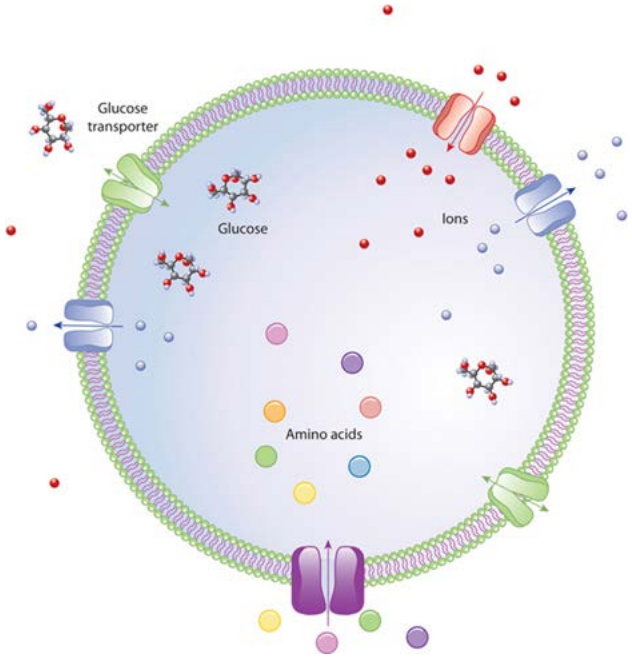
- Crystallography can provide the answer to many structure related questions, from global folds to atomic details of bonding. In contrast to NMR and EM, less size limitation exists for the molecule or complex to be studied. NMR is to prefer for small peptides/proteins and cryo-EM for larger complexes, typically >150 kDa.

- The price for the high accuracy of crystallographic structures is that a good crystal must be found, and that limited information about the molecule's dynamic behavior in solution is available from one single diffraction experiment.

- However, in the core regions of the molecules, X-ray and NMR structures agree very well, and enzymes maintain their activity even in crystals.

- Molecular structures is a prerequisite for rational drug design and for structure based functional studies to aid the development of effective therapeutic agents and drugs.

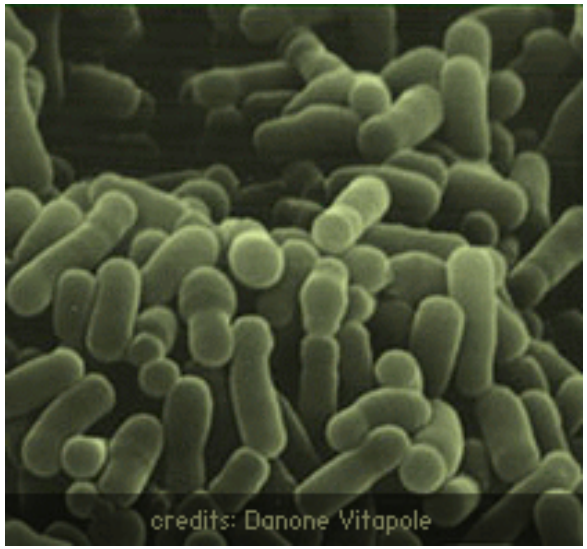
Aquaporins



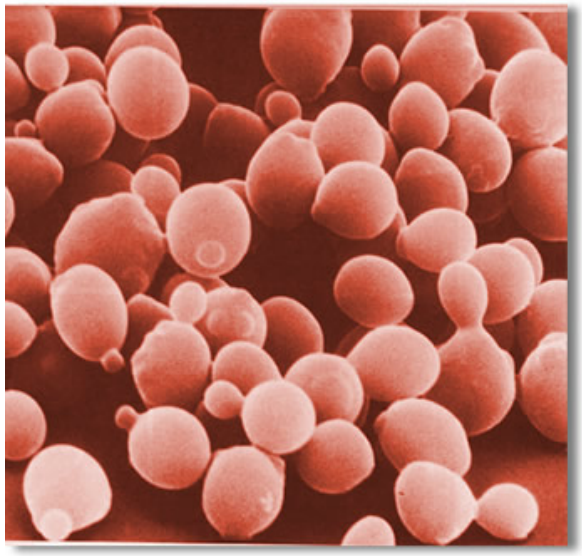
- Peter Agre, Nobel Prize 2003
- Membrane channels selective towards water (aquaporins) and small solutes such as glycerol (aquaglyceroporins)
- Ion- and proton-exclusion
- Common fold known by structures from different organisms
- Elongated N-terminus common in yeast, but of unknown function
- Aquaporins important for freeze-tolerance in yeast

Protein factories

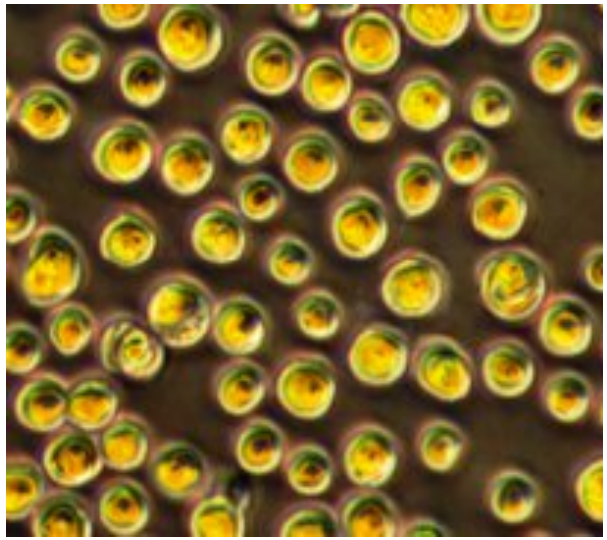
Bacteria (*E. coli*)



Yeast (*S. cerevisiae* or *P. pastoris*)



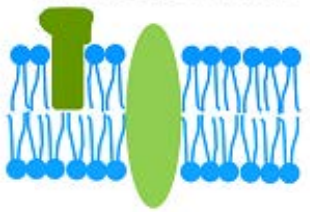
Insect cells (SF9s or Hi5)



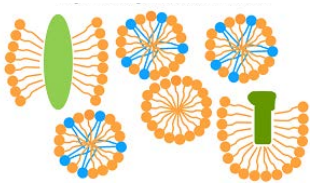
Protein expression, solubilisation and purification

Production

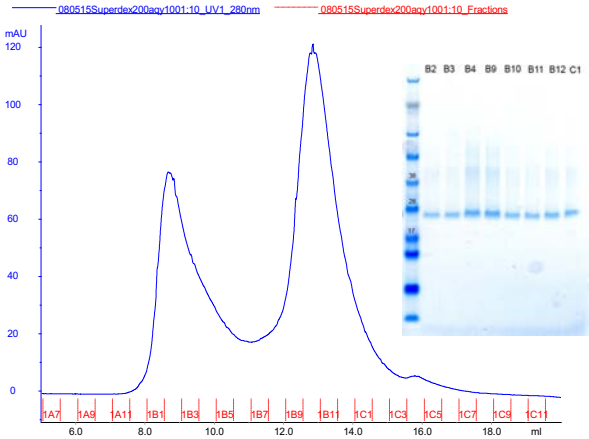
Native membrane



Detergent



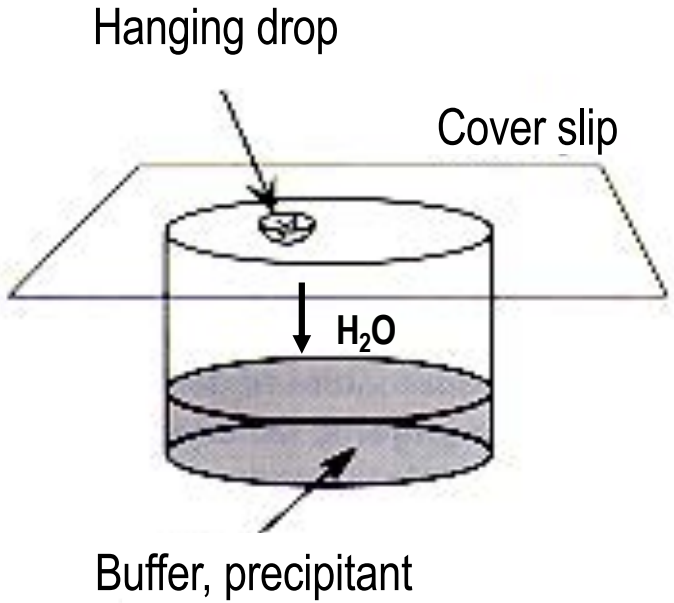
Size exclusion chromatography



MOLECULAR BIOLOGY → PROTEN EXPRESSION → SOLUBILISATION/ PURIFICATION → CRYSTALLISATION → STRUCTURE

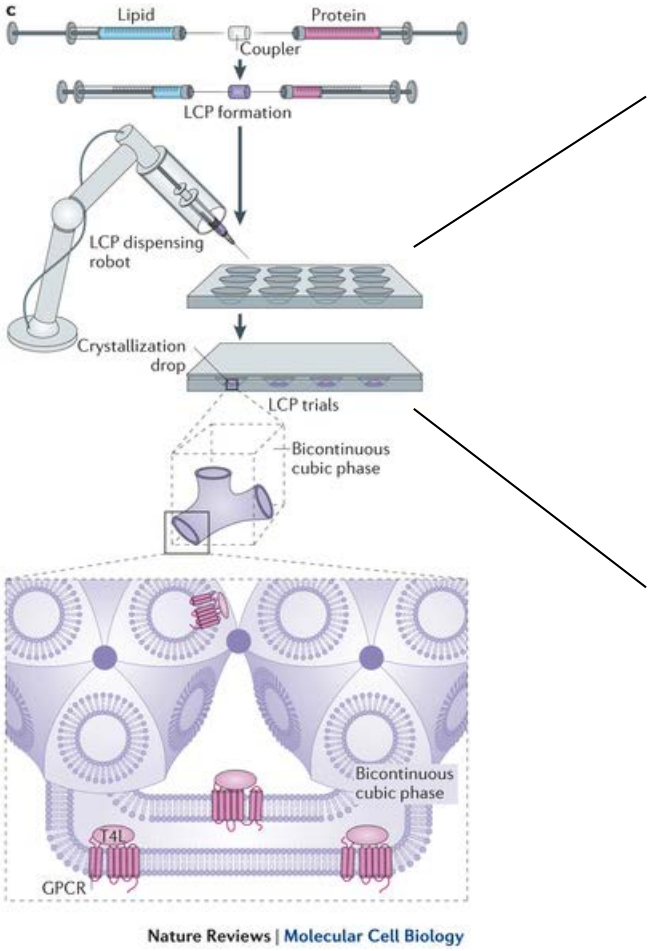
Vapour diffusion crystallisation

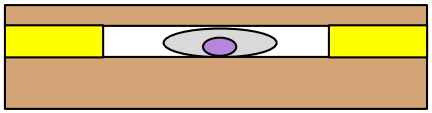
Growing protein crystals using vapour diffusion

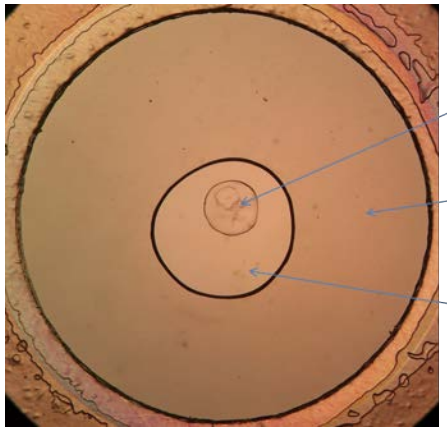


The drop contains 50% protein solution and 50% buffer
water will diffuse to the bottom solution and crystals will be formed in the drop.

Liquid cubic phase crystallisation (LCP)

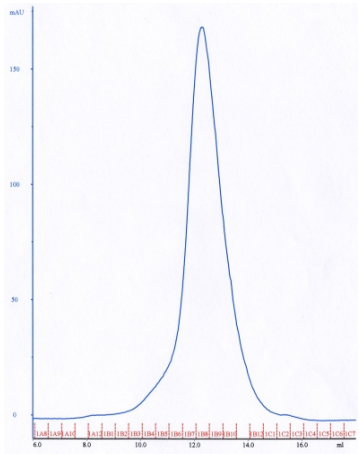


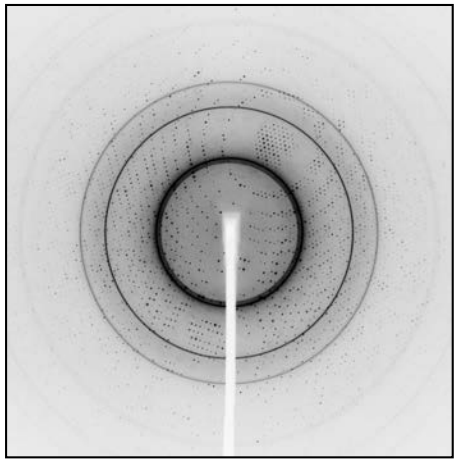


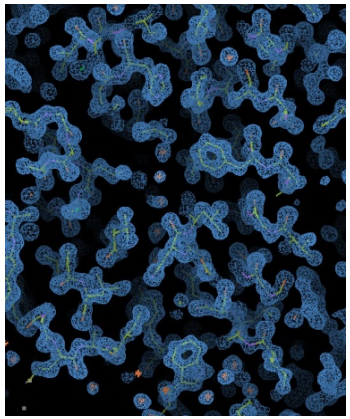


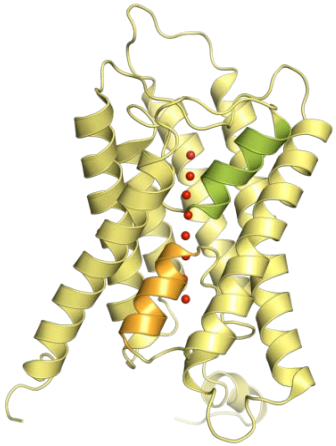
MOLECULAR BIOLOGY → PROTEN EXPRESSION → PURIFICATION → **CRYSTALLISATION** → STRUCTURE

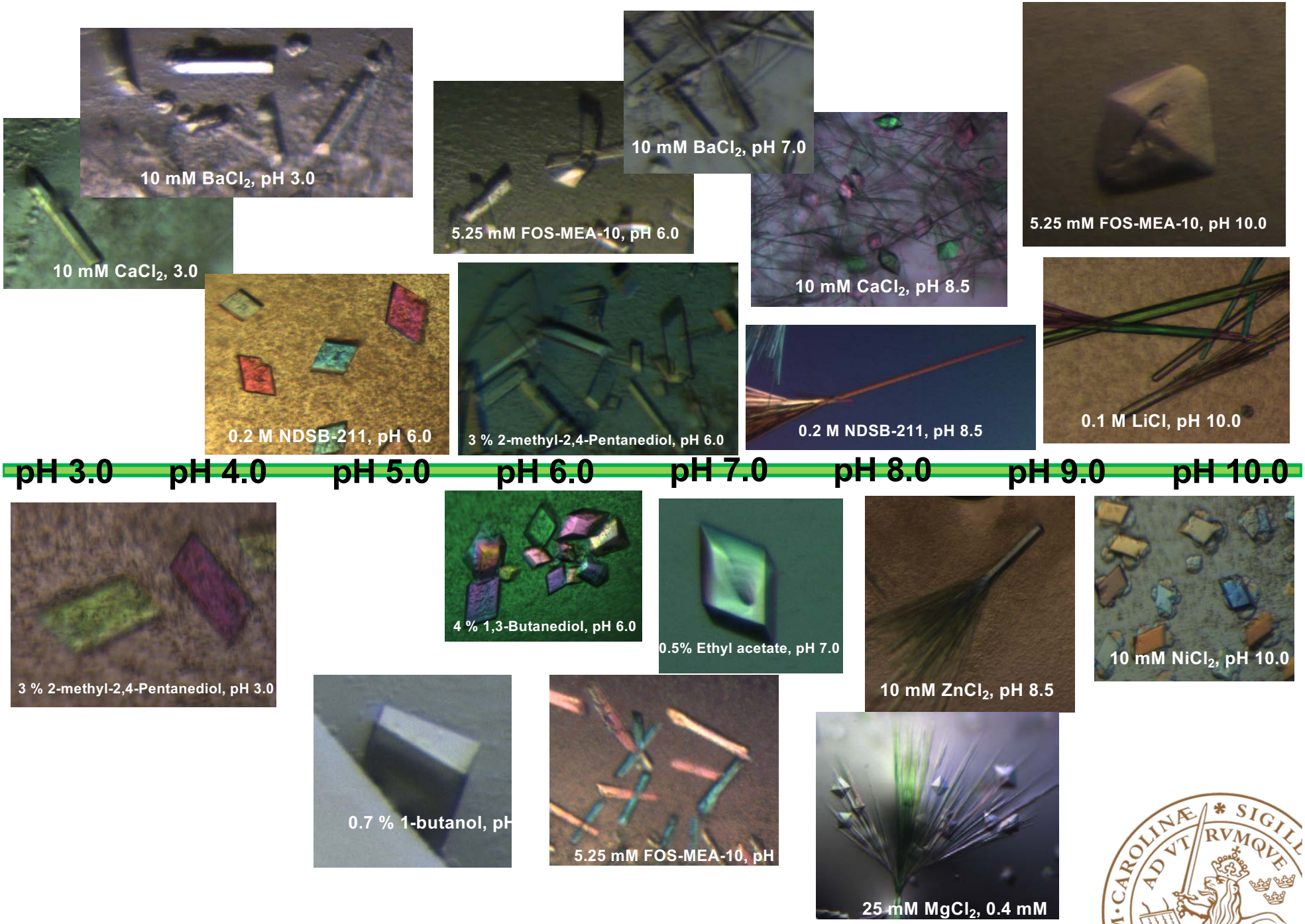
The process in short…





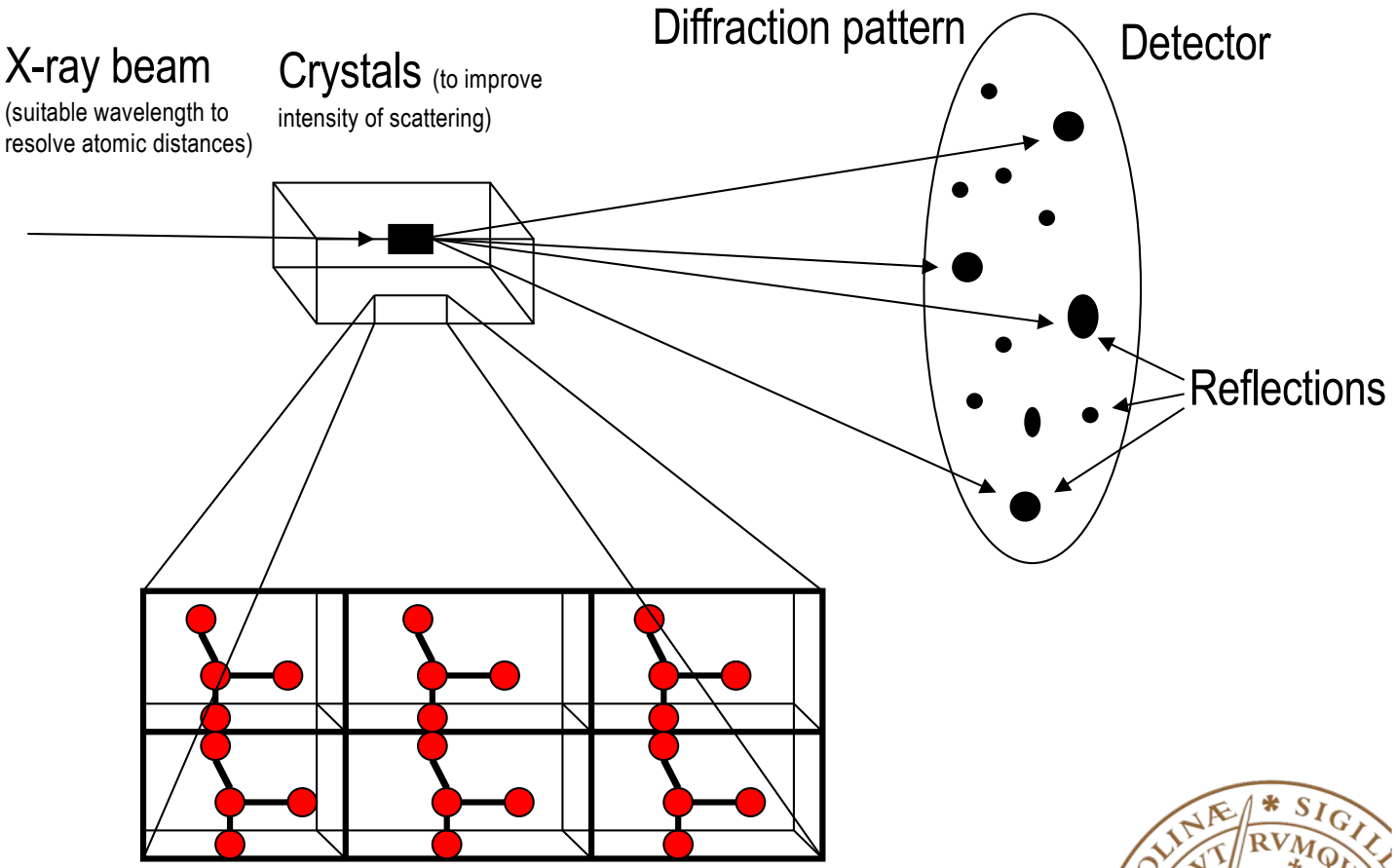




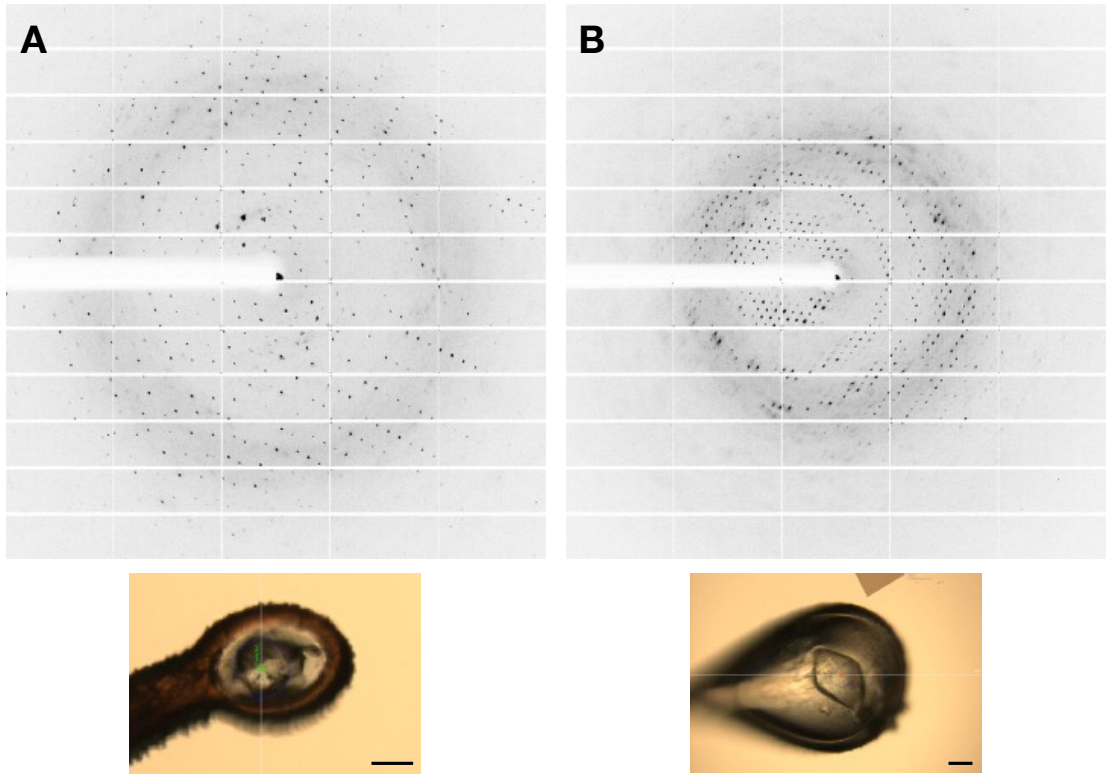

10 mM $BaCl_2$, pH 3.0
10 mM $CaCl_2$, 3.0
5.25 mM FOS-MEA-10, pH 6.0
10 mM $BaCl_2$, pH 7.0
10 mM $CaCl_2$, pH 8.5
5.25 mM FOS-MEA-10, pH 10.0
0.2 M NDSB-211, pH 6.0
3 % 2-methyl-2,4-Pentanediol, pH 6.0
0.2 M NDSB-211, pH 8.5
0.1 M LiCl, pH 10.0
pH 3.0
pH 4.0
pH 5.0
pH 6.0
pH 7.0
pH 8.0
pH 9.0
pH 10.0
3 % 2-methyl-2,4-Pentanediol, pH 3.0
4 % 1,3-Butanediol, pH 6.0
0.5% Ethyl acetate, pH 7.0
10 mM $ZnCl_2$, pH 8.5
10 mM $NiCl_2$, pH 10.0
0.7 % 1-butanol, pH
5.25 mM FOS-MEA-10, pH
25 mM $MgCl_2$, 0.4 mM

The crystals are fished by using loops connected to a top-hat, and froozen in liquid nitrogen.

X-ray crystallography



Glycerol soaking

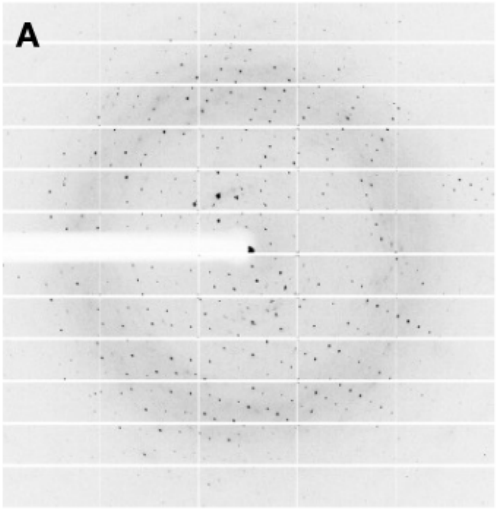


5-20 % glycerol

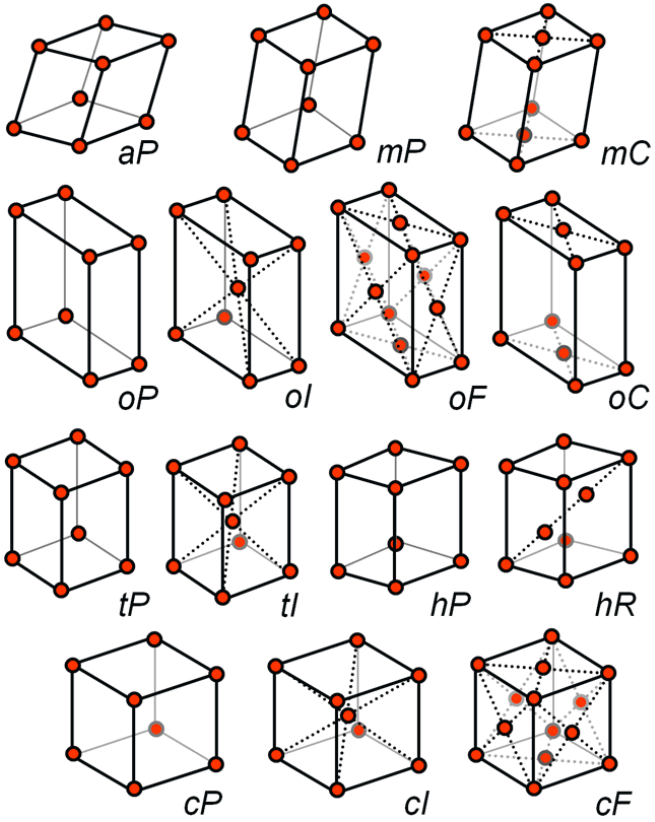
5 % glycerol

Flow scheme for structure determination

Collect data



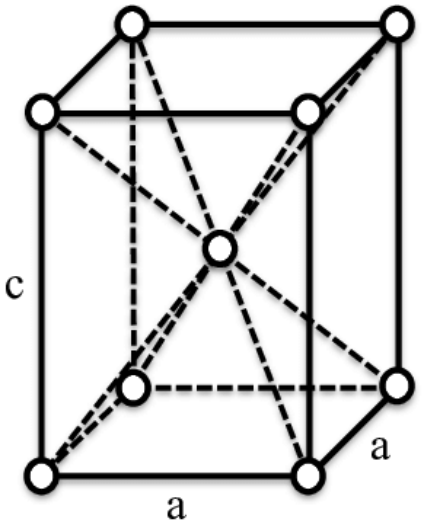
Determine the space group



XDS output- to decide on the space group

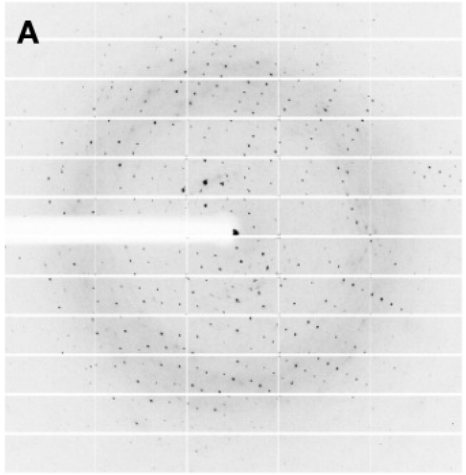
	LATTICE-CHARACTER	BRAVAIS-LATTICE	QUALITY OF FIT	UNIT CELL CONSTANTS (ANGSTROEM & DEGREES) a	b	c	alpha	beta	gamma	REINDEXING TRANSFORMATION											
*	44	aP	0.0	73.6	73.6	73.6	111.2	108.6	108.6	-1	-1	-1	0	1	-1	1	0	-1	1	1	0
*	25	mC	0.0	83.1	121.5	73.6	90.0	124.4	90.0	0	0	2	0	-2	2	0	0	-1	-1	-1	0
*	14	mC	0.0	83.1	121.5	73.6	90.0	124.4	90.0	0	0	-2	0	-2	2	0	0	1	1	1	0
*	17	mC	0.0	119.5	85.9	83.1	90.0	134.1	90.0	0	-2	-2	0	2	0	0	0	0	0	2	0
*	43	mI	0.0	83.1	121.5	73.6	90.0	124.4	90.0	0	0	2	0	-2	2	0	0	-1	-1	-1	0
*	8	oI	0.0	83.1	85.9	85.9	90.0	90.0	90.0	0	0	2	0	0	2	0	0	-2	0	0	0
*	16	oF	0.0	83.1	121.5	121.5	90.0	90.0	90.0	0	0	2	0	-2	2	0	0	-2	-2	0	0
*	6	tI	0.0	85.9	85.9	83.1	90.0	90.0	90.0	0	-2	0	0	-2	0	0	0	0	0	-2	0
*	24	hR	7.3	119.5	121.5	73.6	90.0	88.5	120.5	2	0	-2	0	-2	2	0	0	1	1	1	0
*	4	hR	21.8	119.5	119.5	73.6	88.5	91.5	118.9	0	2	2	0	-2	0	-2	0	-1	-1	1	0
*	7	tI	21.8	85.9	83.1	85.9	90.0	90.0	90.0	-2	0	0	0	0	0	2	0	0	2	0	0
*	5	cI	29.0	85.9	83.1	85.9	90.0	90.0	90.0	0	-2	0	0	0	0	-2	0	2	0	0	0
	31	aP	159.4	73.6	73.6	73.6	68.8	71.4	108.6	-1	-1	-1	0	-1	1	1	0	-1	1	-1	0
	20	mC	181.2	83.1	121.5	73.6	90.0	124.4	90.0	0	0	2	0	2	-2	0	0	1	1	-1	0
	10	mC	181.2	83.1	121.5	73.6	90.0	124.4	90.0	0	0	-2	0	2	-2	0	0	-1	-1	1	0
	3	cP	181.2	73.6	73.6	73.6	108.6	108.6	111.2	-1	1	-1	0	1	-1	-1	0	-1	-1	1	0
	41	mC	228.2	138.7	73.6	73.6	108.6	120.5	81.6	1	-1	-3	0	-1	1	-1	0	1	1	1	0
	37	mC	228.2	138.7	73.6	73.6	108.6	120.5	81.6	1	1	-3	0	-1	-1	-1	0	-1	1	1	0
	39	mC	228.2	138.7	73.6	73.6	108.6	120.5	81.6	1	1	-3	0	-1	-1	-1	0	-1	1	1	0
	23	oC	318.8	83.1	121.5	73.6	90.0	124.4	90.0	0	0	2	0	-2	2	0	0	-1	-1	-1	0
	35	mP	318.8	73.6	73.6	73.6	108.6	111.2	108.6	-1	1	-1	0	1	1	1	0	1	-1	-1	0
	42	oI	318.8	73.6	73.6	119.5	88.5	91.5	108.6	1	1	1	0	-1	1	-1	0	-2	0	2	0
	15	tI	318.8	73.6	73.6	119.5	91.5	88.5	108.6	-1	-1	-1	0	1	-1	1	0	-2	0	2	0
	13	oC	318.8	83.1	121.5	73.6	90.0	124.4	90.0	0	0	-2	0	-2	2	0	0	1	1	1	0
	33	mP	318.8	73.6	73.6	73.6	108.6	111.2	108.6	-1	1	-1	0	1	1	1	0	1	-1	-1	0
	34	mP	318.8	73.6	73.6	73.6	108.6	111.2	108.6	-1	1	1	0	-1	-1	-1	0	1	-1	1	0
	22	hP	387.7	73.6	73.6	73.6	108.6	108.6	111.2	1	-1	1	0	-1	1	1	0	-1	-1	-1	0
	40	oC	387.7	73.6	138.7	73.6	59.5	108.6	98.4	-1	1	-1	0	-1	1	3	0	1	1	1	0
	12	hP	387.7	73.6	73.6	73.6	108.6	108.6	111.2	1	-1	-1	0	-1	1	-1	0	1	1	1	0
	36	oC	387.7	73.6	138.7	73.6	59.5	108.6	98.4	-1	1	-1	0	-1	1	3	0	1	1	1	0
	38	oC	387.7	73.6	138.7	73.6	59.5	108.6	98.4	-1	1	1	0	-1	1	-3	0	-1	-1	-1	0
	30	mC	467.4	73.6	138.7	73.6	59.5	71.4	81.6	-1	1	1	0	1	-1	3	0	1	1	1	0
	29	mC	467.4	73.6	138.7	73.6	59.5	71.4	81.6	-1	-1	-1	0	1	1	-3	0	1	-1	-1	0
	28	mC	467.4	73.6	138.7	73.6	59.5	71.4	81.6	-1	1	1	0	1	-1	3	0	1	1	1	0
	21	tP	500.0	73.6	73.6	73.6	108.6	108.6	111.2	1	-1	1	0	-1	1	1	0	-1	-1	-1	0
	32	oP	500.0	73.6	73.6	73.6	111.2	108.6	108.6	-1	-1	-1	0	1	-1	1	0	-1	1	1	0
	2	hR	500.0	119.5	85.9	142.0	107.6	89.2	135.9	0	-2	-2	0	0	2	0	0	3	-1	1	0
	11	tP	500.0	73.6	73.6	73.6	108.6	108.6	111.2	1	-1	-1	0	-1	1	-1	0	1	1	1	0
	1	cF	568.8	138.7	73.6	142.0	100.8	116.8	98.4	-1	-1	-3	0	-1	-1	1	0	-1	3	1	0
	27	mC	637.7	138.7	73.6	85.9	54.3	108.0	81.6	-1	-1	3	0	1	1	1	0	0	2	0	0
	19	oI	637.7	73.6	85.9	138.7	72.0	59.5	54.3	1	1	1	0	0	2	0	0	-1	1	3	0
	26	oF	672.1	73.6	138.7	184.4	62.4	130.8	98.4	-1	-1	-1	0	-1	-1	3	0	-1	3	3	0
	18	tI	672.1	85.9	138.7	73.6	59.5	125.7	108.0	0	2	0	0	1	-1	-3	0	-1	-1	-1	0
	9	hR	728.2	73.6	83.1	252.0	60.3	132.6	124.4	-1	-1	-1	0	0	0	2	0	-1	5	3	0

I4-tetragonal bodycentered

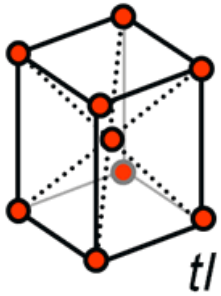


Flow scheme for structure determination

Collect data



Determined the space group



Scale the data
(what data to use)

XDS output- to decide what data to include

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
5.57	664	421	978	43.0%	0.9%	1.7%	425	40.38	1.2%	99.9*	-24	0.473	14
3.95	1173	844	1724	49.0%	1.1%	1.8%	586	34.54	1.5%	99.9*	-60	0.649	6
3.23	1503	1128	2200	51.3%	1.4%	2.0%	709	26.89	1.9%	99.9*	0	0.000	1
2.80	1993	1482	2640	56.1%	3.3%	3.6%	1005	15.87	4.6%	99.7*	0	0.000	0
2.50	2147	1668	2963	56.3%	6.2%	6.3%	958	9.67	8.8%	98.7*	0	0.000	0
2.28	2503	1887	3244	58.2%	10.6%	11.4%	1232	5.65	15.0%	97.4*	0	0.000	0
2.11	2613	2038	3551	57.4%	19.8%	22.2%	1149	2.99	28.0%	88.9*	0	0.000	0
1.98	2911	2231	3787	58.9%	39.1%	44.6%	1360	1.52	55.2%	72.6*	0	0.000	0
1.87	1994	1647	4042	40.7%	68.9%	77.2%	694	0.75	97.4%	45.0*	0	0.000	0
total	17501	13346	25129	53.1%	1.9%	2.6%	8118	10.30	2.7%	100.0*	-54	0.520	21

NUMBER OF REFLECTIONS IN SELECTED SUBSET OF IMAGES	18076
NUMBER OF REJECTED MISFITS	575
NUMBER OF SYSTEMATIC ABSENT REFLECTIONS	0
NUMBER OF ACCEPTED OBSERVATIONS	17501
NUMBER OF UNIQUE ACCEPTED REFLECTIONS	13346

Aimless-does the statistics look good?

Run of AIMLESS on 30/11/2018 at 14:01:42

Result

Summary data for Project: XDSproject Crystal: XDScrystal Dataset: XDSdataset

	Overall	InnerShell	OuterShell
Low resolution limit	60.69	60.69	1.91
High resolution limit	1.87	8.95	1.87
Rmerge (within I+/I-)	0.036	0.019	1.800
Rmerge (all I+ and I-)	0.039	0.020	2.158
Rmeas (within I+/I-)	0.043	0.023	2.390
Rmeas (all I+ & I-)	0.042	0.022	2.559
Rpim (within I+/I-)	0.023	0.013	1.554
Rpim (all I+ & I-)	0.016	0.009	1.338
Rmerge in top intensity bin	0.025	-	-
Total number of observations	160906	1369	3695
Total number unique	24728	231	1270
Mean((I)/sd(I))	19.2	71.8	0.5
Mn(I) half-set correlation CC(1/2)	1.000	0.999	0.201
Completeness	98.7	93.6	80.6
Multiplicity	6.5	5.9	2.9
Mean(Chi^2)	1.00	0.89	0.98
Anomalous completeness	96.4	94.0	47.5
Anomalous multiplicity	3.2	3.1	1.9
DelAnom correlation between half-sets	-0.208	-0.311	0.032
Mid-Slope of Anom Normal Probability	0.944	-	-

The anomalous signal appears to be weak so anomalous flag was left OFF

Estimates of resolution limits: overall
- from half-dataset correlation CC(1/2) > 0.30: limit = 1.90A
- from Mn(I/sd) > 1.50: limit = 1.95A
- from Mn(I/sd) > 2.00: limit = 1.98A

Estimates of resolution limits in reciprocal lattice directions:

Along h k plane
- from half-dataset correlation CC(1/2) > 0.30: limit = 1.95A
- from Mn(I/sd) > 1.50: limit = 2.02A

Along l axis
- from half-dataset correlation CC(1/2) > 0.30: limit = 1.93A
- from Mn(I/sd) > 1.50: limit = 1.93A

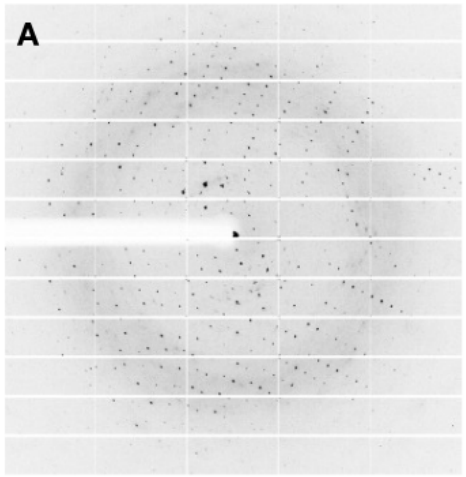
Anisotropic deltaB (i.e. range of principal components), A^2: 2.83

Average unit cell: 85.83 85.83 83.05 90.00 90.00 90.00
Space group: I 4
Average mosaicity: 0.12

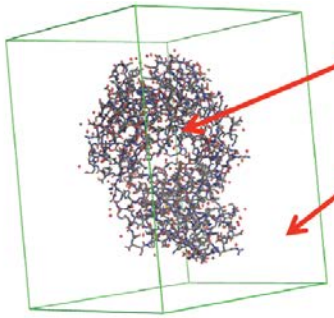
Minimum and maximum SD correction factors: Fulls 0.74 2.65 Partials 0.00 0.00

Flow scheme for structure determination

Collect data



Determine the space group



Scale the data
(what data to use)

Solve the phases
(molecular replacement)

The phase problem

Light detectors, such as CCDs, measure only the intensity of the light that hits them. This measurement is incomplete because a light wave has not only an amplitude (related to the intensity), but also a phase, which is systematically lost in a measurement.

In order to obtain an interpretable electron density map, both amplitude and phase must be known.

Solving the phase problem

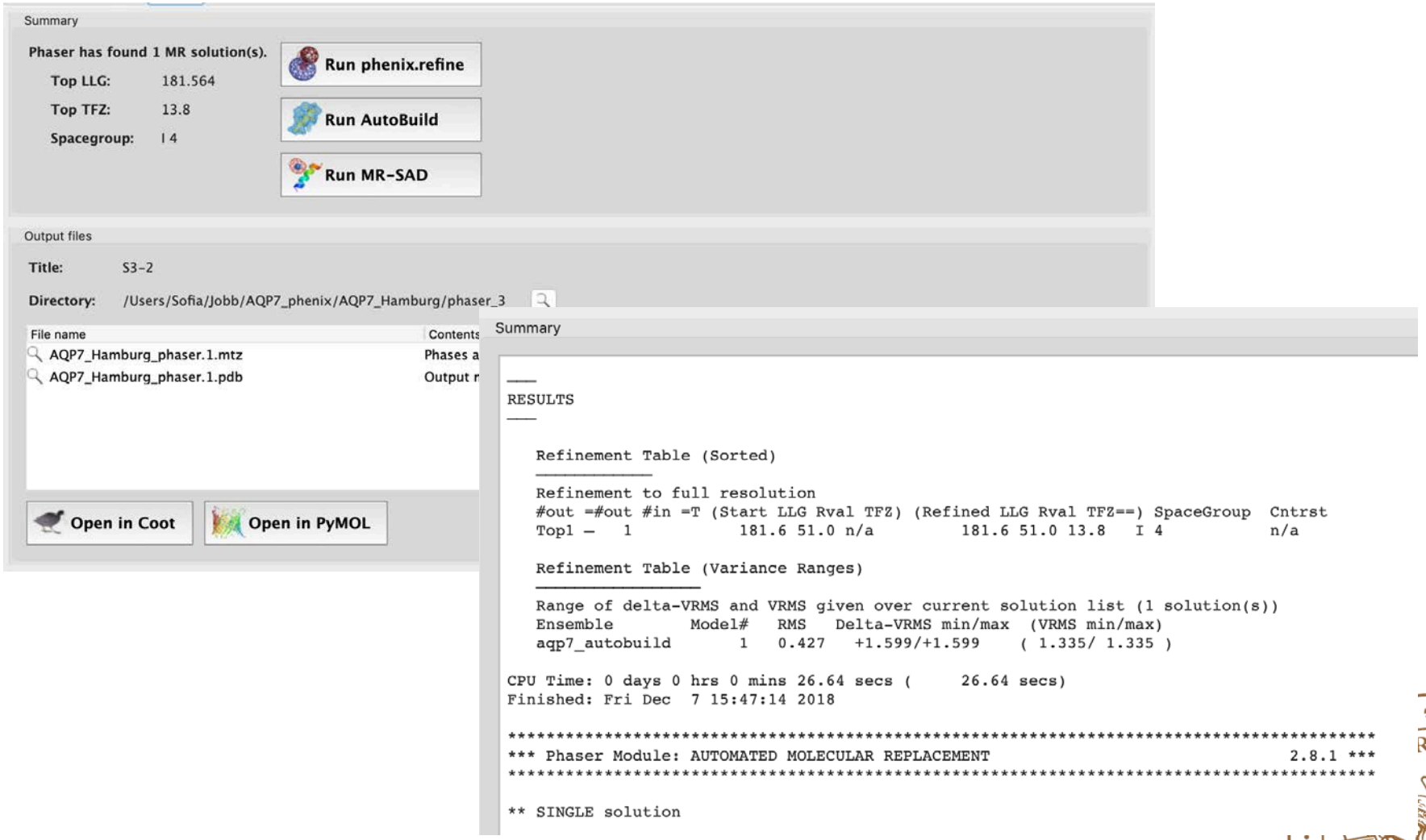
Initial phase estimates can be obtained in a variety of ways:

Molecular replacement– if a related structure is known, it can be used as a search model in molecular replacement to determine the orientation and position of the molecules within the unit cell. The phases obtained this way can be used to generate electron density maps.

Anomalous X-ray scattering (*MAD or SAD phasing)*– by incorporating anomalous scattering atoms into the methionines (seleno-methionine). A MAD experiment can then be conducted around the absorption edge, which should then yield the position of any methionine residues within the protein, providing initial phases, one can solve for the substructure of the anomalously diffracting atoms and use this to solve the structure of the whole molecule.

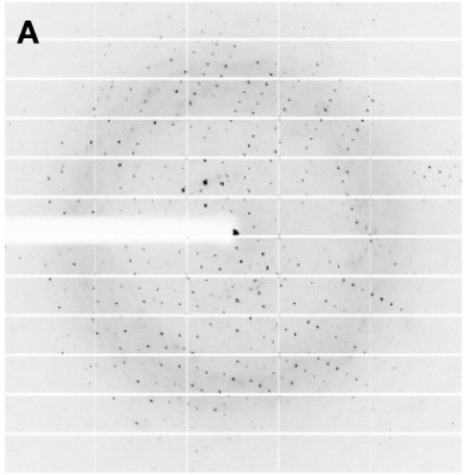
Heavy atom methods (multiple isomorphous replacement, MIR) – If electron-dense metal atoms can be introduced into the crystal, Patterson function can be used to determine their location and to obtain initial phases. Such heavy atoms can be introduced either by soaking the crystal in a heavy atom-containing solution, or by co-crystallization (growing the crystals in the presence of a heavy atom). As in MAD phasing, the changes in the scattering amplitudes can be interpreted to yield the phases. Although this is the original method by which protein crystal structures were solved, it has largely been superseded by MAD phasing with selenomethionine.

Molecular replacement Phaser-MR, Phenix (1.9 Å structure, single solution)

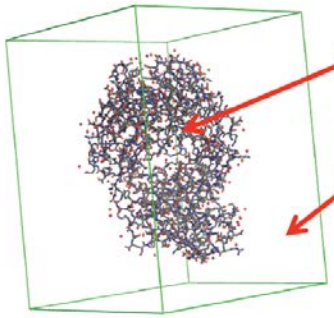


Flow scheme for structure determination

Collect data

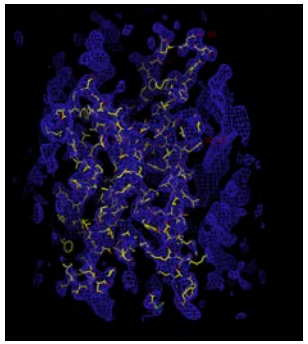


Determine the space group

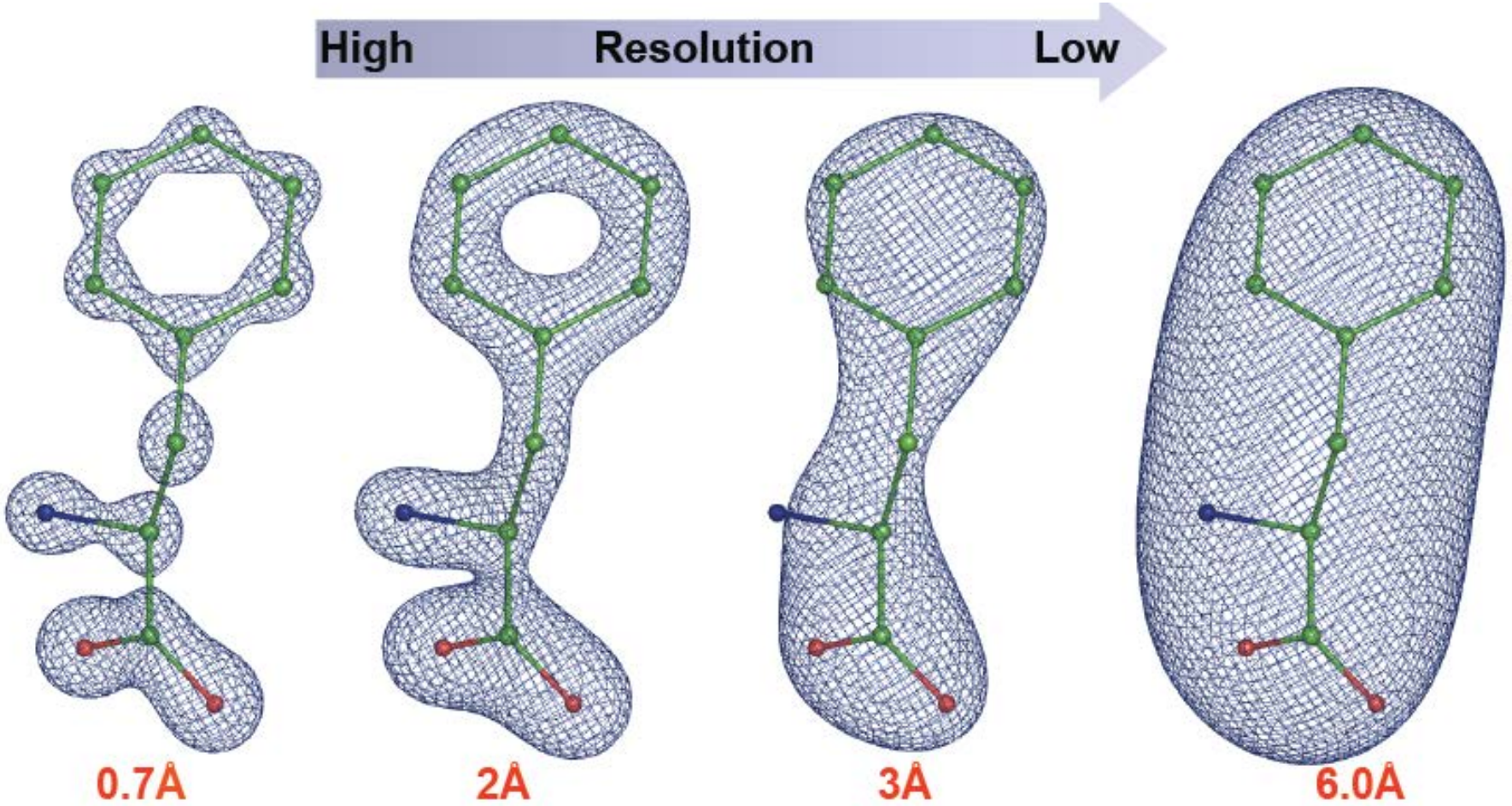


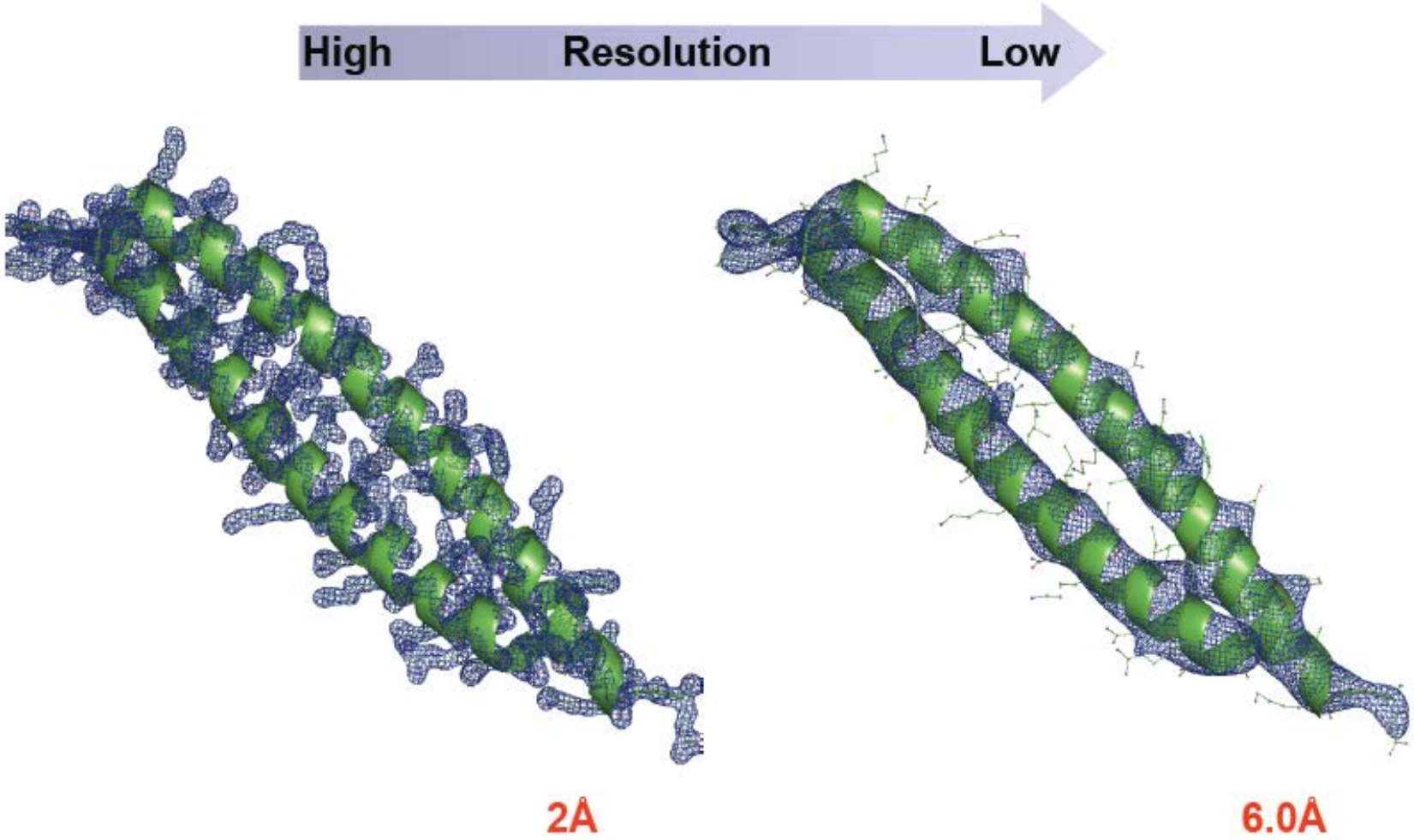
Scale the data
(what data to use)

Fourier transform the data and
calculate the map



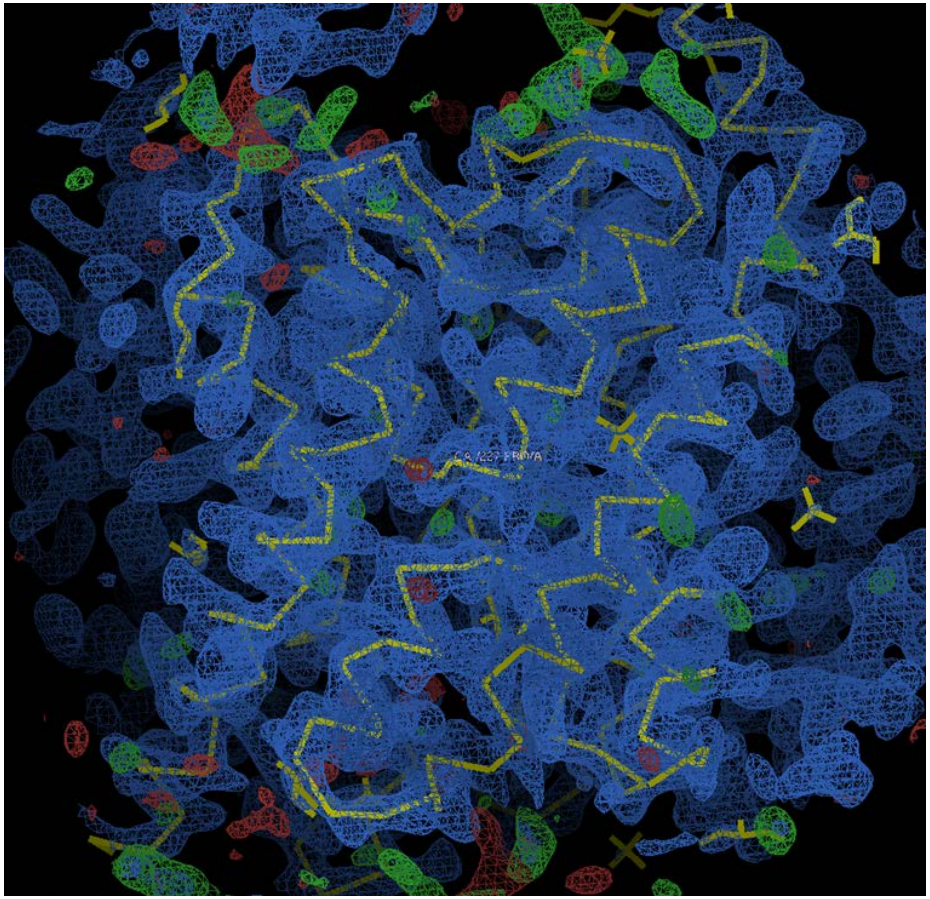
Solve the phases
(molecular replacement)

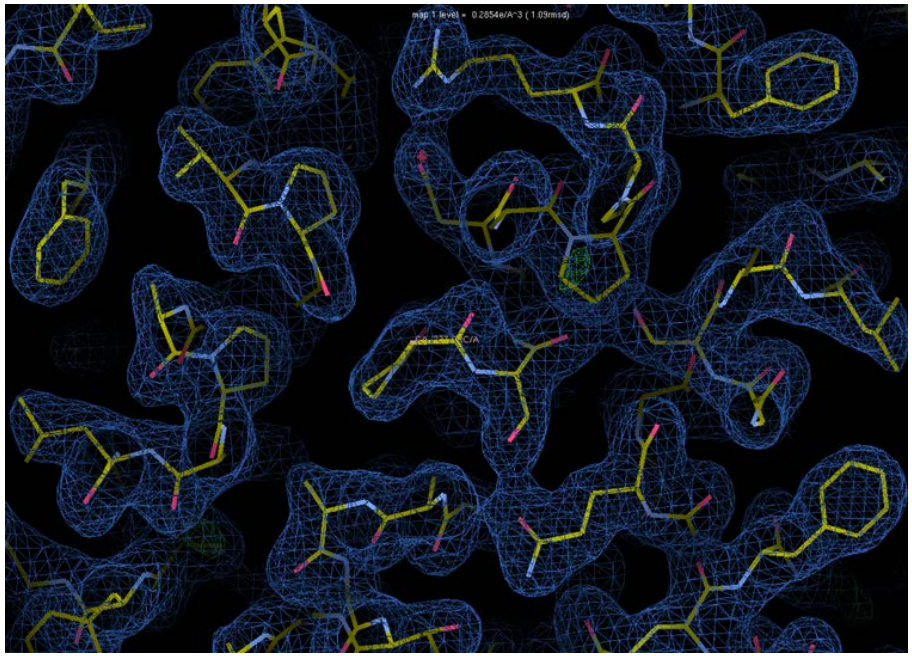

High
Resolution
Low
0.7Å
2Å
3Å
6.0Å


High
Resolution
Low
2Å
6.0Å

Electron density for human AQP7 (1.9 Å)

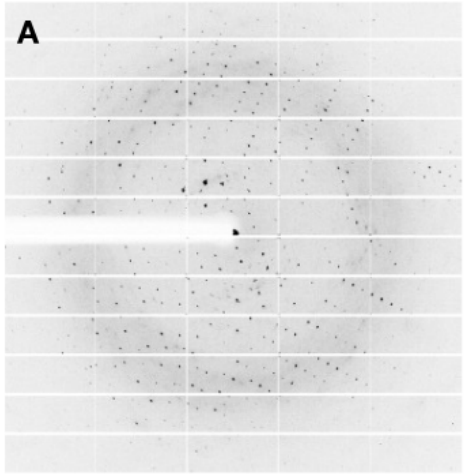
The whole structure



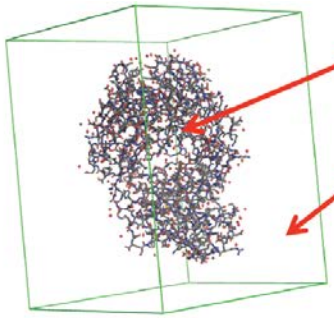


Flow scheme for structure determination

Collect data

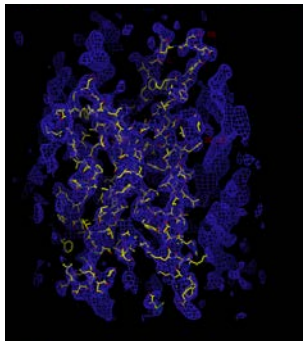


Determine the space group



Scale the data
(what data to use)

Fourier transform the data and
calculate the map

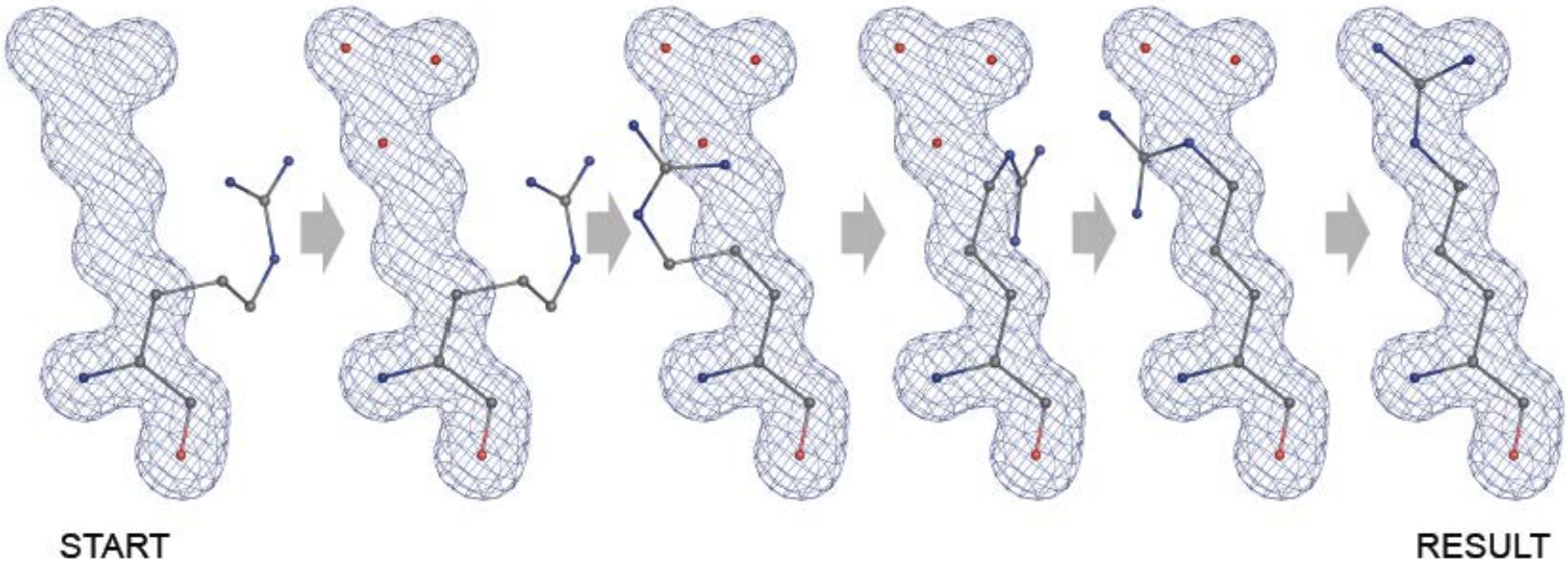


Rebuild the model

Refine the structure

Solve the phases
(molecular replacement)

Refinment of an aminoacid



- Manual building alt. ARP/WARP (high res)
- Rigid body refinement
- Simulated annealing

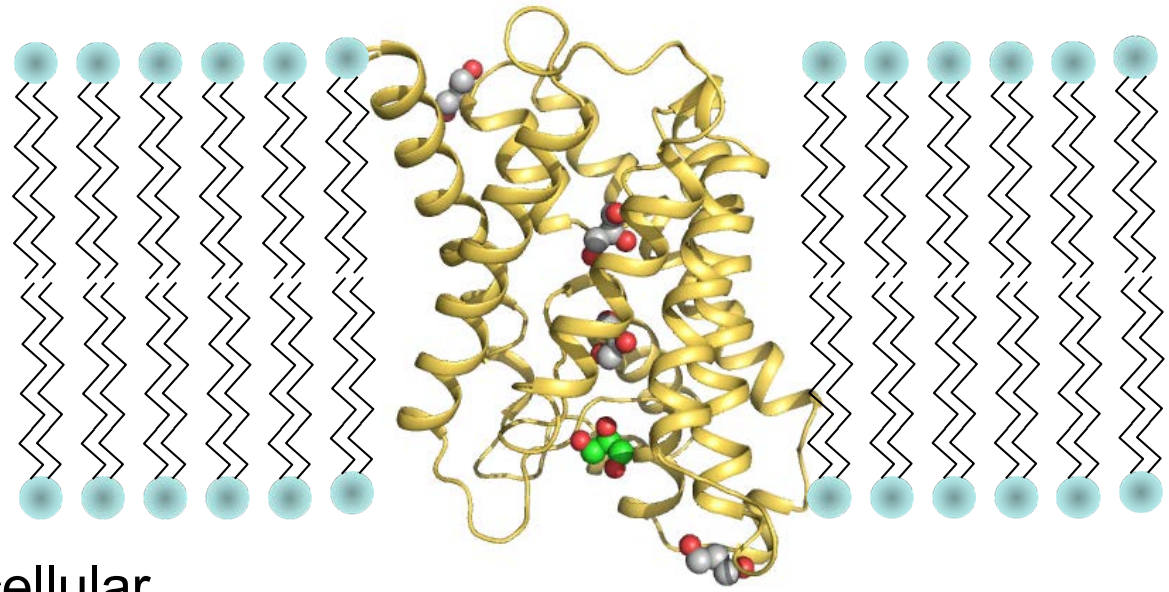
Refinement

In crystallography, the **R-factor** (sometimes called residual factor or reliability factor) is a measure of the agreement between the crystallographic model and the experimental X-ray diffraction data. In other words, it is a measure of how well the refined structure predicts the observed data

Crystallographers also use the **Free R-Factor**, which is computed according to the same formula, but on a small, random sample of data that are set aside for the purpose and never included in the refinement. The Rfree will always be greater than Rwork, but the two statistics should be similar because a correct model should predict *all* the data with uniform accuracy.

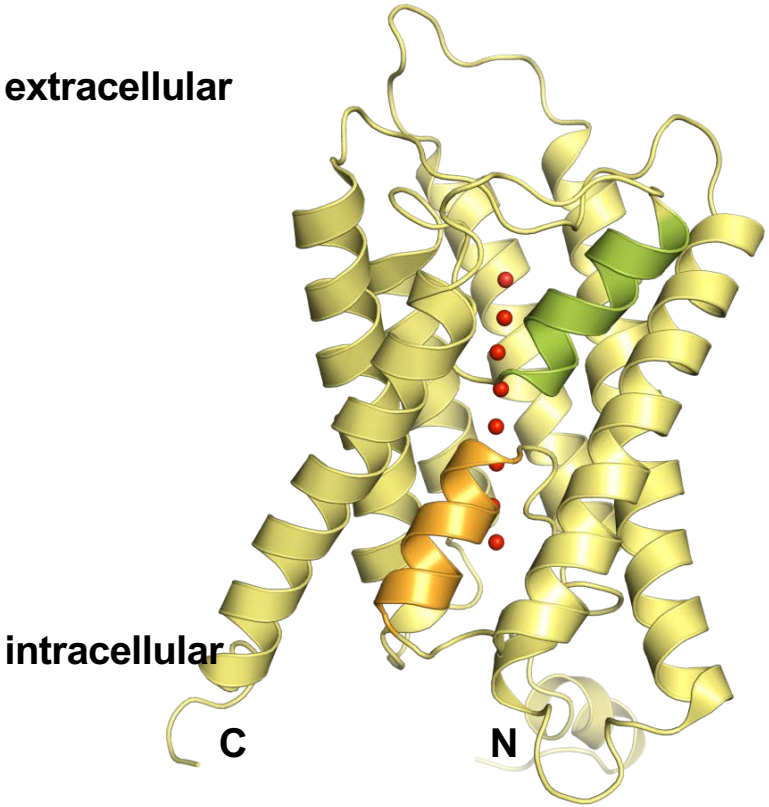
Structural studies of the glycreol channelAQP7

Extracellular



Intracellular

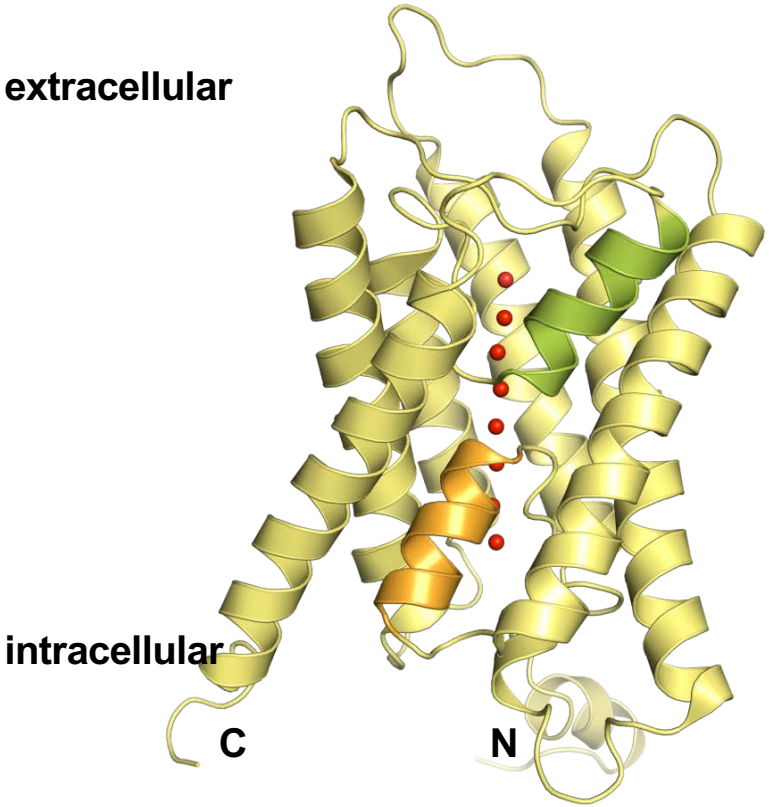
The Structure of *P. pastoris* Aqy1 at 1.15Å resolution



Monomer

	Aqy1 – pH 3.5 Crystal 1	Aqy1 – pH 8.0 Crystal 2
Data collection		
Space group	I4	I4
Cell dimensions		
a, *b*, *c* (Å)	91.4, 91.4, 80.8	90.9, 90.9, 80.6
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	1.15 (1.20-1.15)*	1.40 (1.48-1.40)
R_{sym}	8.0 (73.3)	8.8 (70.6)
$I / \sigma I$	9.3 (2.2)	12.2 (1.8)
Completeness (%)	94.7 (99.1)	100 (100)
Redundancy	3.9 (3.2)	5.3 (3.4)
Refinement		
Resolution (Å)	20.0-1.15	64.5-1.4
No. reflections	111,455	61,170
R_{work} / R_{free}	14.4/17.0	16.2/18.1
R_{work} / R_{free} (Fo>4σ)	11.7/14.0	
No. atoms		
Protein	2,059	2,080
β-OG	32	80
Cl^-	3	2
Water	141	133
B-factors		
Protein	15.9	10.4
β-OG	28.2	24.3
Cl^-	19.1	17.1
Water	29.8	24.1
R.m.s. deviations		
Bond lengths (Å)	0.023	0.008
Bond angles (°)	2.6	1.07

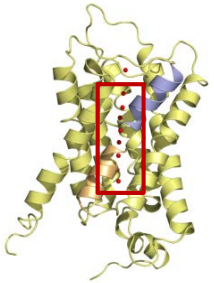
The Structure of *P. pastoris* Aqy1 at 1.15Å resolution

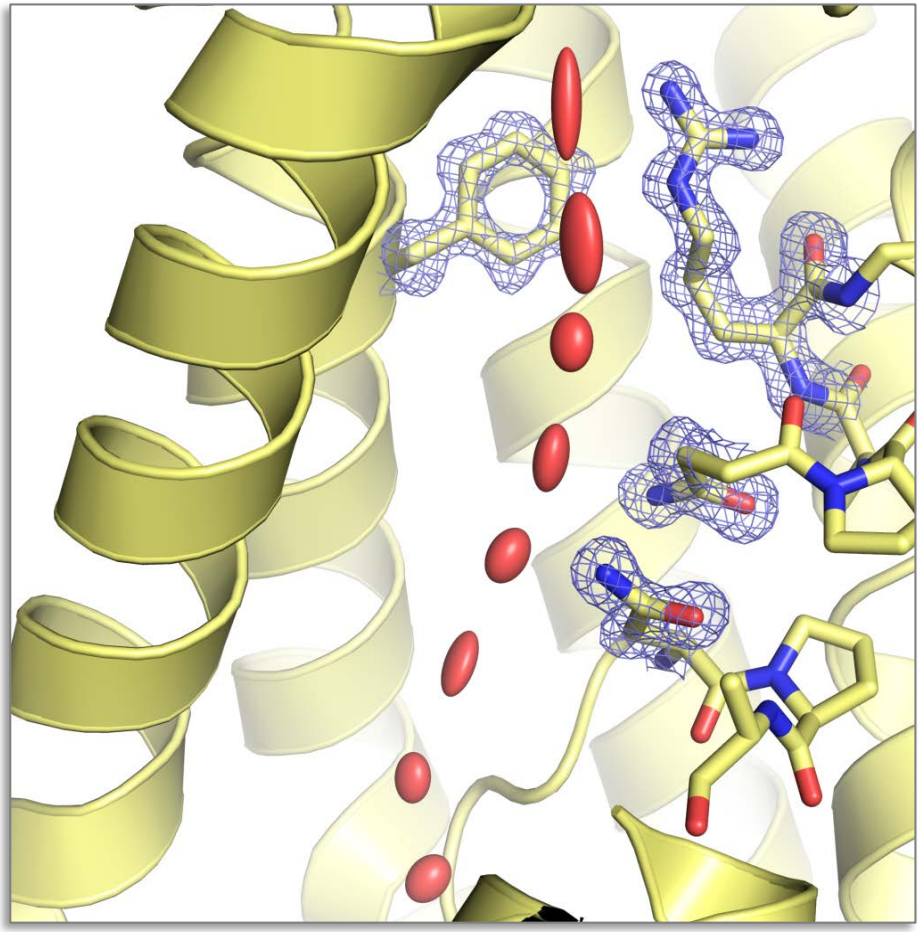


Monomer

- Hour-glass structure
- 6 TM helices
- Loops B and E form a pseudo TM helix
- NPA-motif
- ar/R constriction region

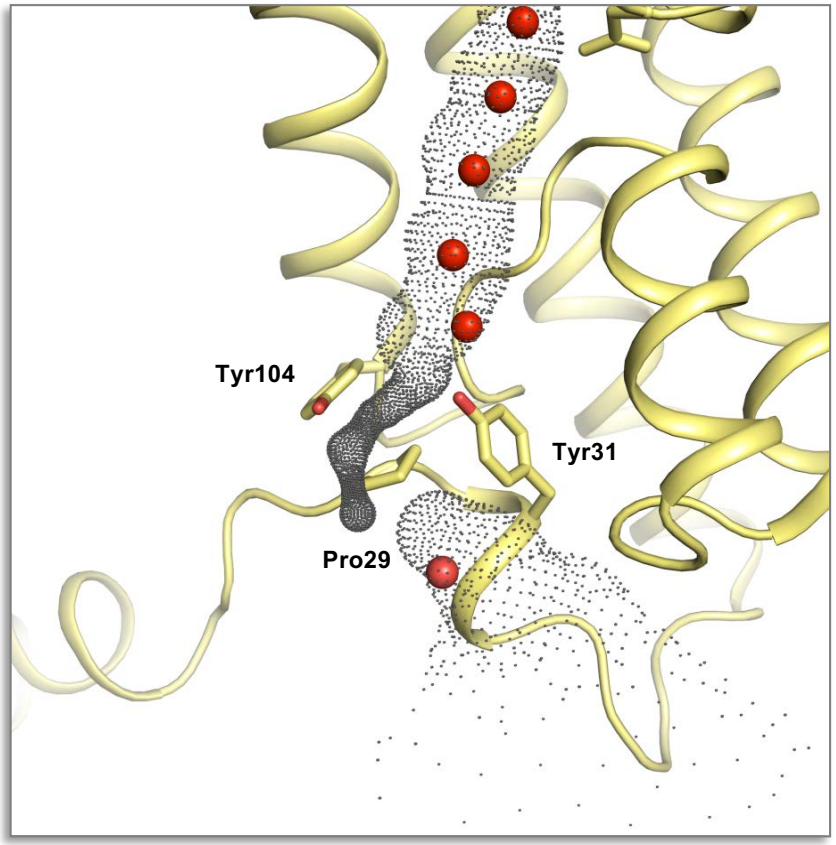
The constriction region of the water pore

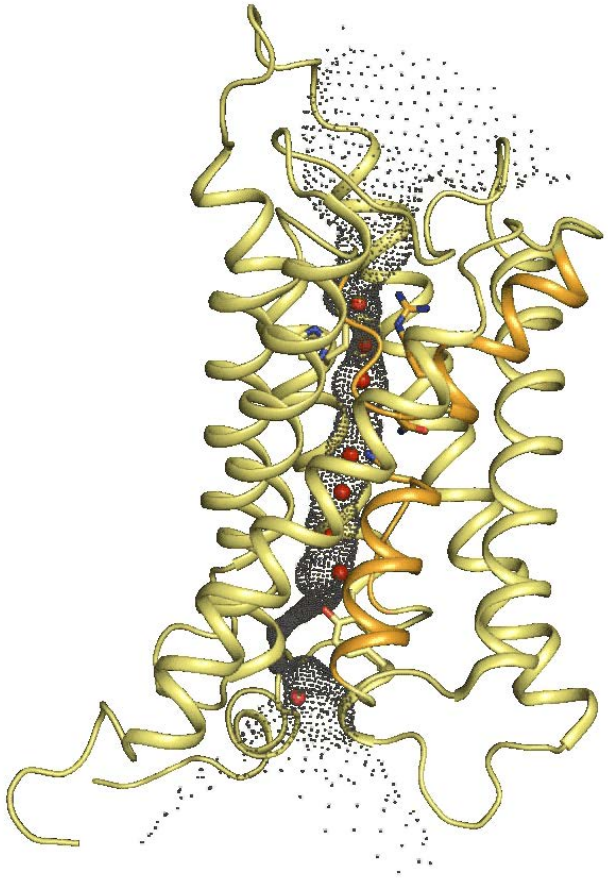




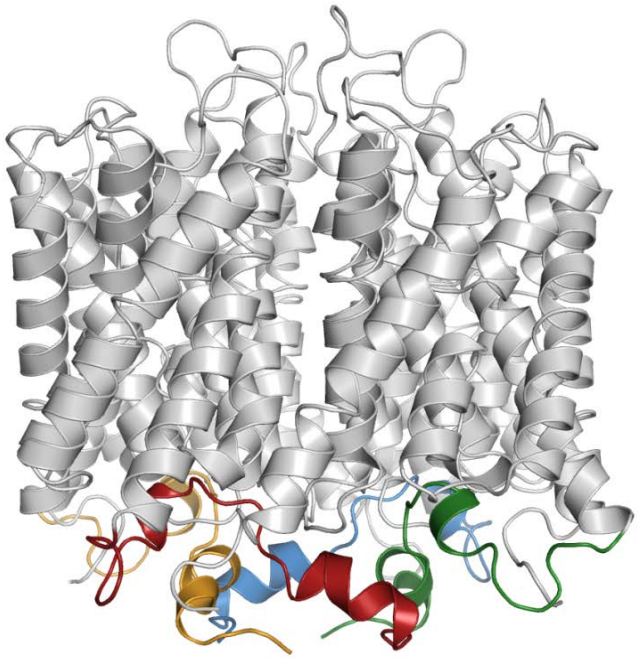
The waters are represented as ellipsoids showing the anisotropic thermal motions being parallel with the pore.

Aqy1 is closed in the structure

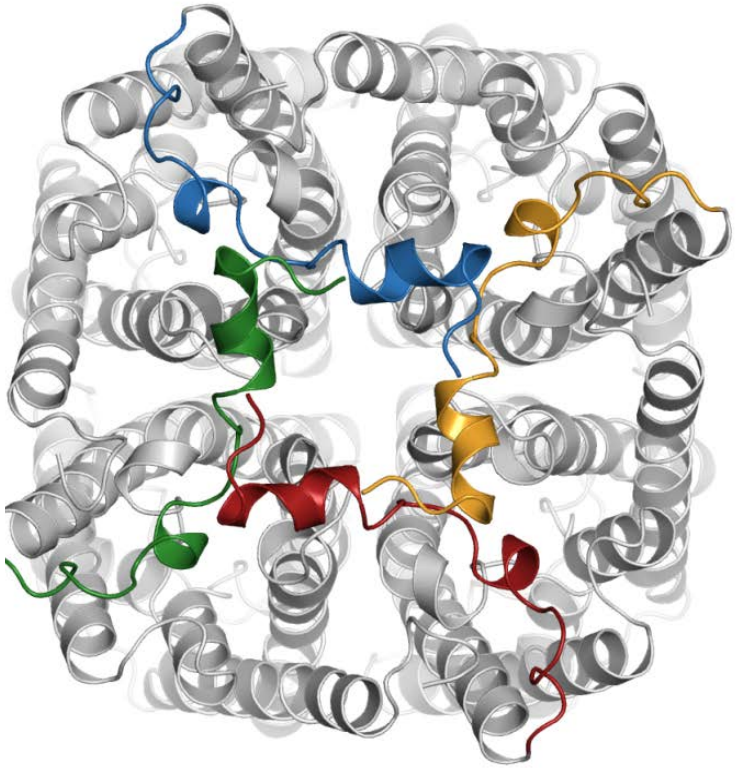




The N-terminal Bundle
- stabilizes the closed channel

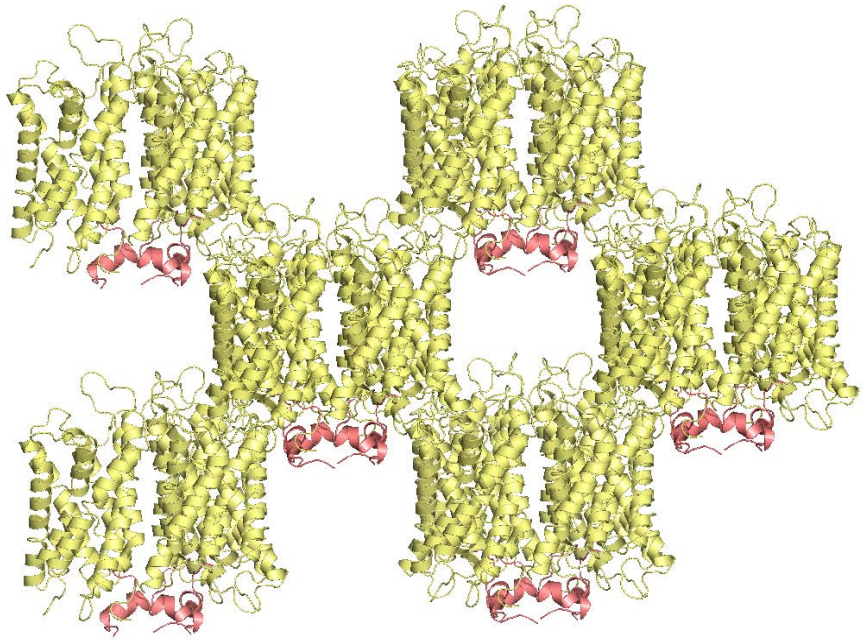


side-view

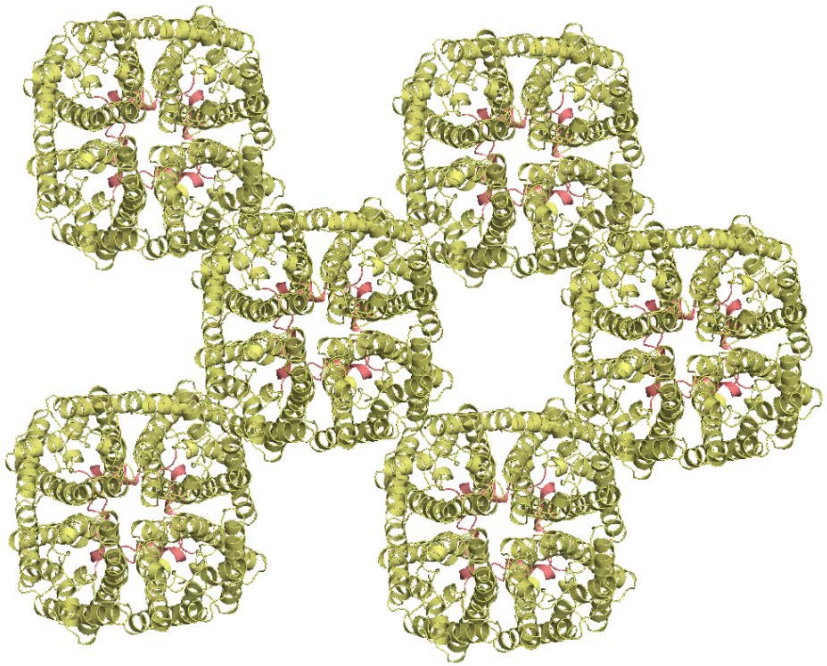


top-
view

Crystal Packing



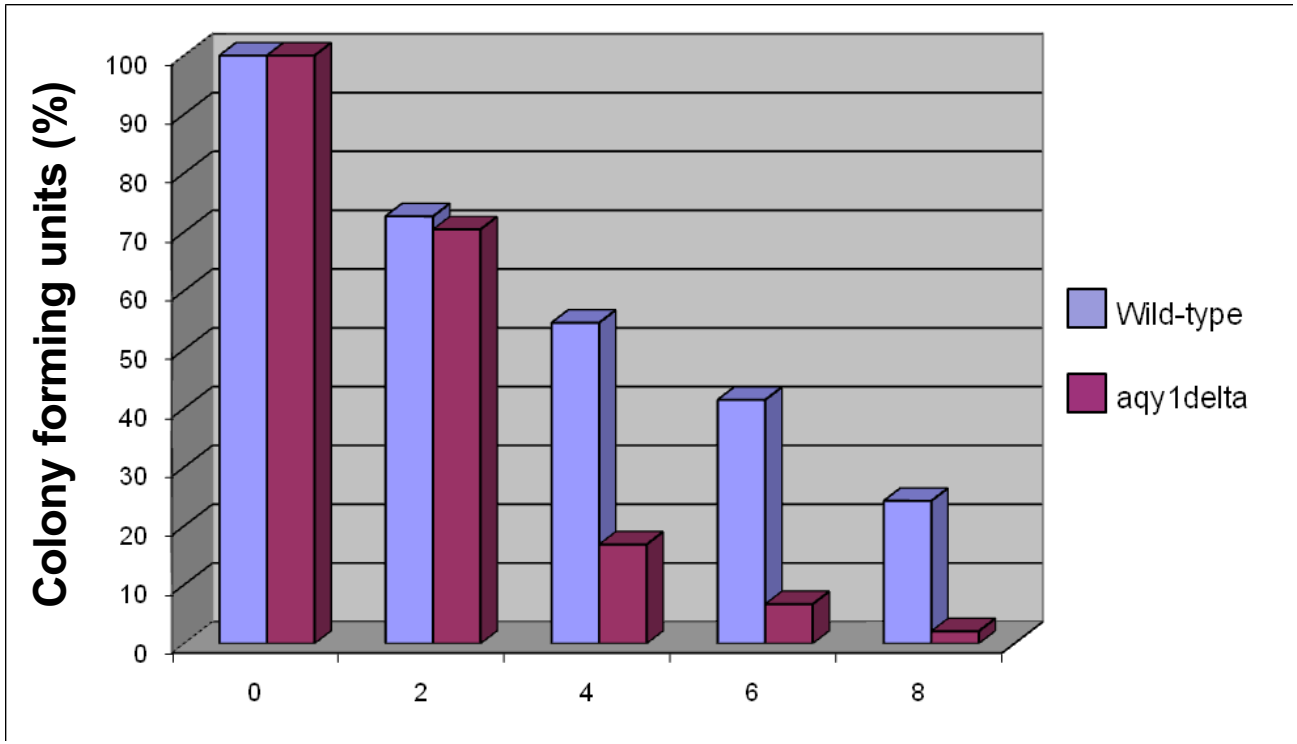
side-view

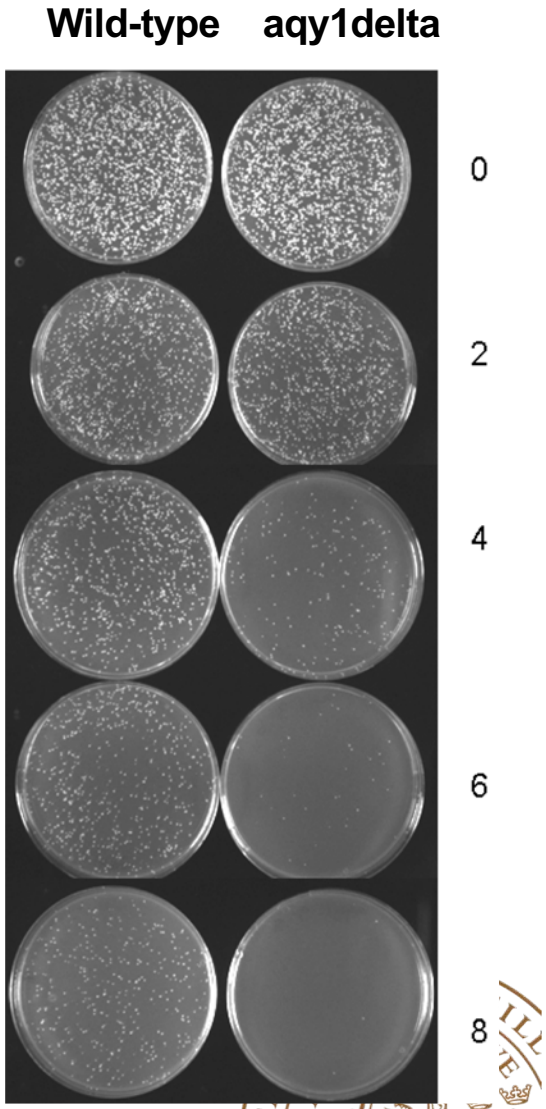


top-view

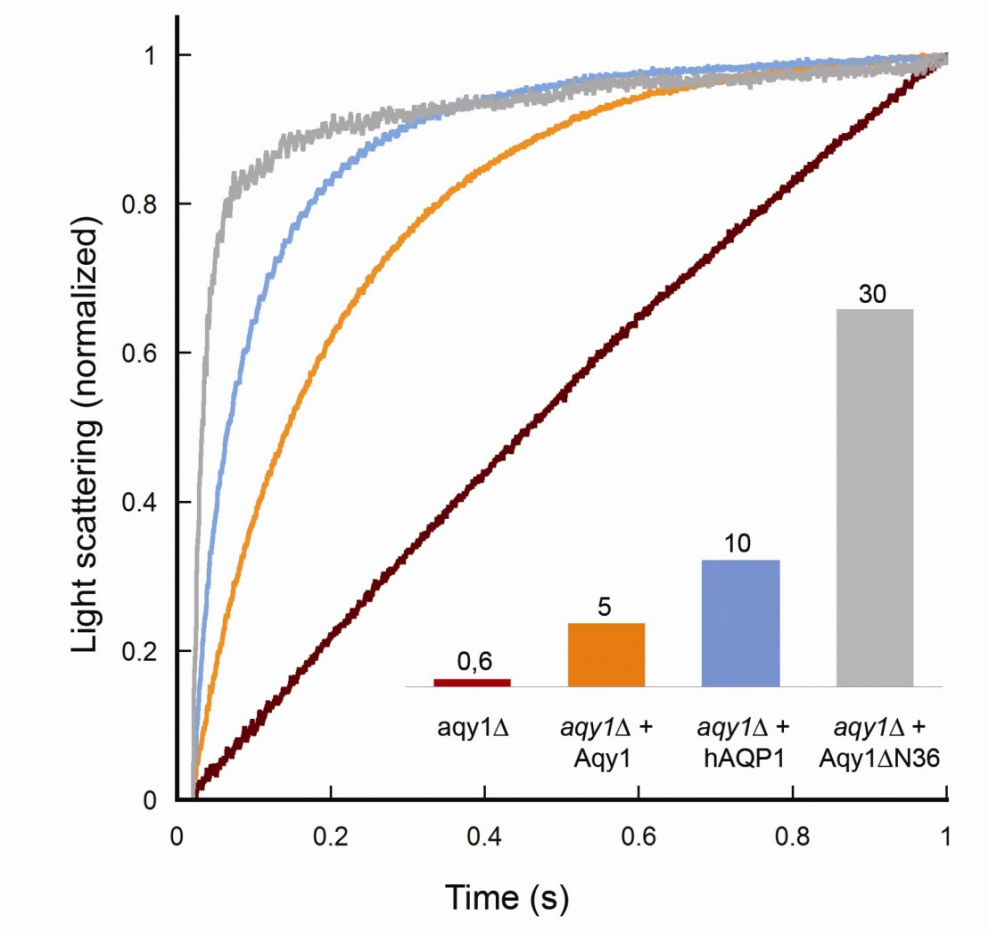
Function and activity

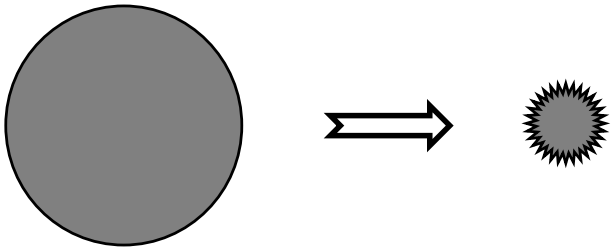
Aqy1 increases freezing/thawing tolerance





Activity assay



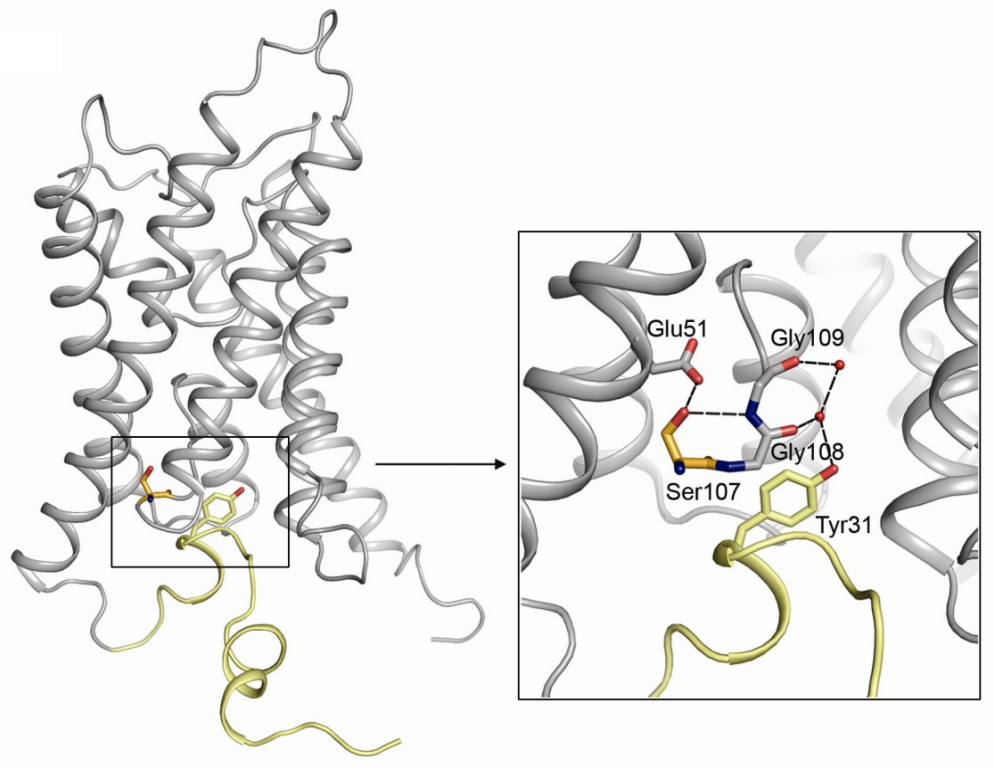


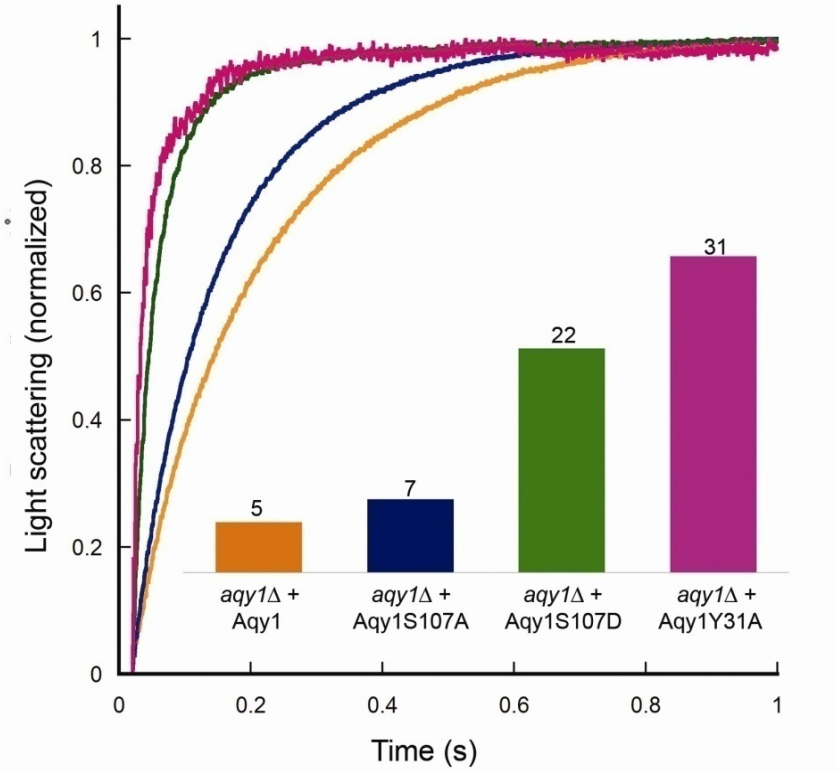
Iso-osmolal solution

hyper-osmolal solution

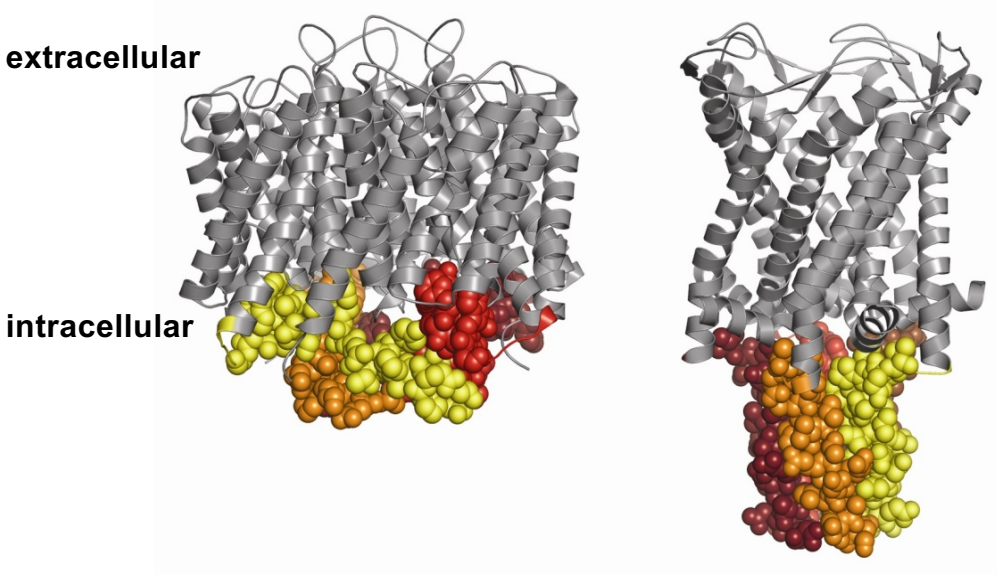
Aqy1 is gated, but how?
Gating mechanism

Gating by phosphorylation

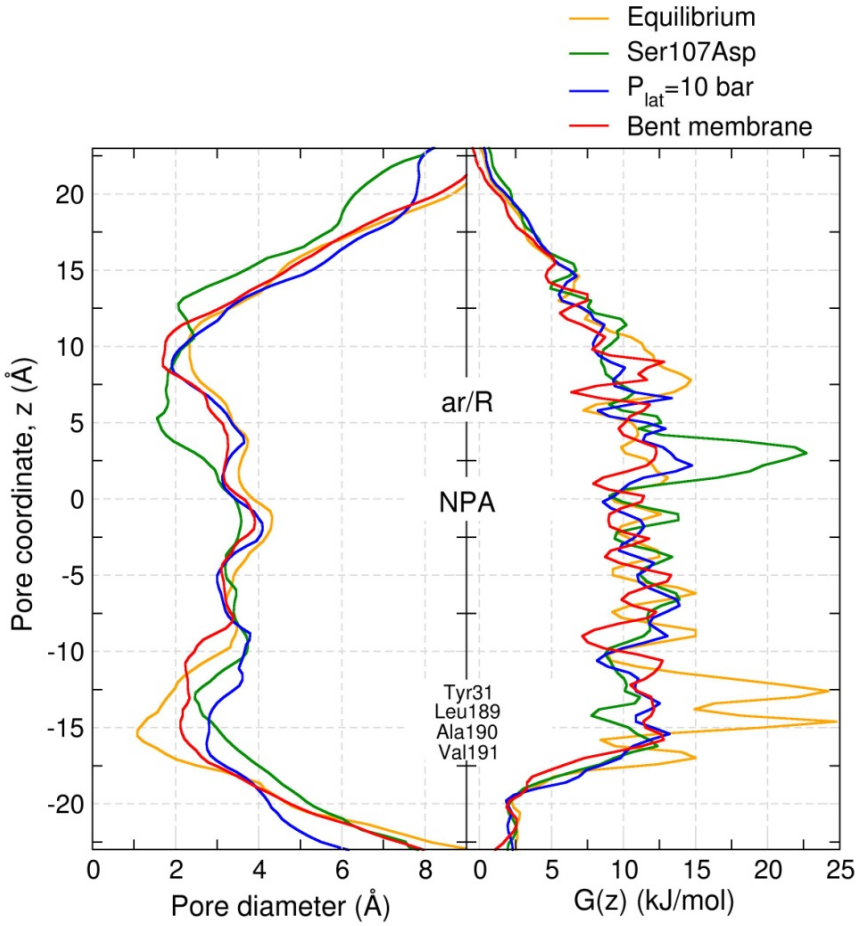




Mechanosensitive gating



Aqy1 MscL



Gating mechanisms proposed for this yeast channel

phosphorylation-and- mechanosensitive