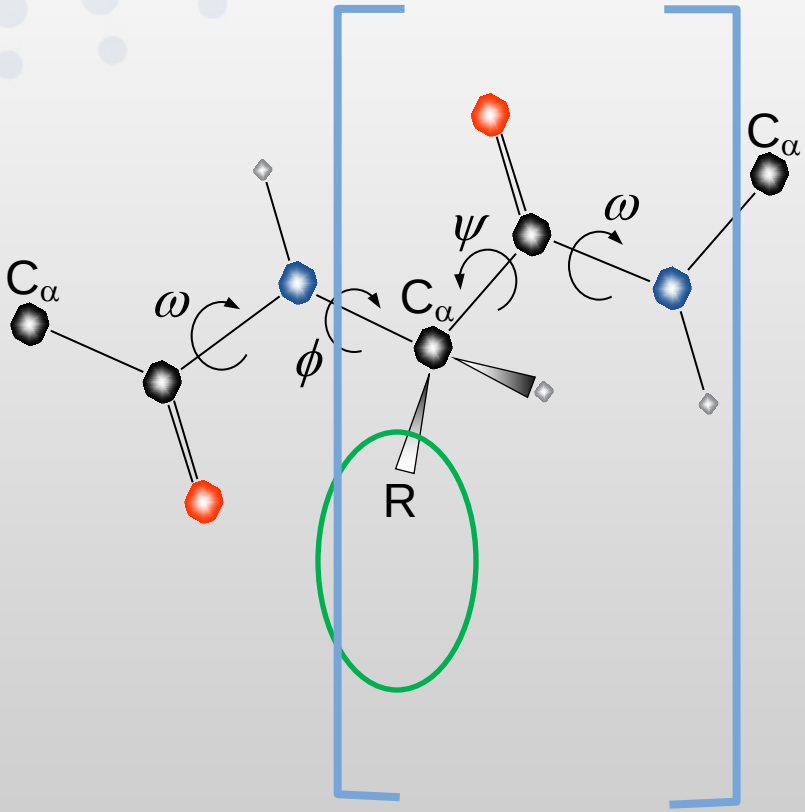
Powder Diffraction with Proteins

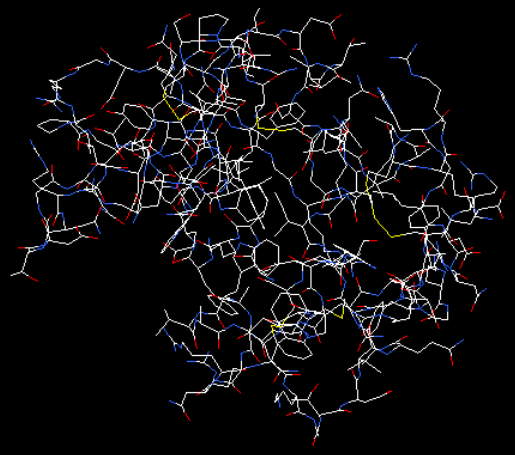
Jon Wright, Irene Margiolaki, Andy Fitch and Yves Watier

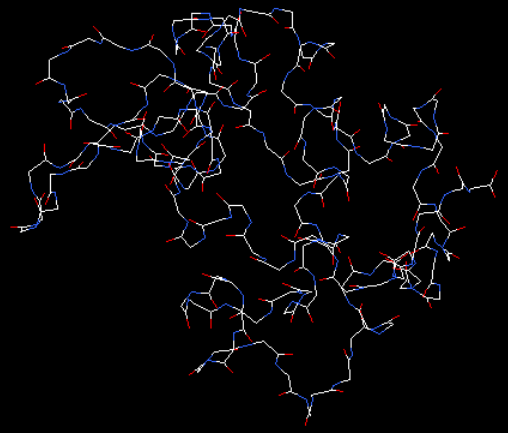
Larger systems…. proteins

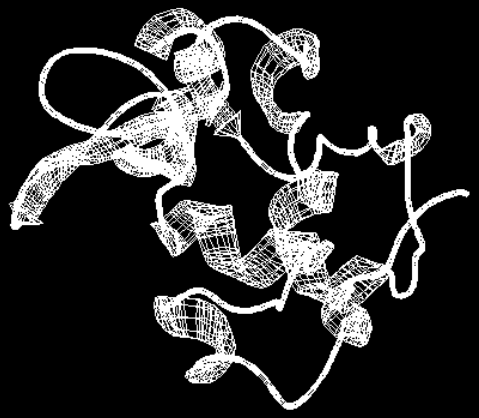
- Fundamental differences to small molecule crystallography:
  - Well known geometric constraints (polypeptide chains)
  - only ϕ, ψ and sidechain conformations which are unknown

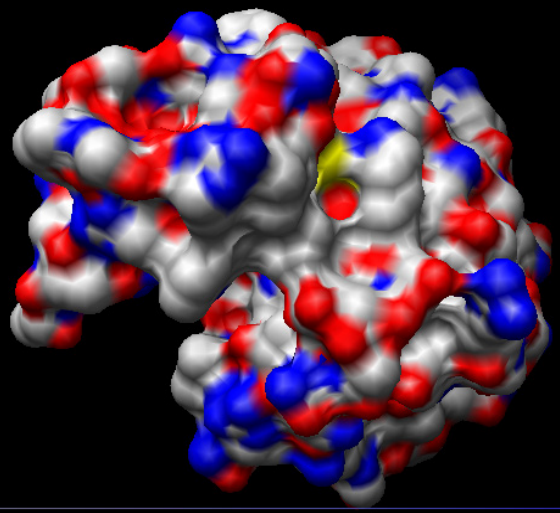


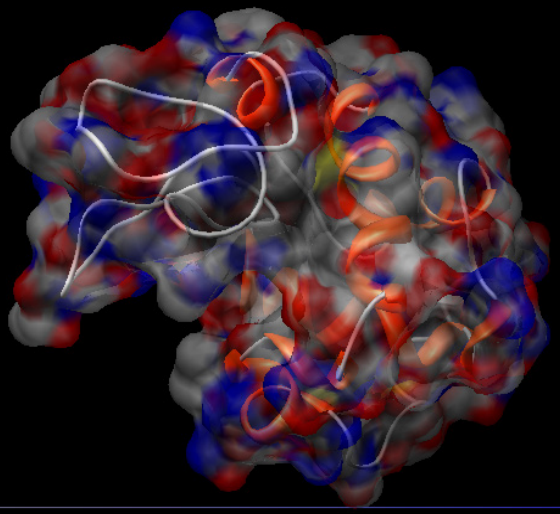
Structure representations

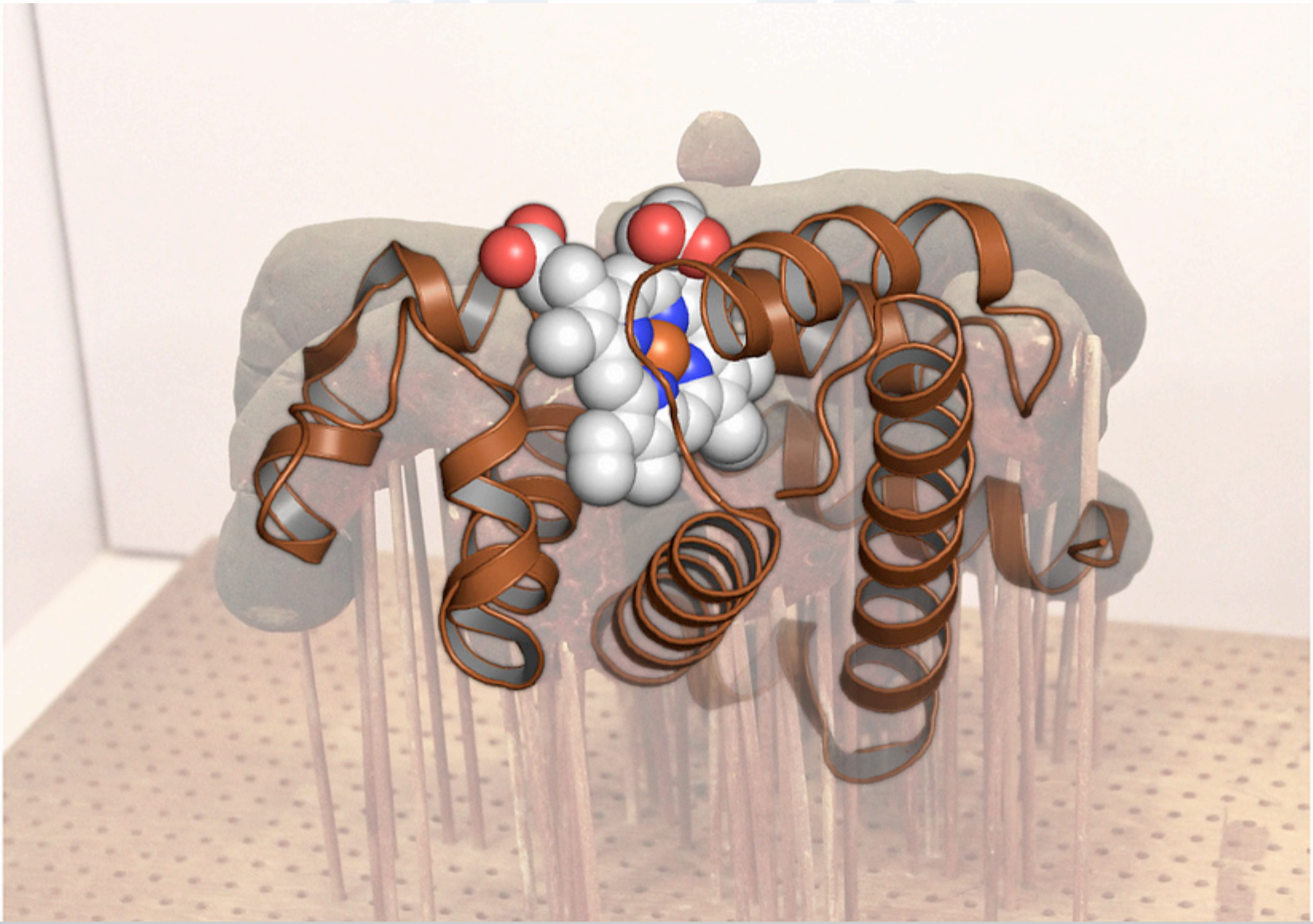




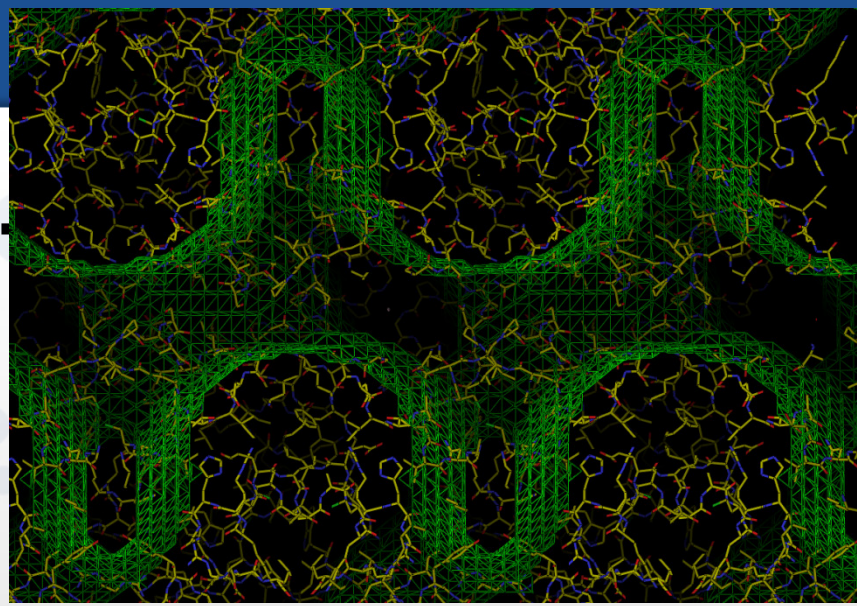








Low resolution structures..



- Much of crystal is liquid water
- Use heavy atom method to “solve” phase problem
- What do we see in electron density maps?

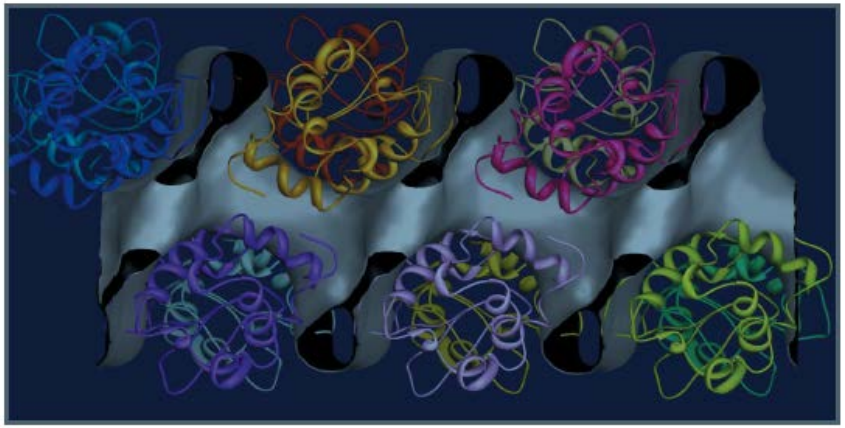


Figure 9
Solvent channels in HEWL crystals. The molecular envelope derived from the experimental data reported in this work is represented as a grey surface. The figure shows the linear solvent channel, which directly traverses the crystal parallel to the *c* axis (horizontal display direction). The protein crystal structure, represented as a main-chain ribbon model, is superimposed on this map.

J. Appl. Cryst. (2008). **41**, 329–339

Molecular envelopes derived from protein powder diffraction data

Jonathan P. Wright,[a]‡ Céline Besnard,[b]‡ Irene Margiolaki,[a] Sebastian Basso,[b] Fabrice Camus,[b] Andrew N. Fitch,[a] Gavin C. Fox,[a] Philip Pattison[b,c] and Marc Schiltz[b]*

J. Appl. Cryst. (2008). **41**

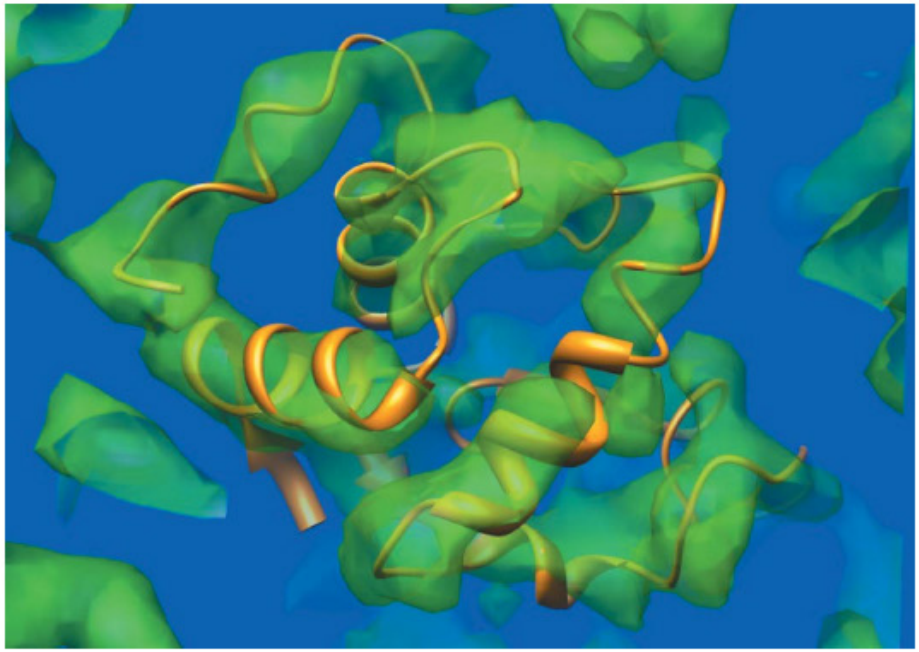


Figure 3
Electron-density map for HEWL (green) obtained from MIR data after density modification superimposed onto the known molecular structure (orange) obtained from the PDB (PDB code 6lyt; Young *et al.*, 1993).

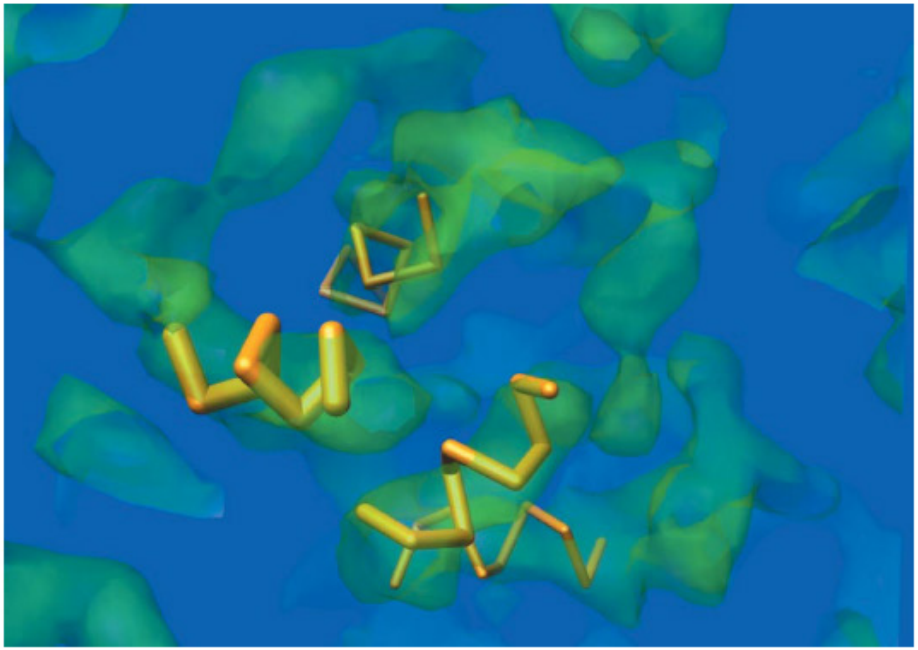
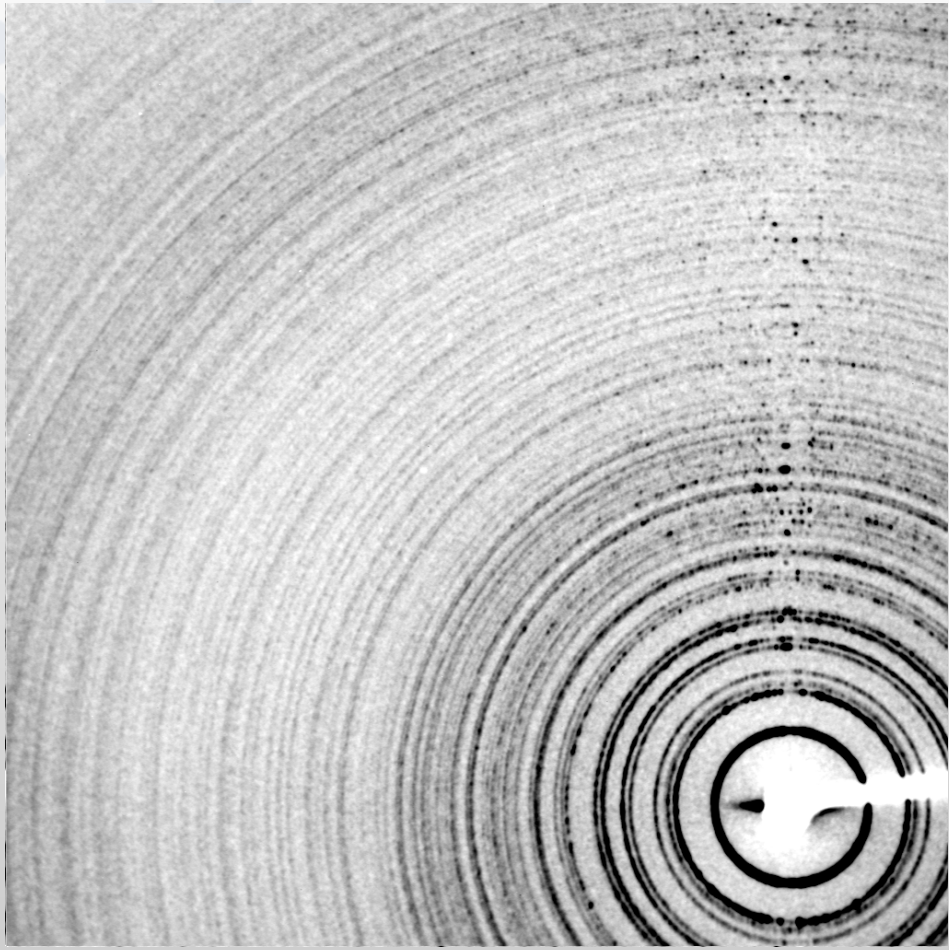


Figure 4
The four helices (orange) found by the program *FFFEAR* superimposed on the electron-density map for HEWL (green) obtained from MIR data after density modification. The thickness of each helix is representative of the accuracy of the fit between the search target and the electron density: the thicker the helix, the lower its score and the more accurate the fit.

Many grains make a powder

- 1
- 2
- 3
- 5
- 10
- 20
- 50

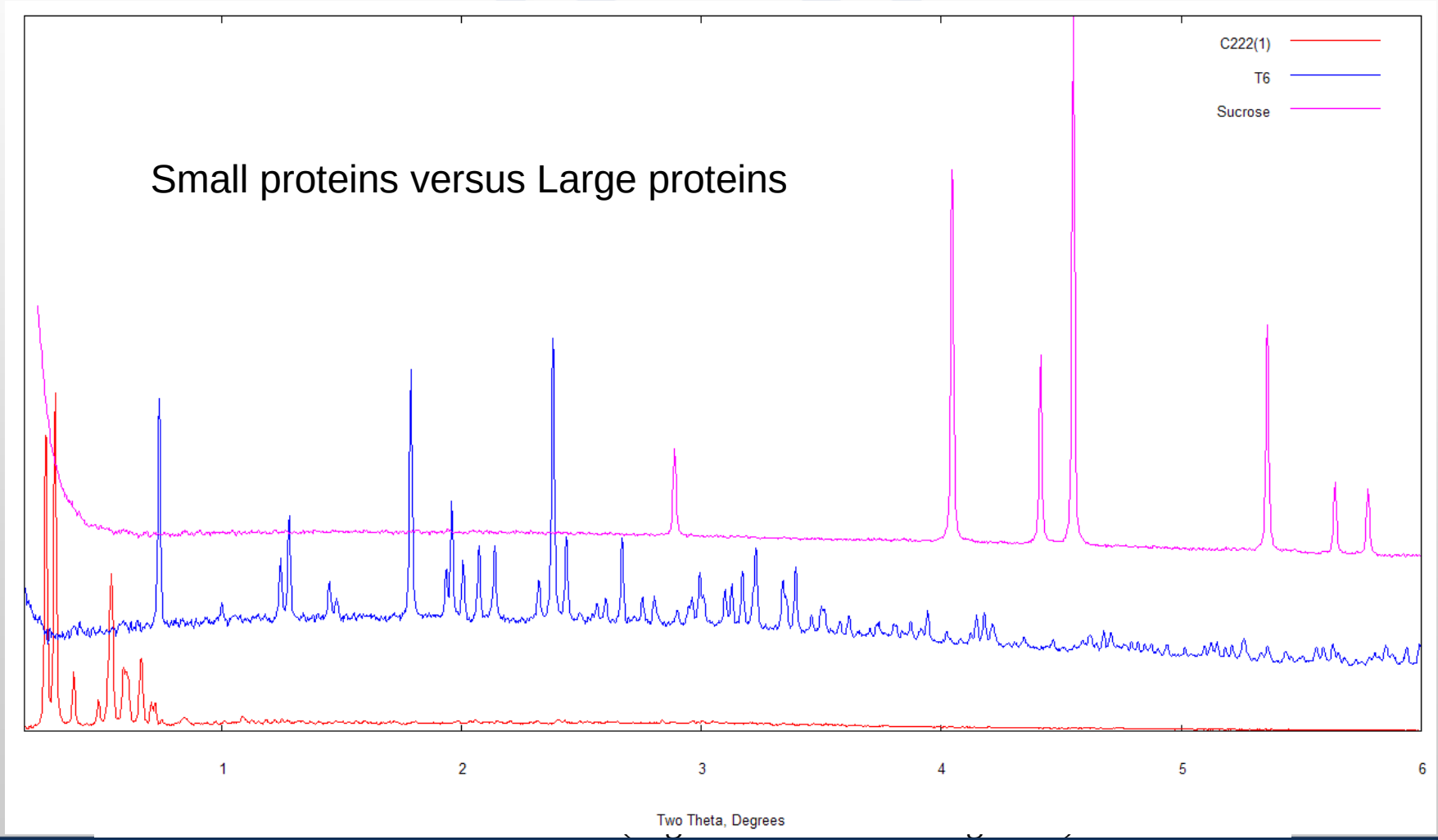


- Spots cover spheres in 3D reciprocal space

- 2D area detector takes a slice
- (on Ewald sphere)
- 1D powder scan measures distance from origin
- Proteins like that give *amazing* powder data

More demanding on the instrument!

- “Conventional” pattern together with protein



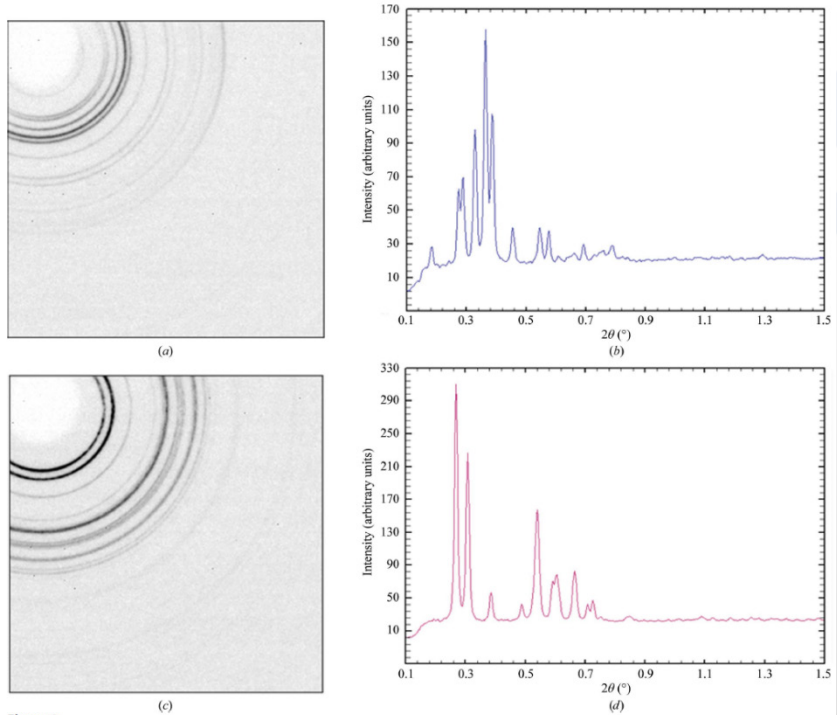


Figure 3
Area-detector images and integrated profiles obtained using the *Fit2D* software of human insulin cocrystallized with phenol at pH 5.75 and 6.78 corresponding to the (*a*, *b*) $P2_{1(\alpha)}$ and (*c*, *d*) $C222_1$ phases. The data were collected at RT [ID11; λ = 0.5339 (5) Å].

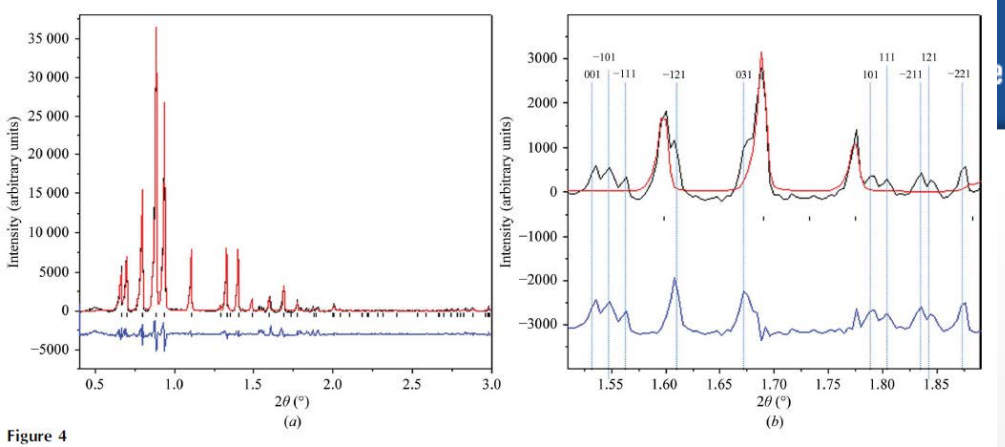


Figure 4

New polymorph of insulin identified from powder data

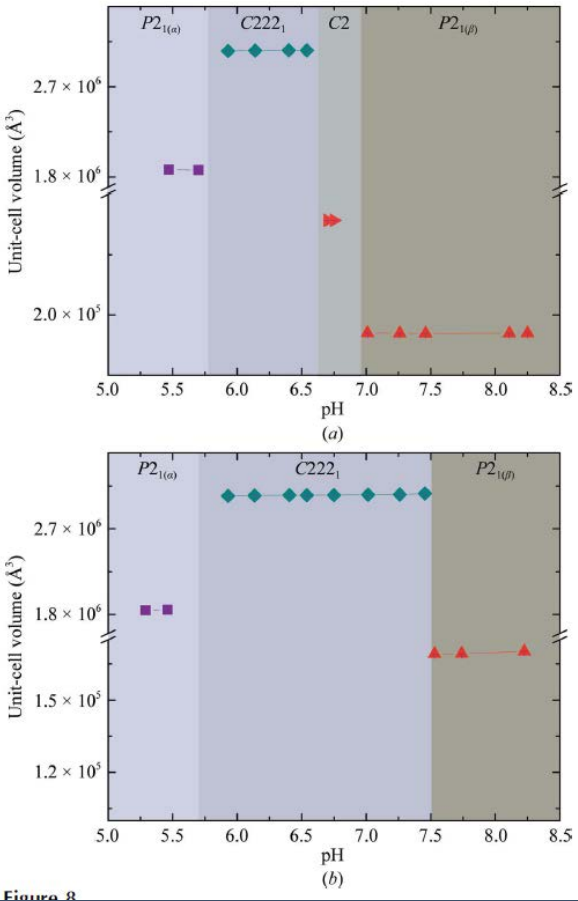


Figure 8

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural studies of human insulin cocrystallized with phenol or resorcinol *via* powder diffraction

Fotini Karavassili,[a] Anastasia E. Giannopoulou,[a] Eleni Kotsiliti,[a] Lisa Knight,[b] Mathias Norrman,[c] Gerd Schluckebier,[c] Lene Drube,[c] Andrew N. Fitch,[b] Jonathan P. Wright[b] and Irene Margiolaki[a]*

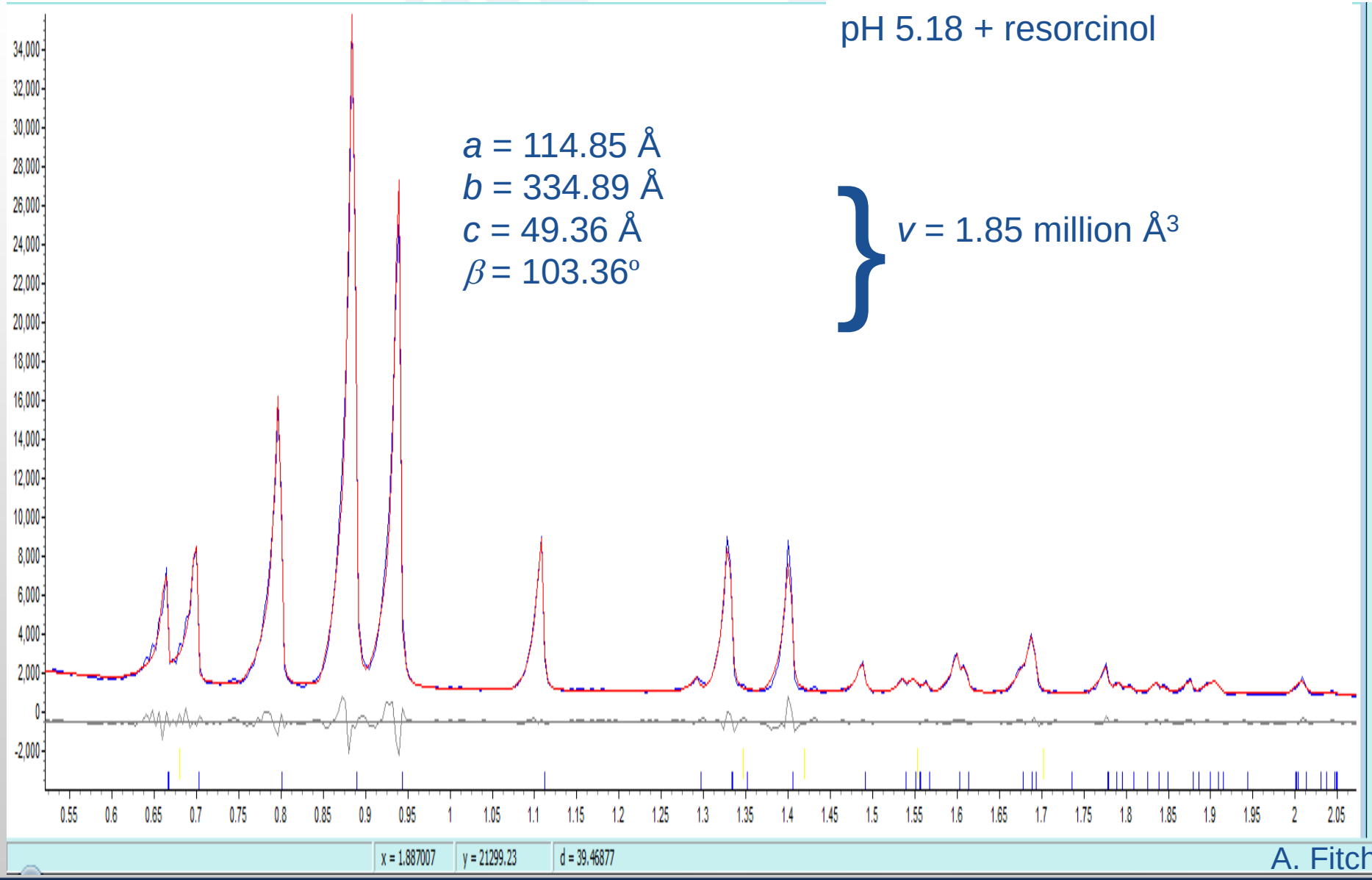
The effects of the ligands phenol and resorcinol on the crystallization of human insulin have been investigated as a function of pH. Powder diffraction data were used to characterize several distinct polymorphic forms. A previously unknown polymorph with monoclinic symmetry ($P2_1$) was identified for both types of ligand with similar characteristics [the unit-cell parameters for the insulin–resorcinol complex were a = 114.0228 (8), b = 335.43 (3), c = 49.211 (6) Å, β = 101.531 (8)°].

Received 1 August 2012
Accepted 14 September 2012

pH 5.18 + resorcinol

a = 114.85 Å
b = 334.89 Å
c = 49.36 Å
β = 103.36°

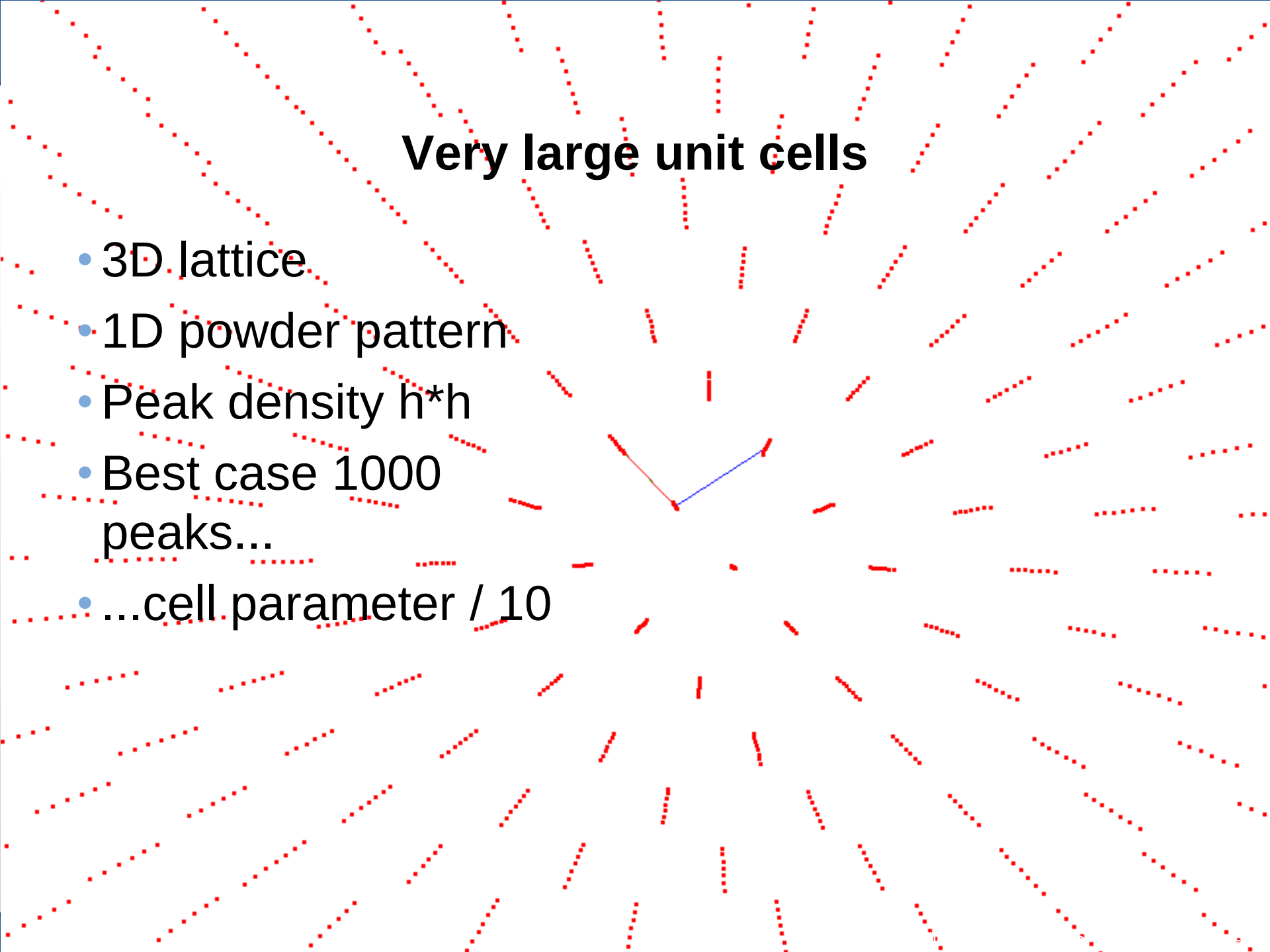
} v = 1.85 million Å^3



A. Fitch

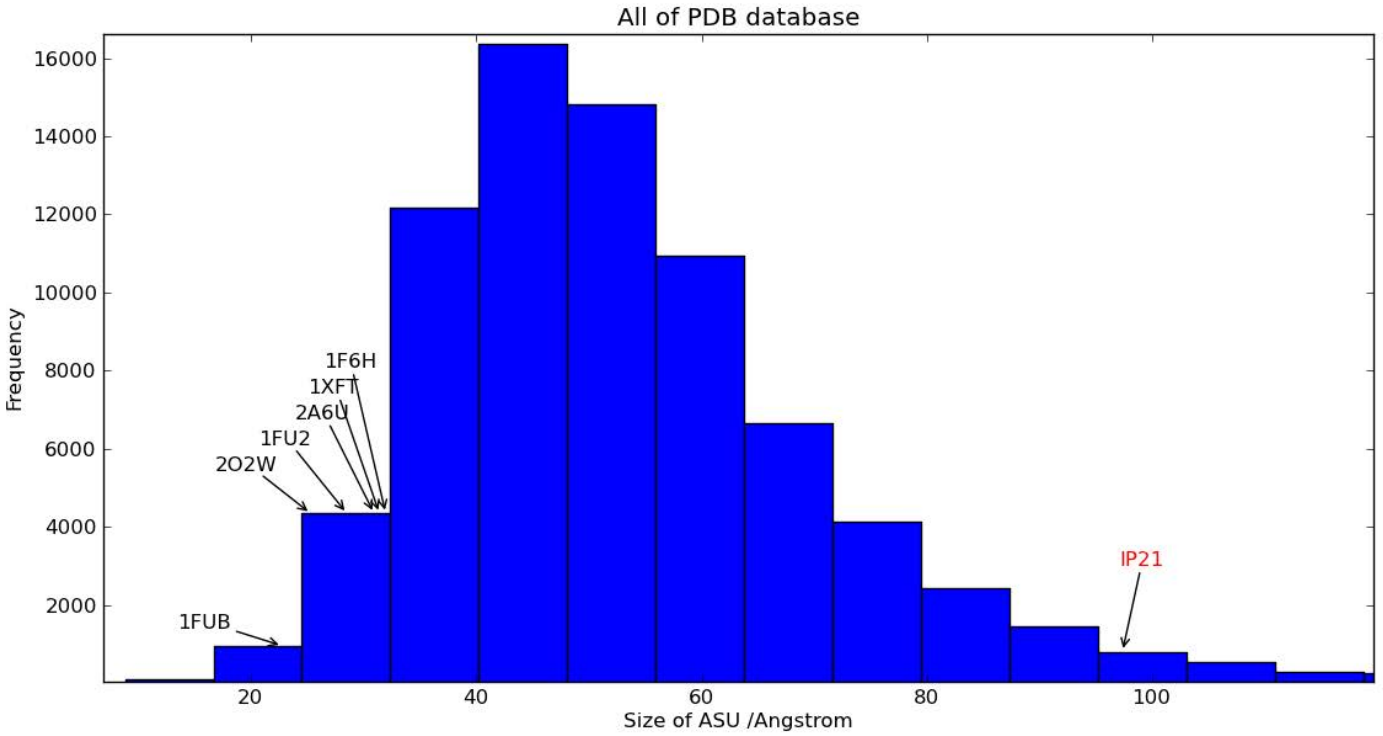
Very large unit cells

- 3D lattice
- 1D powder pattern
- Peak density h*h
- Best case 1000 peaks...
- ...cell parameter / 10



Refinements on relatively small proteins…

- 1FUB 22.8, T3R3 insulin
- 2O2W 25.3, ponsin
- 1FU2 28.52, T3R3'
- 2A6U 30.94, tetragonal lysozyme
- 1XFT 31.41, turkey lysozyme
- 1F6H 31.97, myoglobin



Polymorphism of microcrystalline urate oxidase from *Aspergillus flavus*

Ines Collings,[a] Yves Watier,[a] Marion Giffard,[b]* Sotonye Dagogo,[a] Richard Kahn,[c] Francoise Bonneté,[b] Jonathan P. Wright,[a] Andrew N. Fitch[a] and Irene Margiolaki[a,d]*

- polymorphism in protein drug crystals

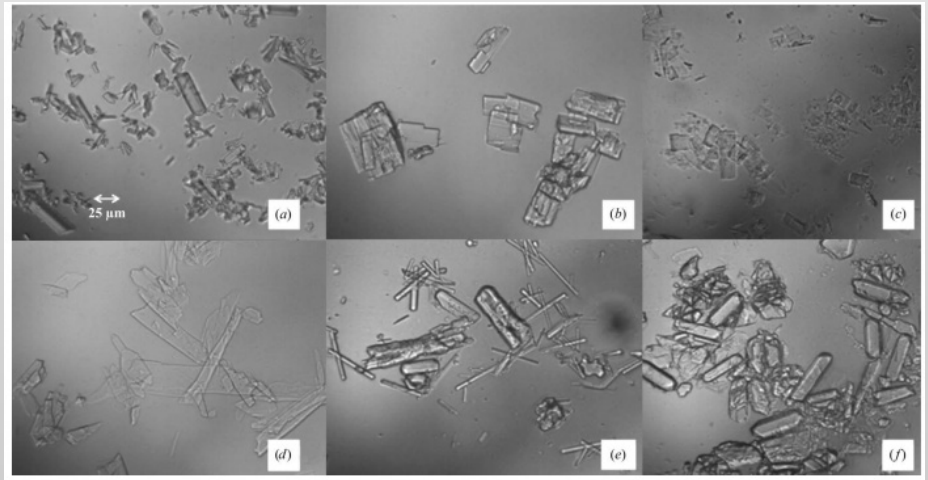
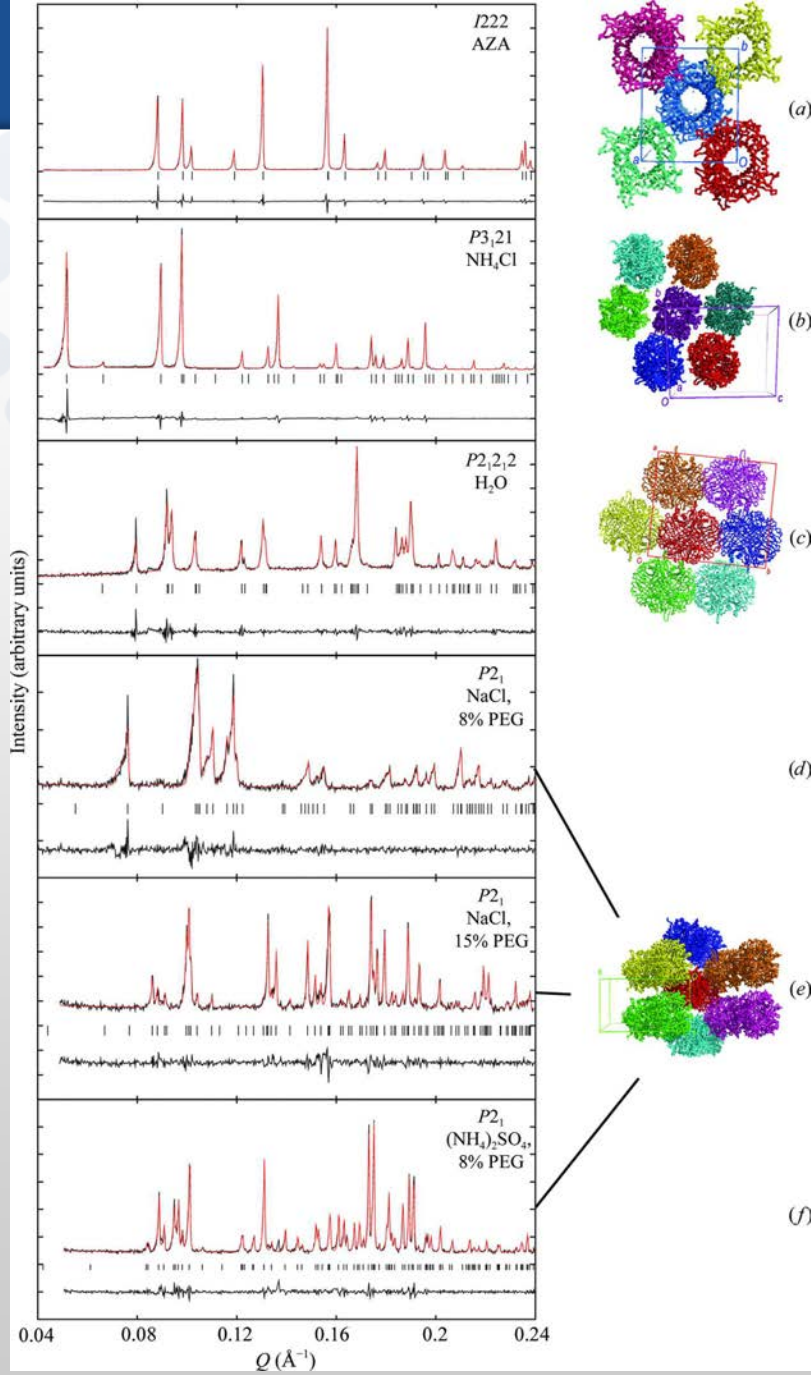


Figure 1
Optical microscopy images of Uox microcrystals prepared under different conditions. (*a*) Ligand-free Uox with NH_4Cl and 15% PEG 8000, (*b*) ligand-free Uox with NaCl and 15% PEG 8000, (*c*) ligand-free Uox with $(NH_4)_2SO_4$ and 15% PEG 8000, (*d*) ligand-free Uox with 10% PEG, (*e*) ligand-free Uox with KCl and 10% PEG 8000, (*f*) Uox complexed to AZA with NaCl and 10% PEG 8000.

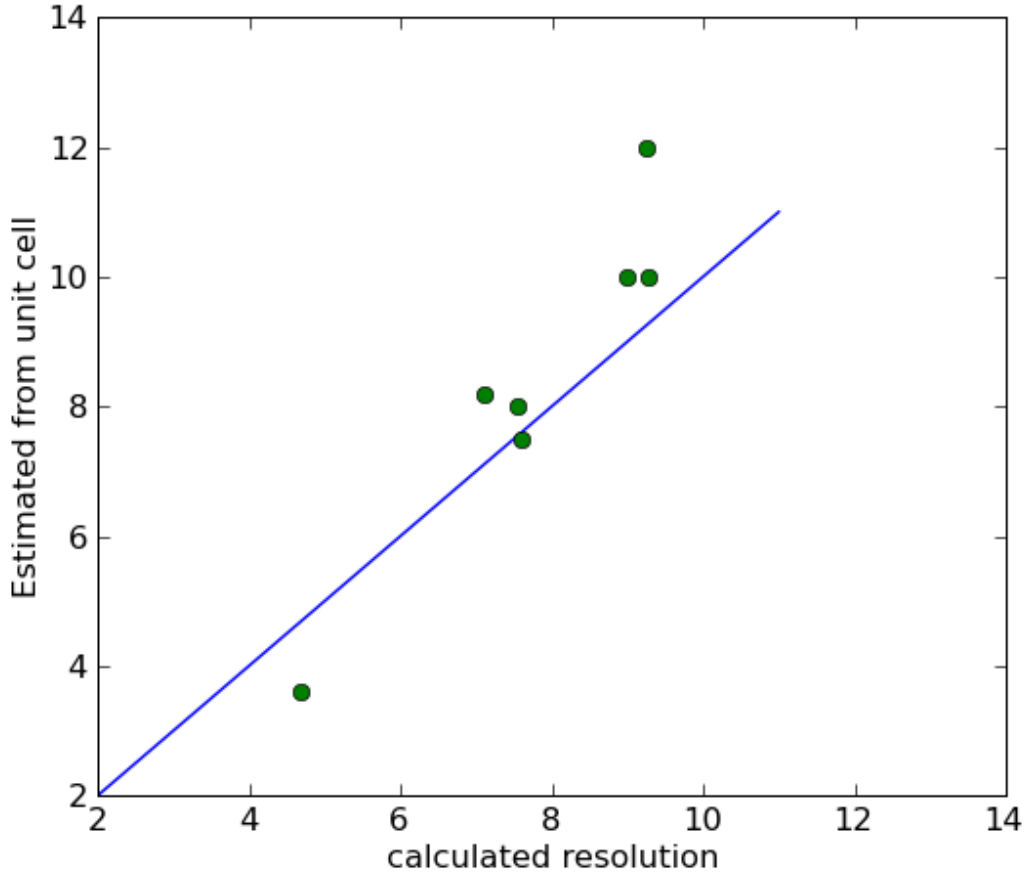


Urate Oxidase

able 1

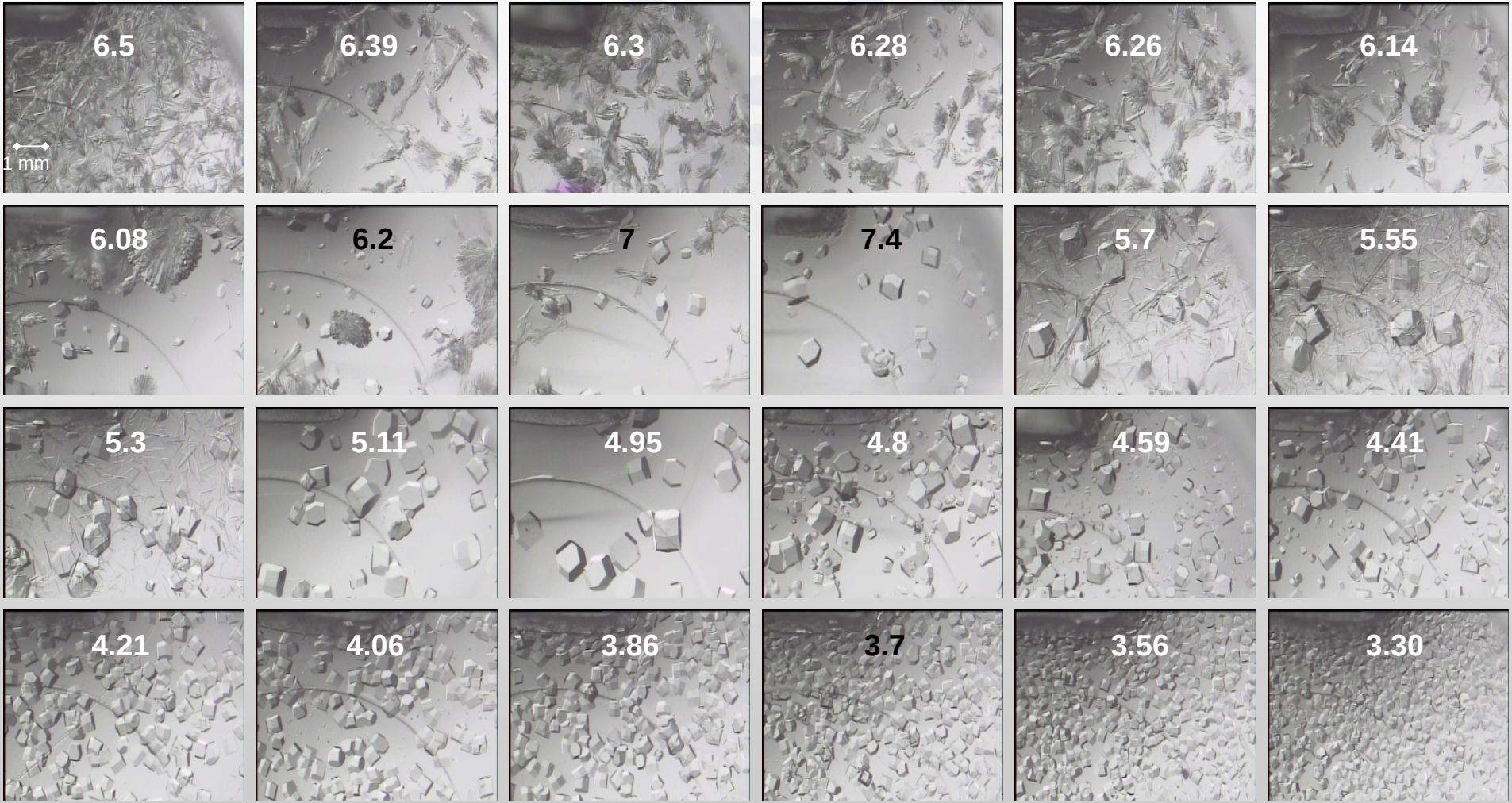
etails of sample crystallization, data collection and Le Bail analysis (Le

ox crystallized with	NH_4Cl	NaCl	Na				
EG concentration (%)	**15**	**15**	**8**				
ata collection							
Wavelength (Å)	1.30000 (6)	1.30000 (6)	1.:				
Exposure time per scan (s)	60	60	60				
eBail refinement							
Space group	$P3_121$	$P2_1$	$P2$				
Unit-cell parameters							
a (Å)	140.4002 (9)	81.8712 (9)	82				
b (Å)	140.4002 (9)	124.7628 (13)	14				
c (Å)	151.1053 (13)	142.9454 (15)	13				
α (°)	90	90	90				
β (°)	90	93.7280 (6)	92				
γ (°)	120	90	90				
Volume ($Å^3$)	2579560 (40)	1457020 (30)	1577490 (70)	1595740 (30)	1433510 (30)	2621400 (100)	815340 (5)
Matthews coefficient ($Å^3$ Da^{-1})	3.14	2.66	2.88	2.90	2.60	3.19	2.95
Solvent content (%)	60.8	53.8	57.3	57.6	52.7	61.4	58.3
No. of monomers per ASU	4	8	8	8	4	4	1
Resolution range (Å)	121.6-8.0	142.6-10.0	134.8-12.0	149.0-10.0	95.3-8.2	121.5-7.5	106.4-3.6

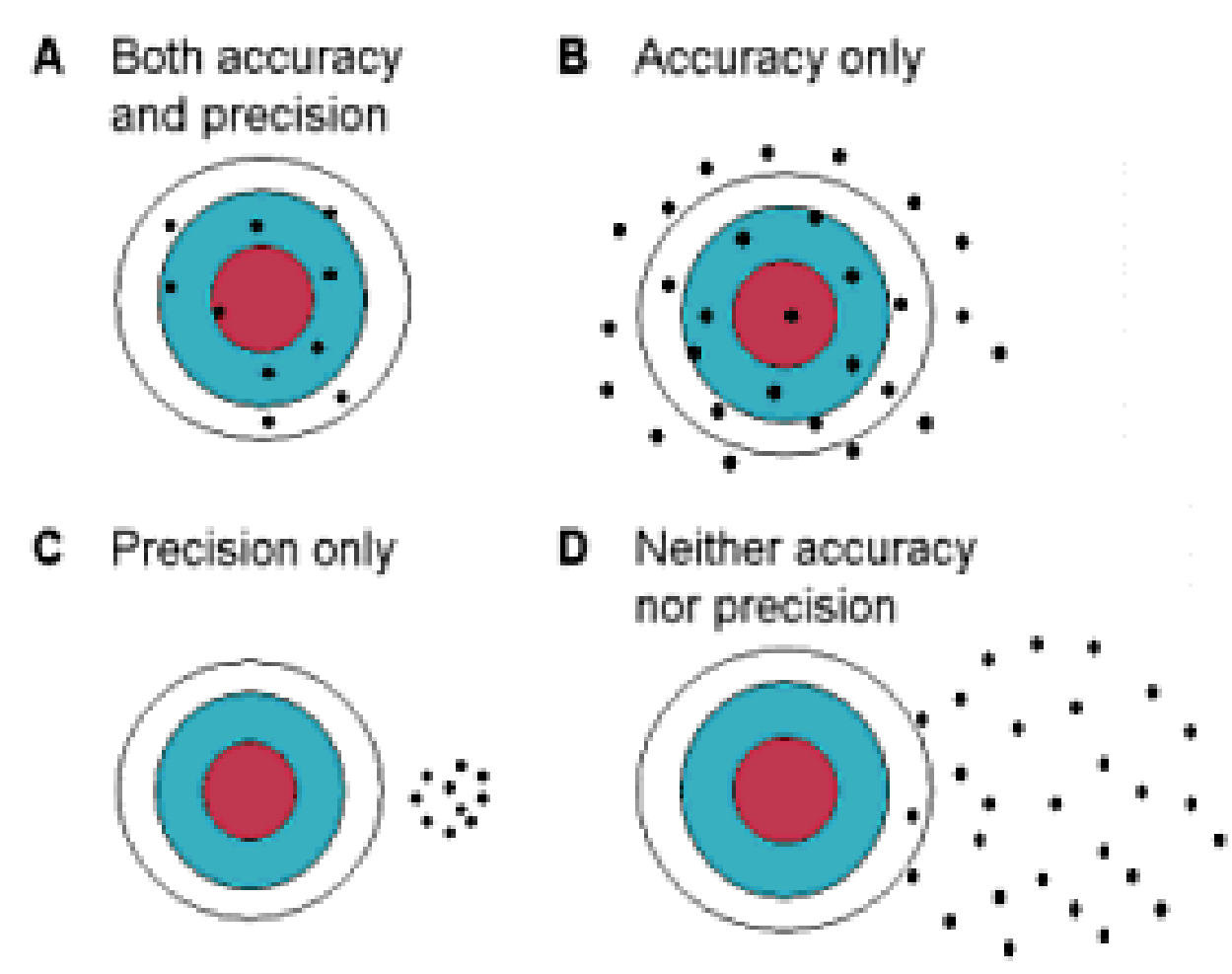


Acta Cryst. (2010). D66, 539-548 [doi:10.1107/S0907444910005354]
Polymorphism of microcrystalline urate oxidase from Aspergillus flavus
I. Collings, Y. Watier, M. Giffard, S. Dagogo, R. Kahn, F. Bonneté, J. P. Wright, A. N. Fitch and I. Margiolaki

Lysozyme crystallised at RT, vary pH



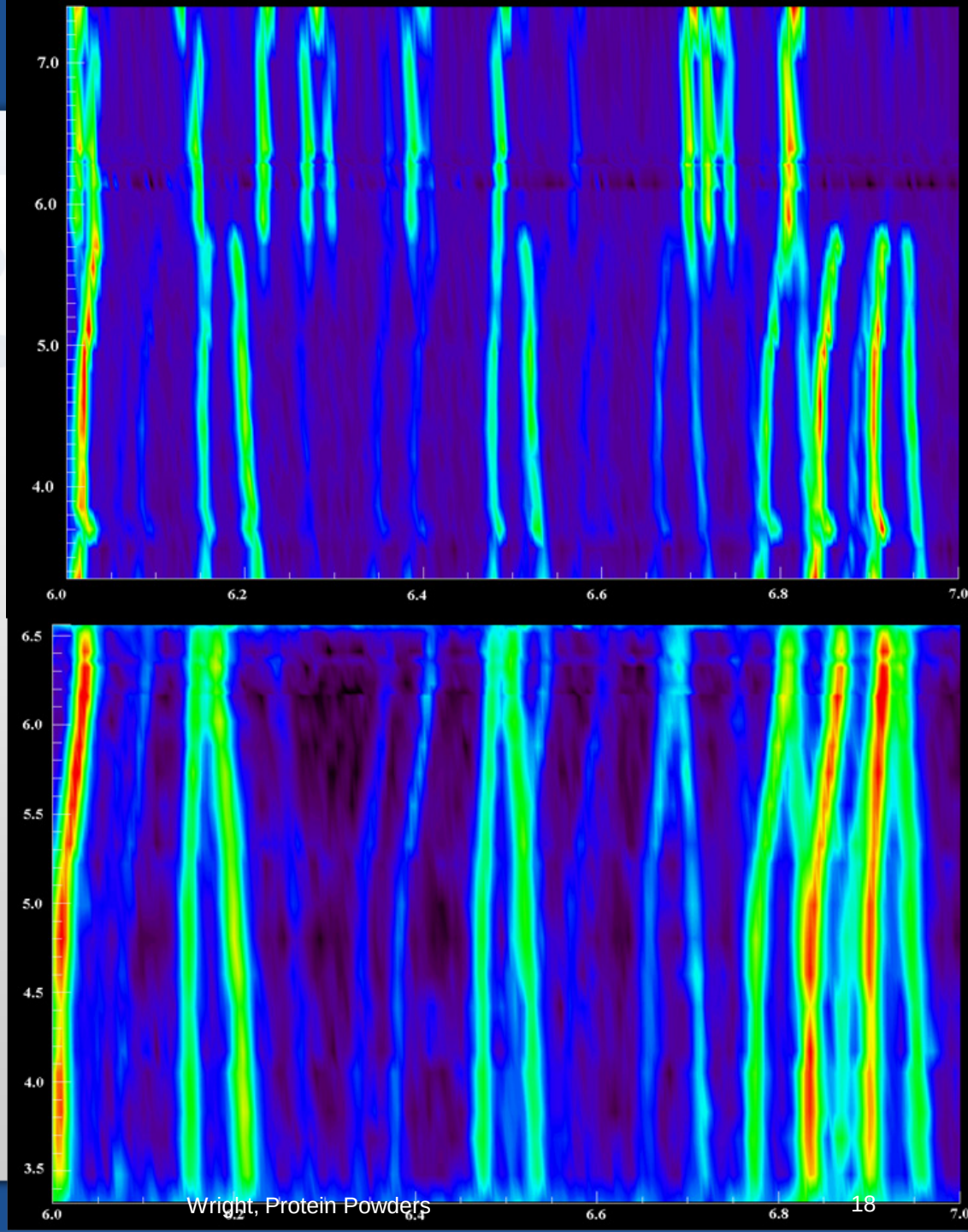
Accuracy of unit cell parameters?



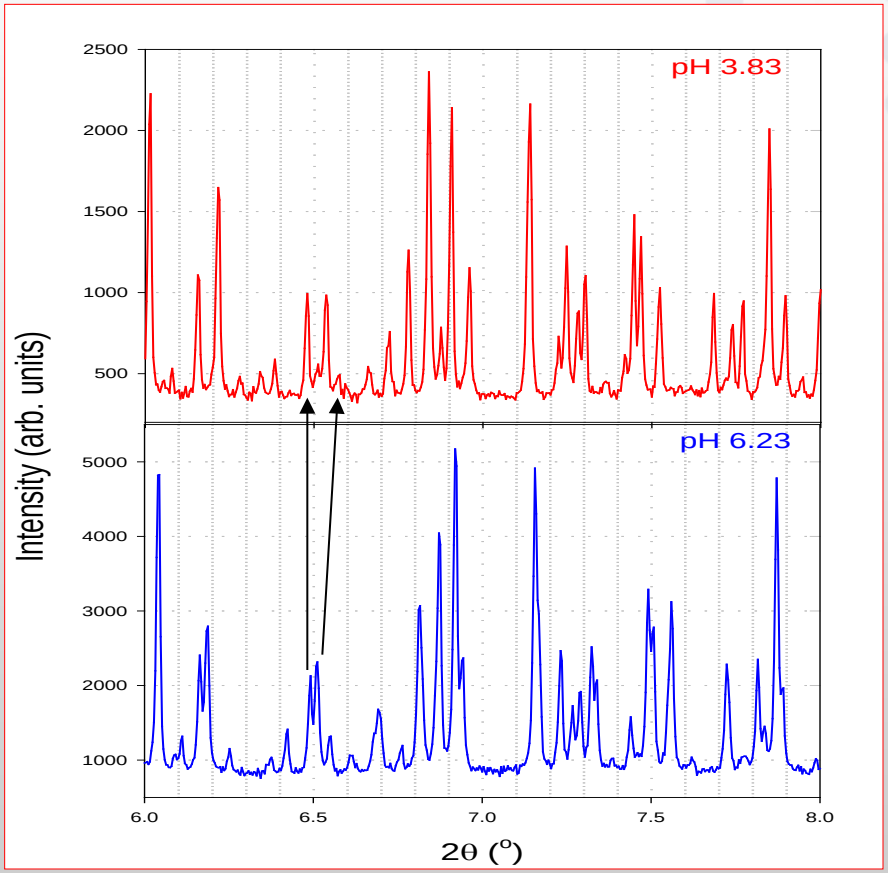
S. Basso et al.,
Acta Cryst. D61,
1612-1625 (2005)

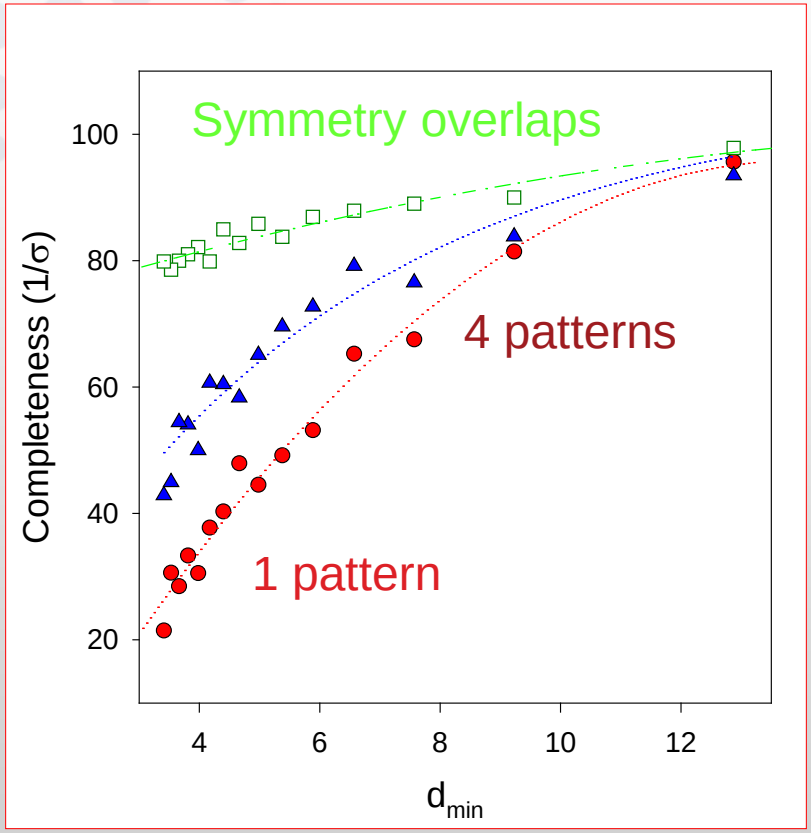
- Upper panel: RT
- Lower panel: 4°C
- Orthorhomic – tetragonal phase transition
- Variation of unit cell exploited for refinement

- Controlled:
  - pH
  - Temperature

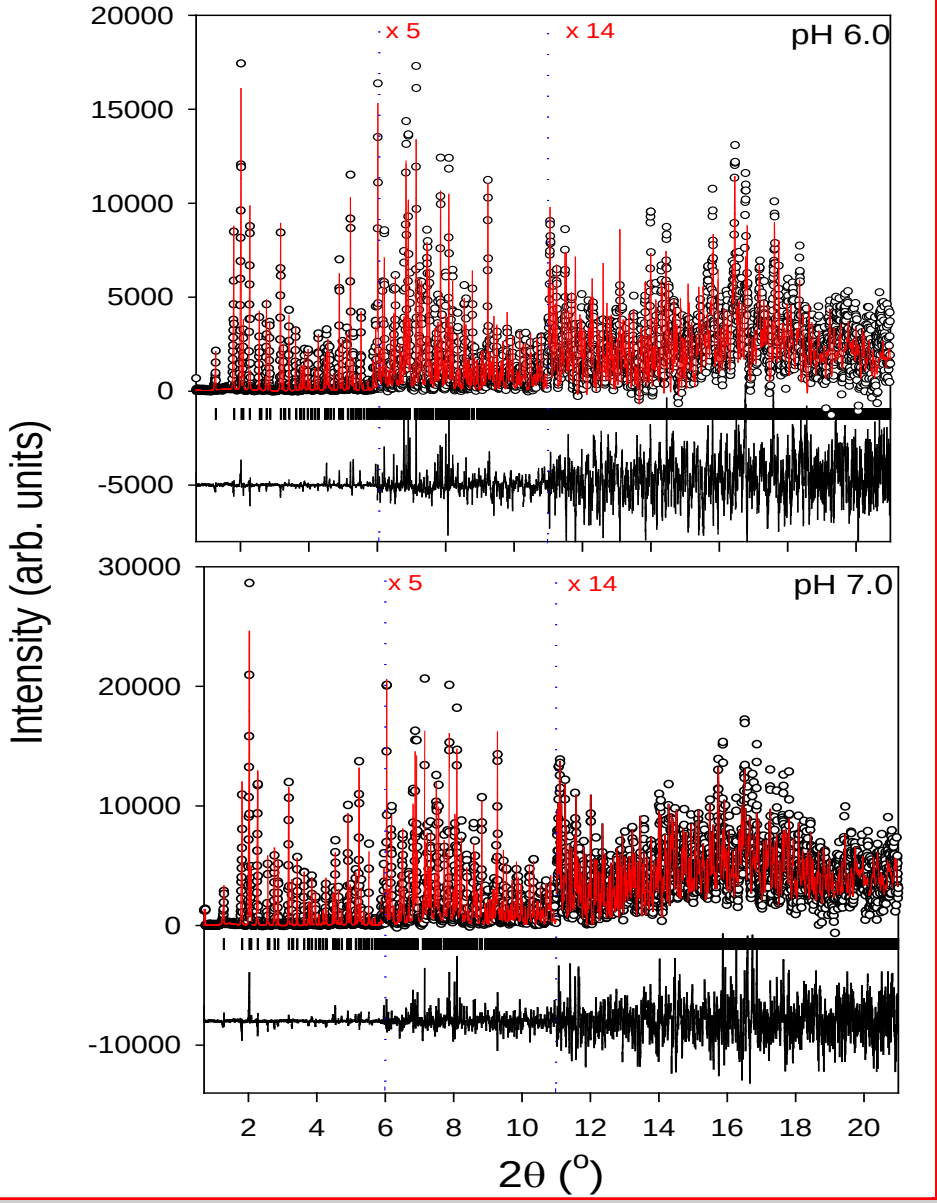


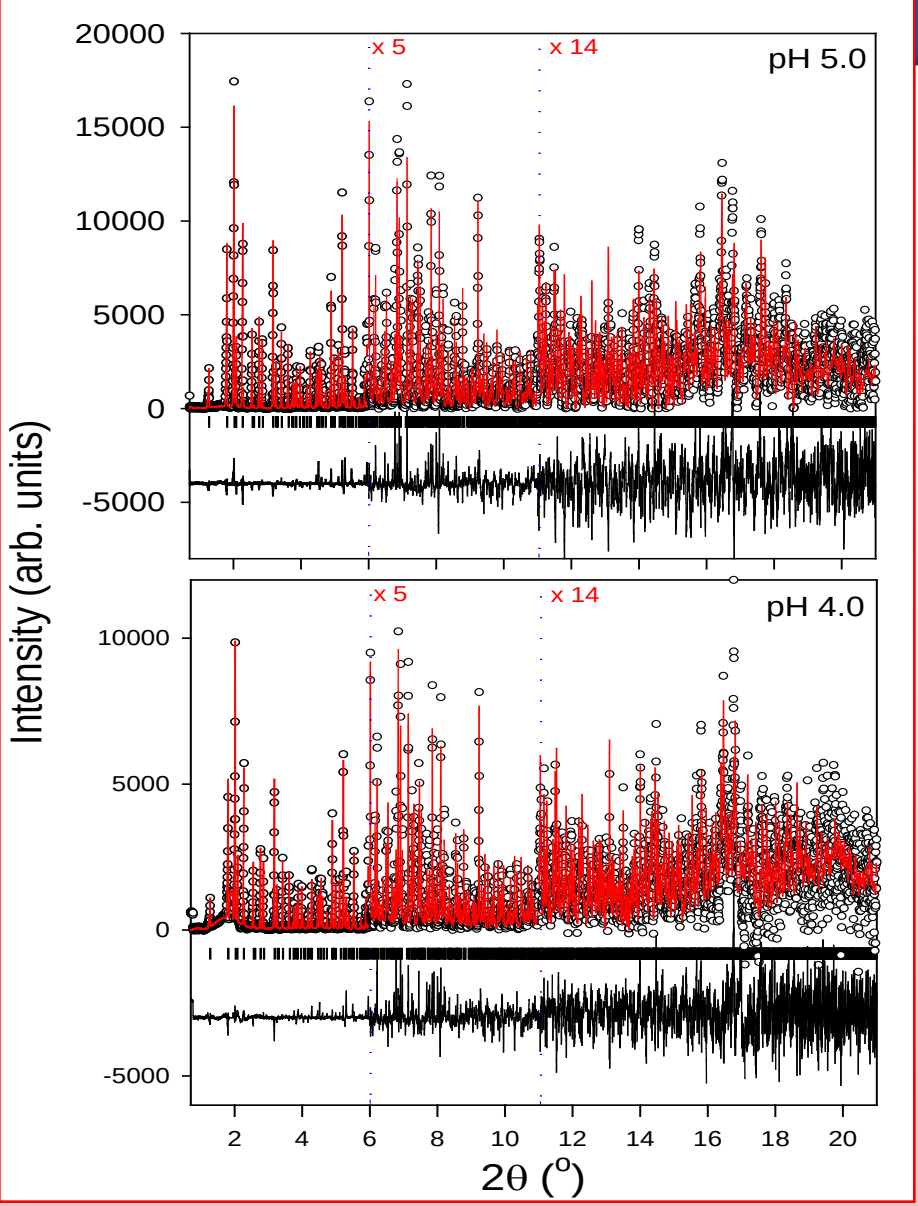
Untangling overlapped peaks





An effective completeness for combined data sets is proposed as the fraction of "peaks" having I/σ(I) greater than some threshold.



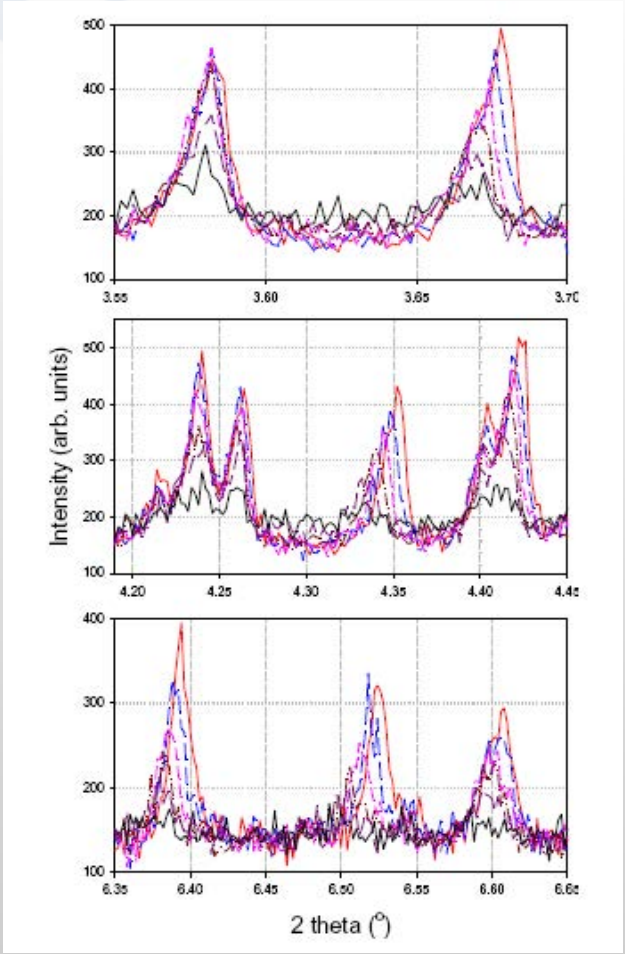


Acta Cryst. (2005). D61, 1612-1625

S. Basso, A. N. Fitch, G. C. Fox, **I. Margiolaki** and J. P. Wright

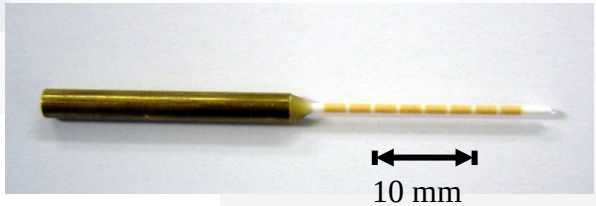
Radiation Damage…

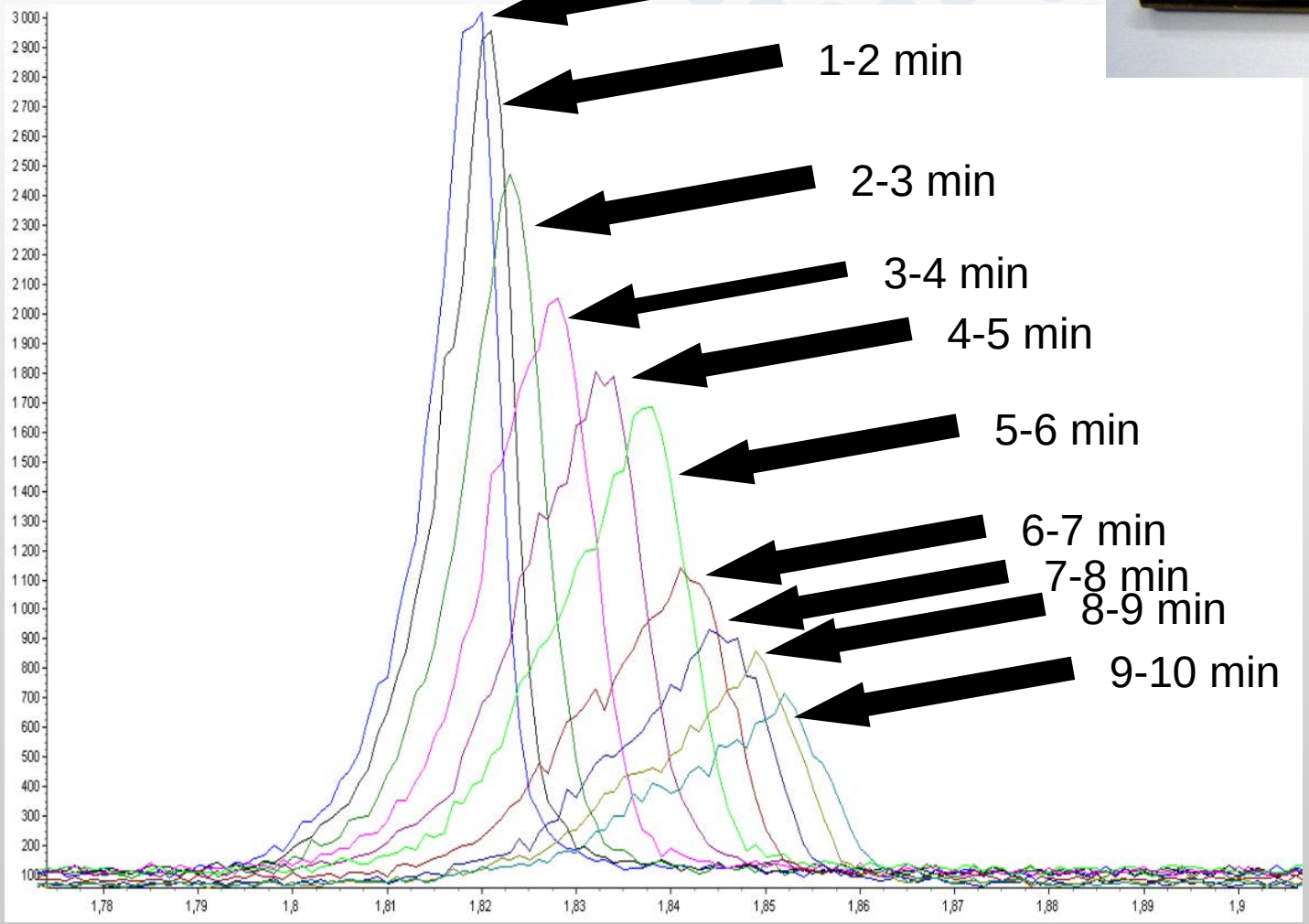
- Depends on absorbed dose
- Large body of literature for single crystals
- Peaks shift, disappear
- About 80X worse at RT than cryo



(Ponsin, ID31)

Effect of radiation damage at Room temperature

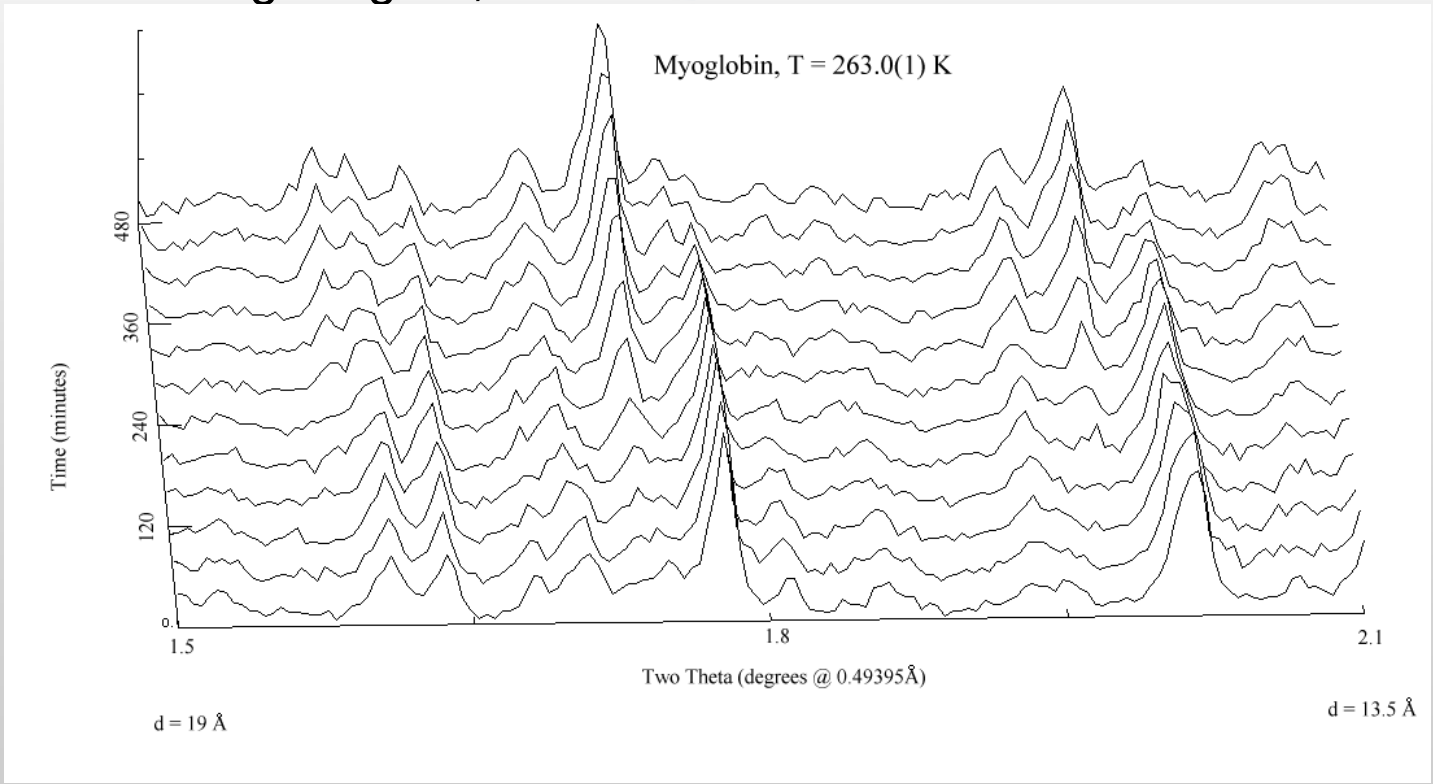




Courtesy: Yves Watier

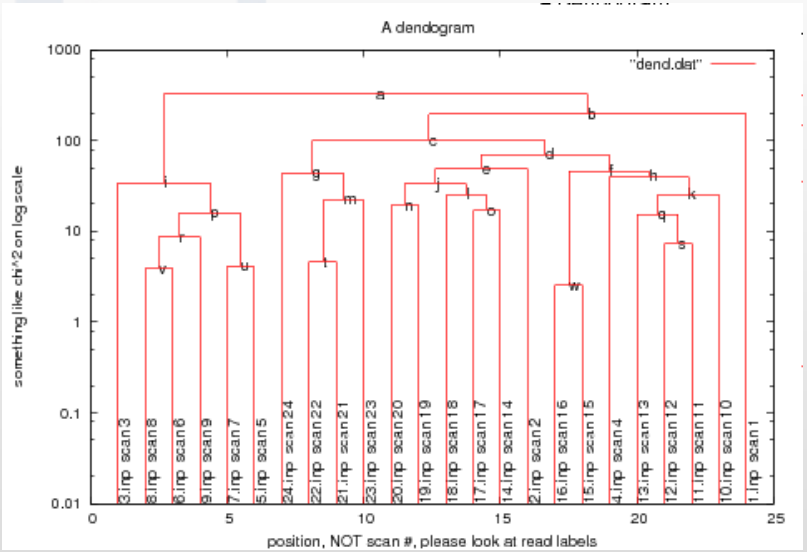
Slow kinetics – no radiation damage

- Myoglobin Phase I>II kinetics at 263 K (fast cool followed by hold at constant temperature)
- ESRF bending magnet, BM16



Merging scans with radiation damage

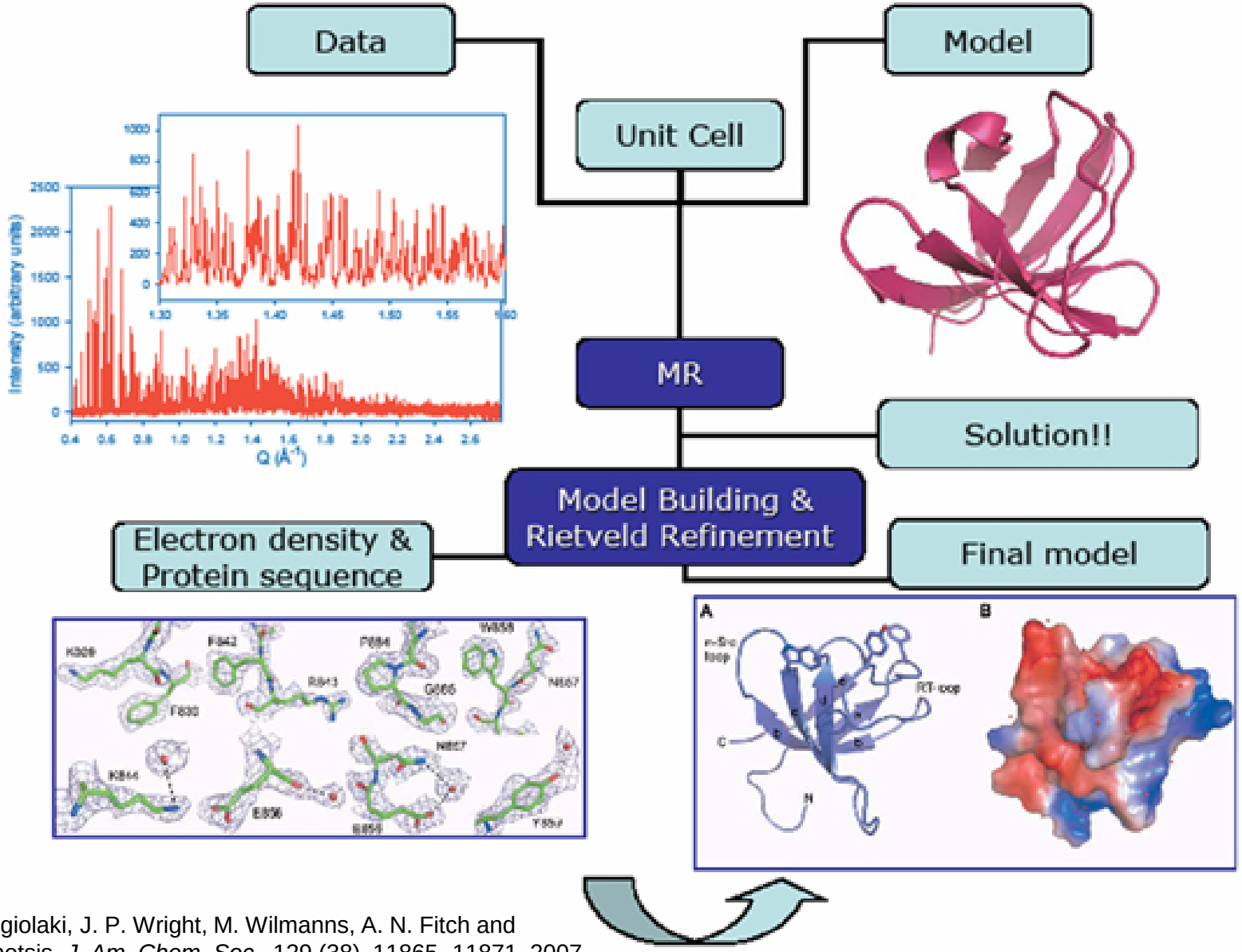
- Anisotropic peak shifts with radiation damage
- Position on capillary
- Classification using pycluster (cluster.py script)
- Sum up similar scans (typically chi2 statistic, CC also useful)



$$\frac{(Y_1 - Y_2)}{\mathrm{sqrt}(e_1^2 + e_2^2)}$$

Sum of :
Difference/error

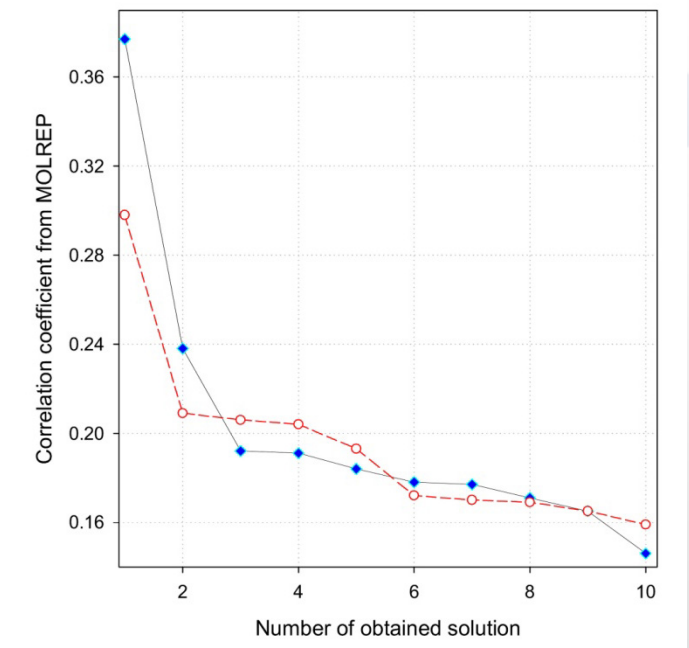
Structure solution of SH3.2, ponsin



I. Margiolaki, J. P. Wright, M. Wilmanns, A. N. Fitch and N. Pinotsis, *J. Am. Chem. Soc.*, 129 (38), 11865 -11871, 2007

Molecular replacement

Solution Score for two different models



Two models were tested
Solution always obvious

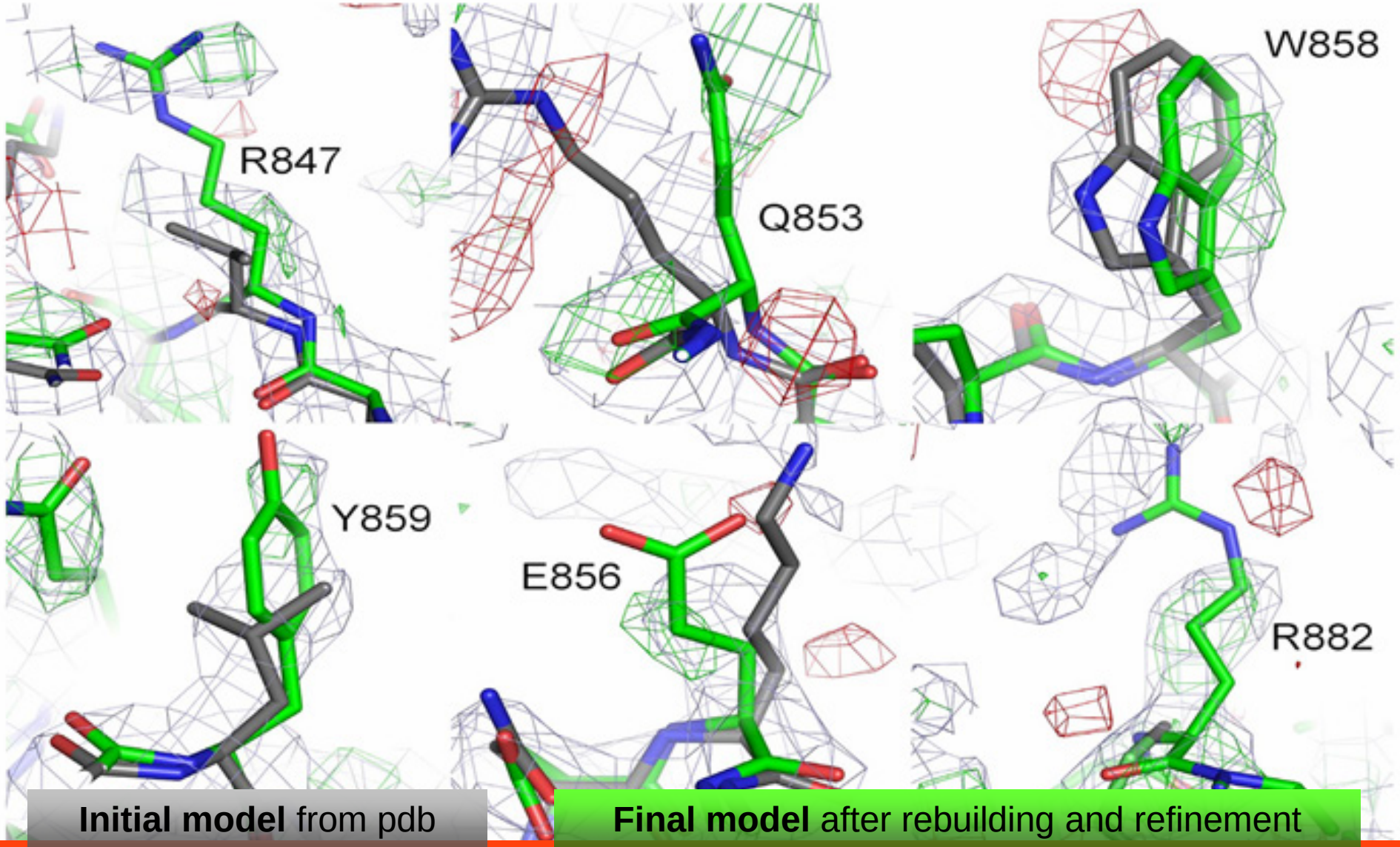
MOLREP searches orientation and position of the model in the unit cell

Molrep results:

```
--- Summary ---
```

S_ RF TF		theta	phi	chi	tx	ty	tz	TFcnt	Rfac	Scor
S___1__1	1	58.07	144.99	102.00	0.161	0.266	0.181	3.47	0.527	0.377
S___6_15	2	145.90	144.44	92.50	0.361	0.166	0.366	5.14	0.595	0.238
S___2_10	3	125.77	-81.54	96.49	0.212	0.061	0.165	2.29	0.627	0.192
S___4__7	4	51.52	58.03	125.28	0.456	0.267	0.351	3.25	0.612	0.191
S___7__5	5	132.09	-179.53	72.05	0.167	0.366	0.214	1.40	0.621	0.184
S___8_10	6	175.99	-179.51	84.91	0.406	0.101	0.391	2.09	0.615	0.178
S__10_13	7	71.94	-151.69	19.36	0.378	0.216	0.318	1.78	0.625	0.177
S___5__5	8	62.17	-139.62	106.19	0.279	0.485	0.198	2.21	0.624	0.171
S___9_11	9	80.99	-179.11	79.75	0.284	0.452	0.132	1.62	0.616	0.165
S___3_13	10	30.37	76.91	178.57	0.346	0.253	0.137	2.17	0.627	0.146

Initial electron density maps



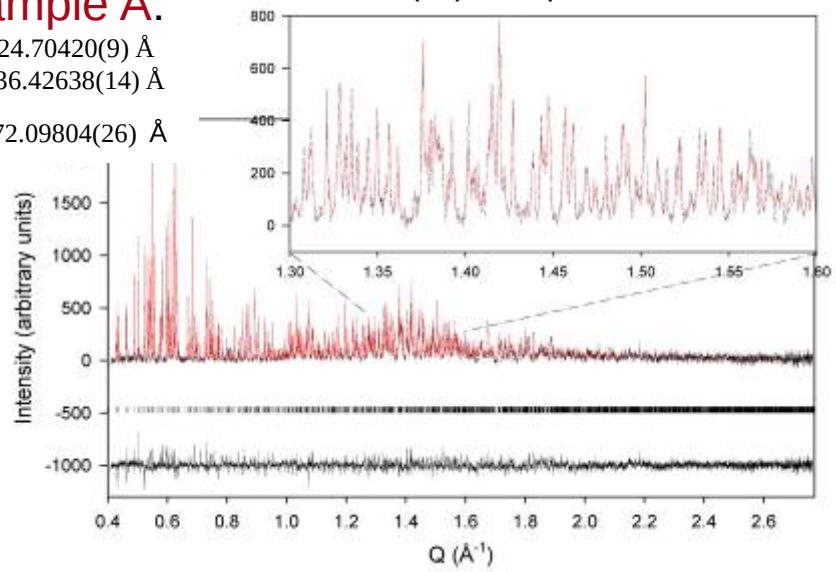
2Fo-1Fc (blue at 1σ) and 1Fo-Fc (red at -2.5σ and green at 2.5σ) electron density maps, as determined directly after the molecular replacement. The residues represented in grey stick carbon atoms correspond to the molecular replacement model used for the calculation of the maps, while the residues in green color carbon atom sticks represent the final refined model.

4 dataset restrained refinement, ponsin

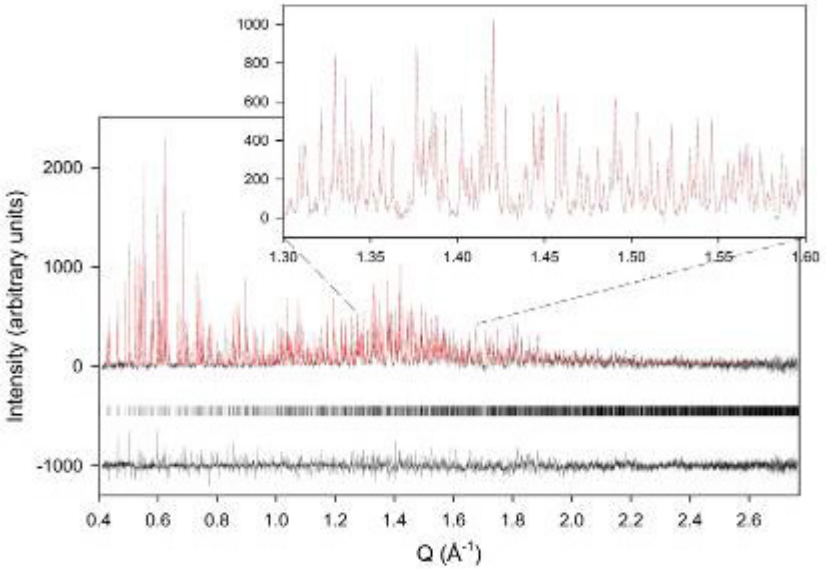
Sample A:

a=24.70420(9) Å
b= 36.42638(14) Å
c= 72.09804(26) Å

λ= 1.252481(32) Å, exposure time: **2 min**.



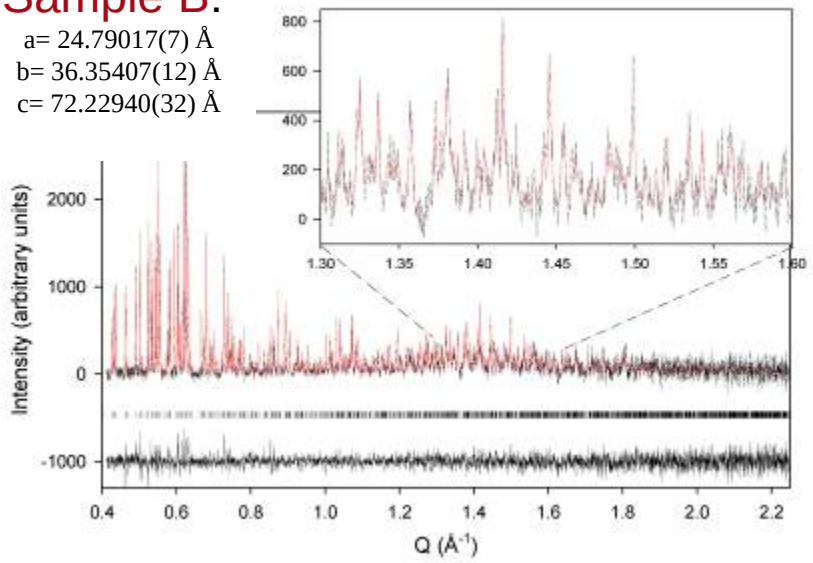
λ= 1.252481(32) Å, exposure time: **4 min**.



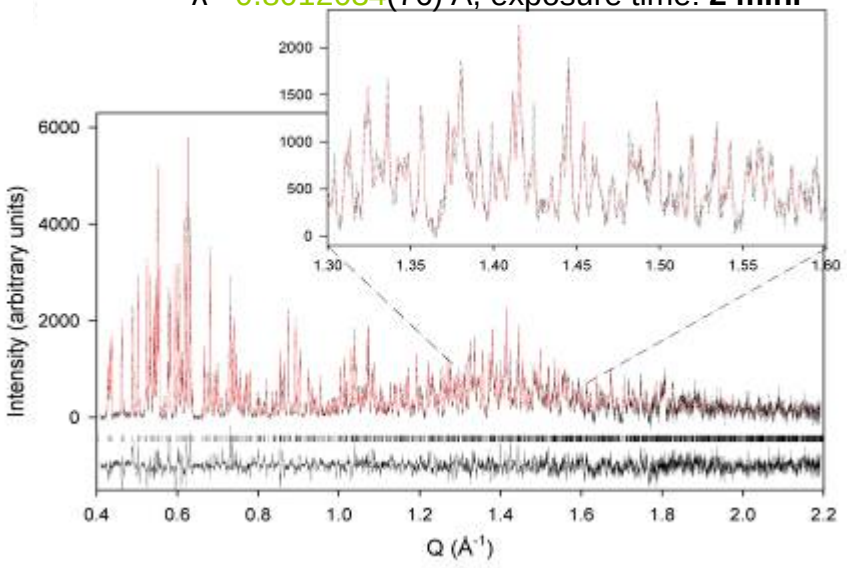
Sample B:

a= 24.79017(7) Å
b= 36.35407(12) Å
c= 72.22940(32) Å

λ= 1.251209(40) Å, exposure time: **2 min.**

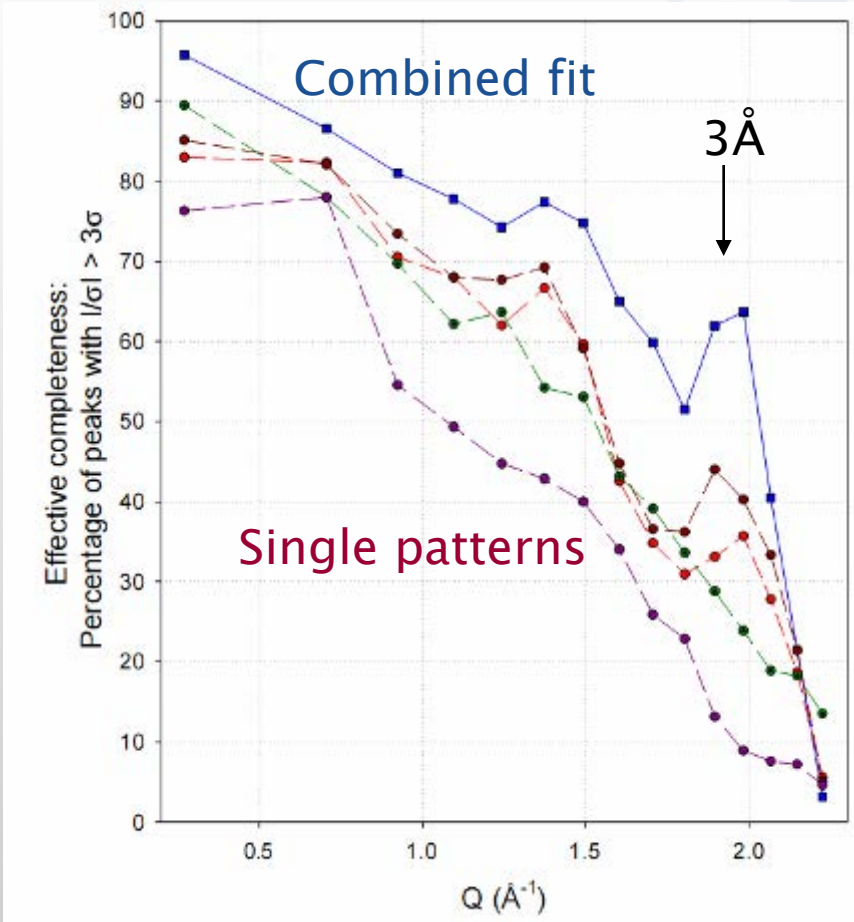


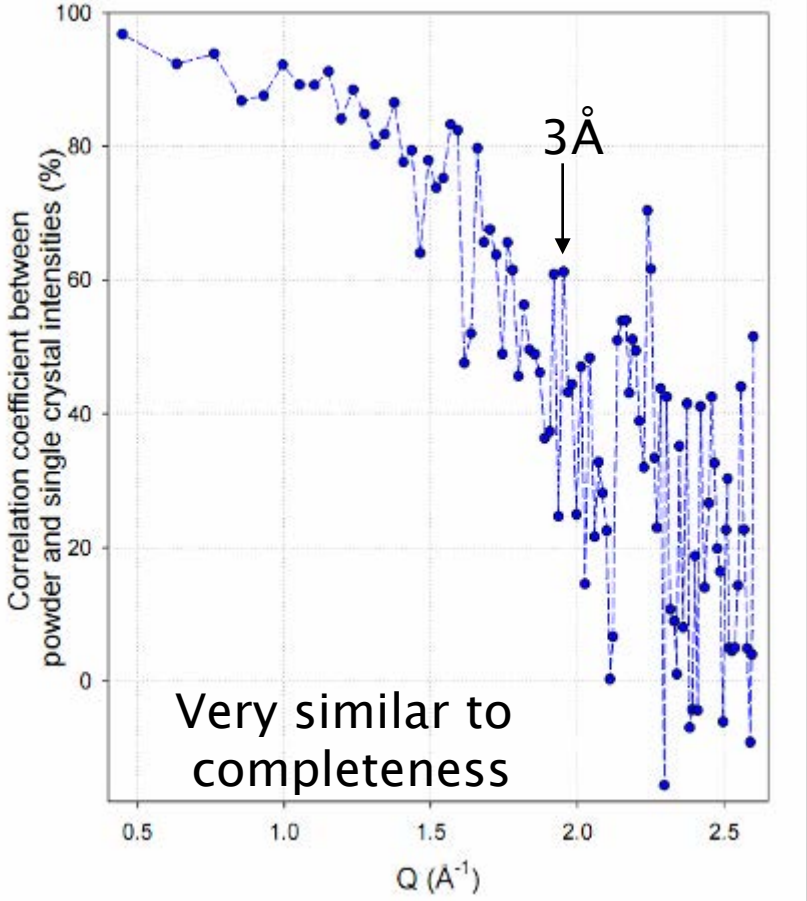
λ= 0.8012034(76) Å, exposure time: **2 min.**



Quality of extracted intensities

Completeness & correlations to SX



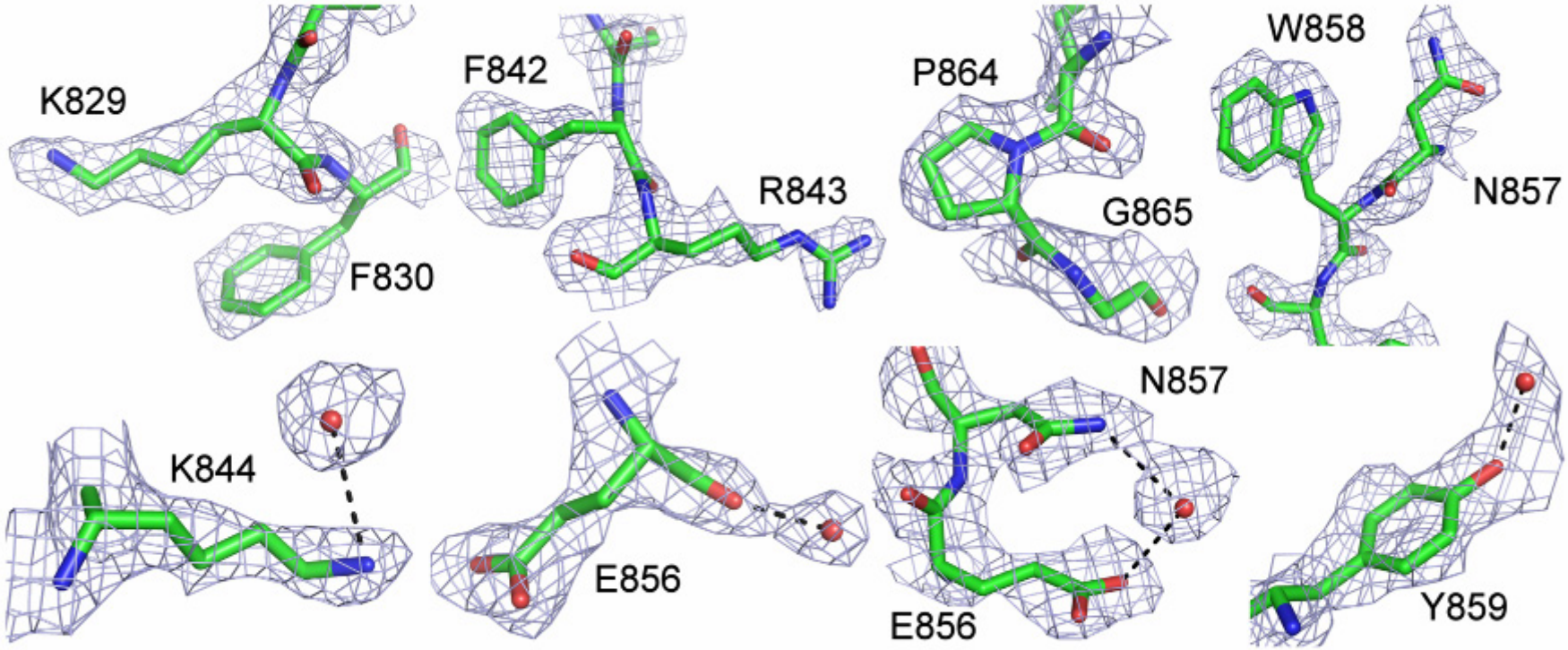


Completeness from eigenvalues of matrix of I/sigma, eg, Pawley weight matrix multiplied by intensities.

The Second SH3 domain of Ponsin

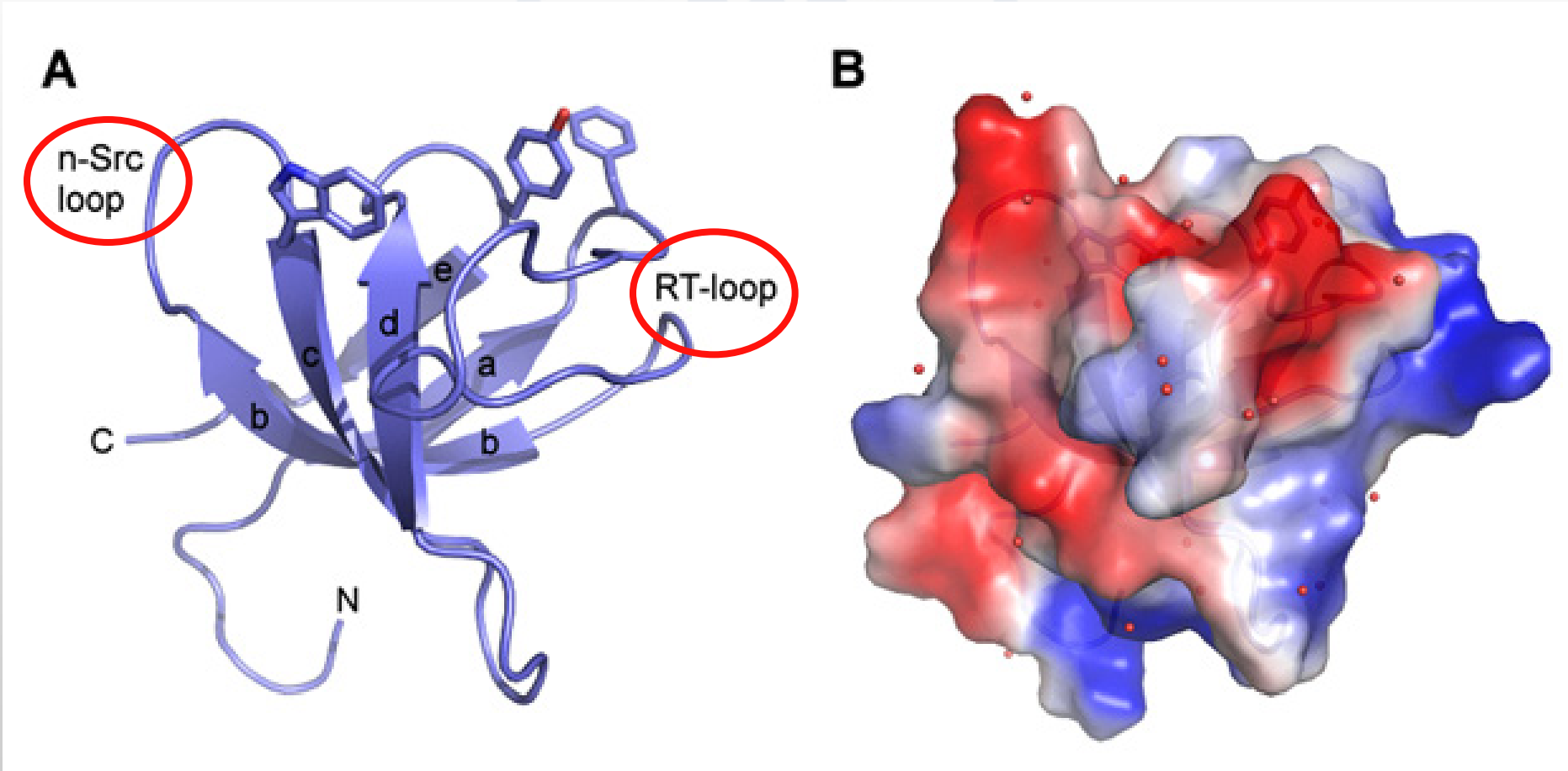
Selected regions of the final refined structural model in stick representation and the corresponding total omit map contoured at 1σ.

544 protein atoms and 36 water molecules were identified



SH3.2: The final model

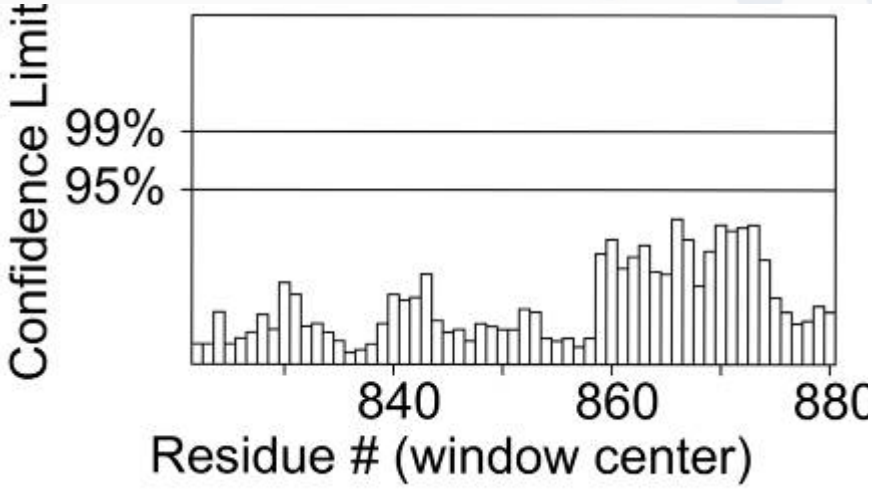
Powder-diffraction structure of the ponsin SH3.2 domain. (A) Ribbon representation of the SH3.2 indicating the secondary structure elements of the domain. The main hydrophobic residues of the binding interface as well as the positions of the n-Src and RT loops are indicated. (B) Electrostatic potential representation of domain identifying additionally the water molecules as red spheres.



Structure Validation

ERRAT:

http://nihserver.mbi.ucla.edu/ERRATv2/

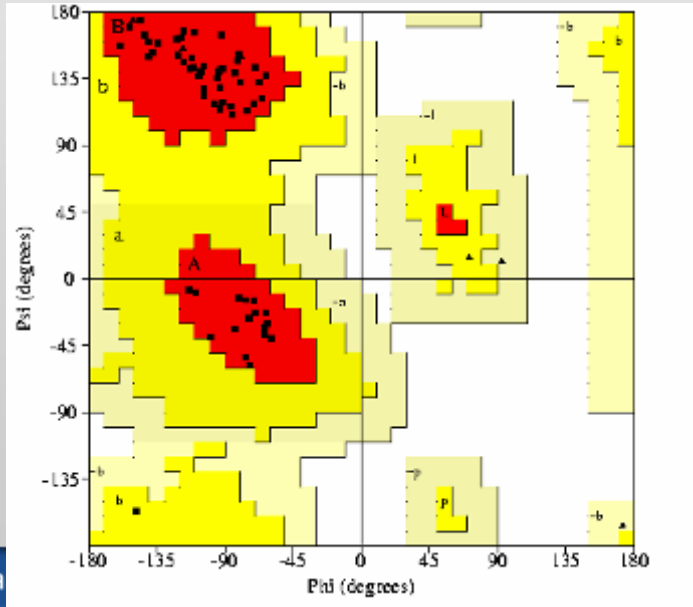


PROCHECK:

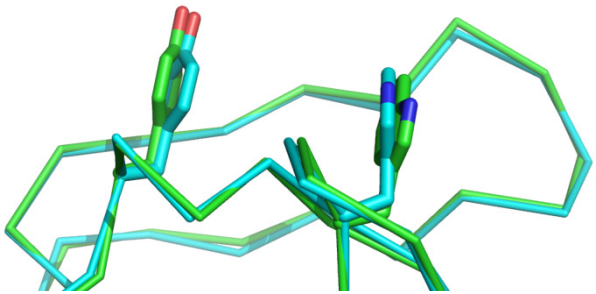
http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html

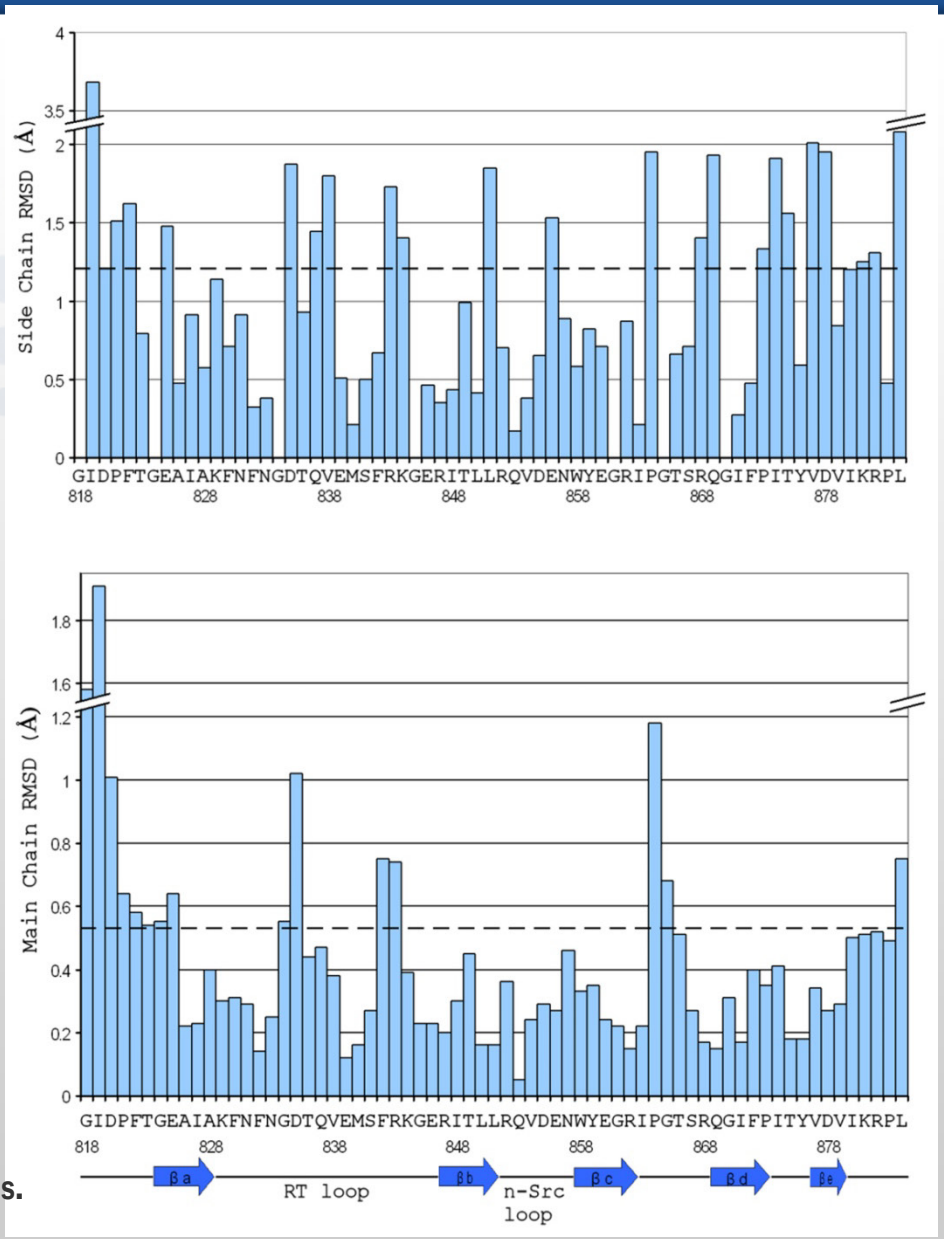
```
 +----------<<< P R O C H E C K    S U M M A R Y >>>----------+
 | final2   2.8                                        67 residues |
 | Ramachandran plot:   98.2% core    1.8% allow     .0% gener     .0% disall |
 | Gly & Pro Ramach:     0 labelled residues (out of  10)                    |
 | Chi1-chi2 plots:      0 labelled residues (out of  43)                    |
 | Main-chain params:    6 better     0 inside      0 worse                 |
 | Side-chain params:    5 better     0 inside      0 worse                 |
 | Residue properties: Max.deviation:     2.7             Bad contacts:    0 |
+|                     Bond len/angle:    4.8    Morris et al class:  1  1  3 |
 | G-factors           Dihedrals:   -.24  Covalent:   -.17    Overall:   -.18 |
 | M/c bond lengths:100.0% within limits    .0% highlighted                 |
 | M/c bond angles:  80.1% within limits  19.9% highlighted                 |
+| Planar groups:    95.5% within limits   4.5% highlighted                 |
 +-----------------------------------------------------------------------+

   + May be worth investigating further.  * Worth investigating further.
```



Comparison with the single crystal model





I. Margiolaki, J. P. Wright, M. Wilmanns, A. N. Fitch & N. Pinotsis.
JACS 129 (38): 11865-11871 SEP 26 2007

Solvent scattering

- Fits with and without solvent contribution
- "Babinet's principle" gives modified atomic scattering factors:

$$f = f_0 - A\exp\left(\frac{-8\pi^2 U \sin^2\theta}{\lambda^2}\right)$$

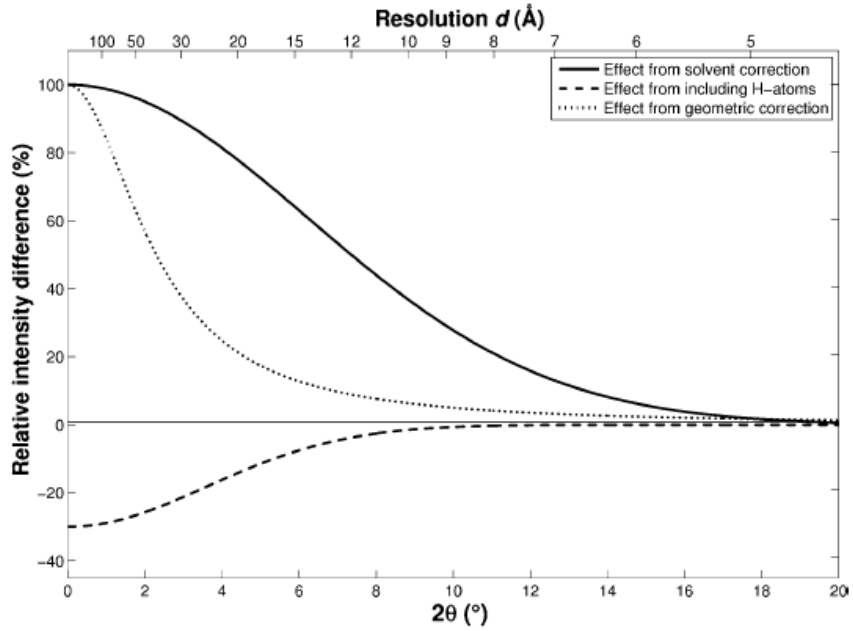
- Here A=6.63e$^-$ and U=1.18Å^2 (so carbon effectively has no electrons at low angles!)

In-house characterization of protein powder

Christian Grundahl Hartmann,[a] Ole Faurskov Nielsen,[b] Kenny Ståhl[a] and Pernille Harris[a]*

[a]Department of Chemistry, Technical University of Denmark, Kemitorvet 207, DK-2800 Kongens Lyngby, Denmark, and [b]Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark. Correspondence e-mail: ph@kemi.dtu.dk

J. Appl. Cryst. (2010). **43**, 876–882



GSAS : Refinement software

- Automatically generates restraints for amino acids and protein chain
- Band matrix (speed) + Marquardt damping (stability)
- Reads & writes PDB files and electron density maps

R. B. Von Dreele, “Combined Rietveld and stereochemical restraint refinement of a protein crystal structure.” *J. Appl. Cryst*. **32**, 1084-1089 (1999).

R. B. Von Dreele, “Binding of N-acetylglucosamine to chicken egg lysozyme: a powder diffraction study.” *Acta Cryst.* D**57**, 1836-1842 (2001).

Restraints / Constraints in GSAS

- All chemical information can be introduced with restraints (planes, chiral volumes, torsions)

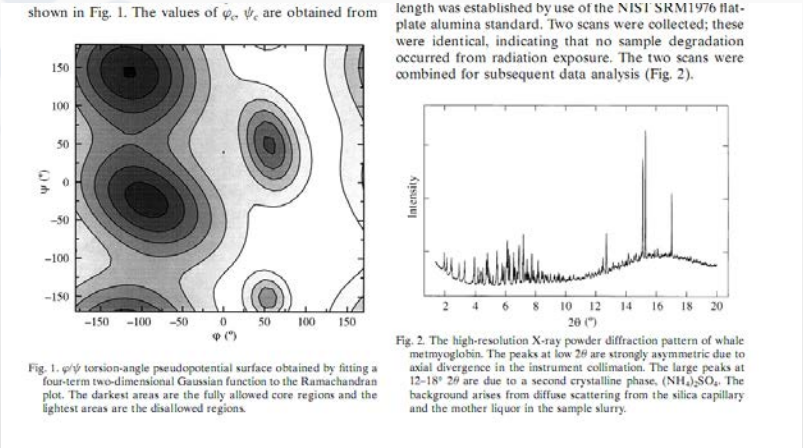

shown in Fig. 1. The values of φ_c, ψ_c are obtained from

Fig. 1. φ/ψ torsion-angle pseudopotential surface obtained by fitting a four-term two-dimensional Gaussian function to the Ramachandran plot. The darkest areas are the fully allowed core regions and the lightest areas are the disallowed regions.

length was established by use of the NIST SRM1976 flat-plate alumina standard. Two scans were collected; these were identical, indicating that no sample degradation occurred from radiation exposure. The two scans were combined for subsequent data analysis (Fig. 2).

Fig. 2. The high-resolution X-ray powder diffraction pattern of whale metmyoglobin. The peaks at low 2θ are strongly asymmetric due to axial divergence in the instrument collimation. The large peaks at 12–18° 2θ are due to a second crystalline phase, $(NH_4)_2SO_4$. The background arises from diffuse scattering from the silica capillary and the mother liquor in the sample slurry.

- Bond distances and angles now via flexible rigid bodies.

were employed in our starting model. A novel approach for refining protein structures using powder diffraction data is presented. In this approach, each amino acid is represented by a flexible rigid body (FRB). The FRB model requires a significantly smaller number of refinable parameters and restraints than a fully free-atom refinement. A total of 1542 stereochemical restraints were imposed in order to refine the positions of 800 protein atoms, two Zn atoms and 44 water

Acta D, in press

Irene Margiolaki,[a]* Anastasia E. Giannopoulou,[a] Jonathan P. Wright,[b]* Lisa Knight,[b] Mathias Norrman,[c] Gerd Schluckebier,[c] Andrew N. Fitch[b] and Robert B. Von Dreele[d]

High-resolution powder X-ray data reveal the T_6 hexameric form of bovine insulin

Extraction and use of correlated integrated intensities with powder diffraction data

Jon P. Wright*

Z. Kristallogr. 219 (2004) 791–802

Adding the nuisance parameters to the model being refined suggested a novel approach for generating a free *R*-factor statistic from powder data. If some peak intensities are treated like the nuisance parameters and ref variables along with the structural param estimated according to the dat equivalent approach in allow some pe their int Those int with the m lated from crystal free

Table 3. Agre weight matrix. non-diagonal ce *R* factor.

Model	I1		A1	AF
Pars	4	5	5	6
GooF	0.4464	0.4458	0.3925	0.3923
R_{free}	0.4553	0.4596	0.3000	0.3100
B Site	1.0	1.00(1)	1.0	1.00(1)
B *x*	0.1995(3)	0.1995(3)	0.1995(3)	0.1995(3)
$10^3 U_{11}$ B	3.1(3)	3.1(3)	2.7(3)	2.8(3)
$10^3 U_{22}$ B	$= U_{11}$	$= U_{11}$	4.2(2)	4.2(2)
$10^3 U_{11}$ La	5.06(2)	5.06(2)	5.09(2)	5.06(2)

The case for open computer programs

Darrel C. Ince, Leslie Hatton & John Graham-Cumming

Affiliations | Contributions | Corresponding author

Nature 482, 485–488 (23 February 2012) | doi:10.1038/nature10836
Received 09 May 2011 | Accepted 05 January 2012 | Published online 22 February 2012

ngs to come out of protein

flections from refinement

alculated values to observed

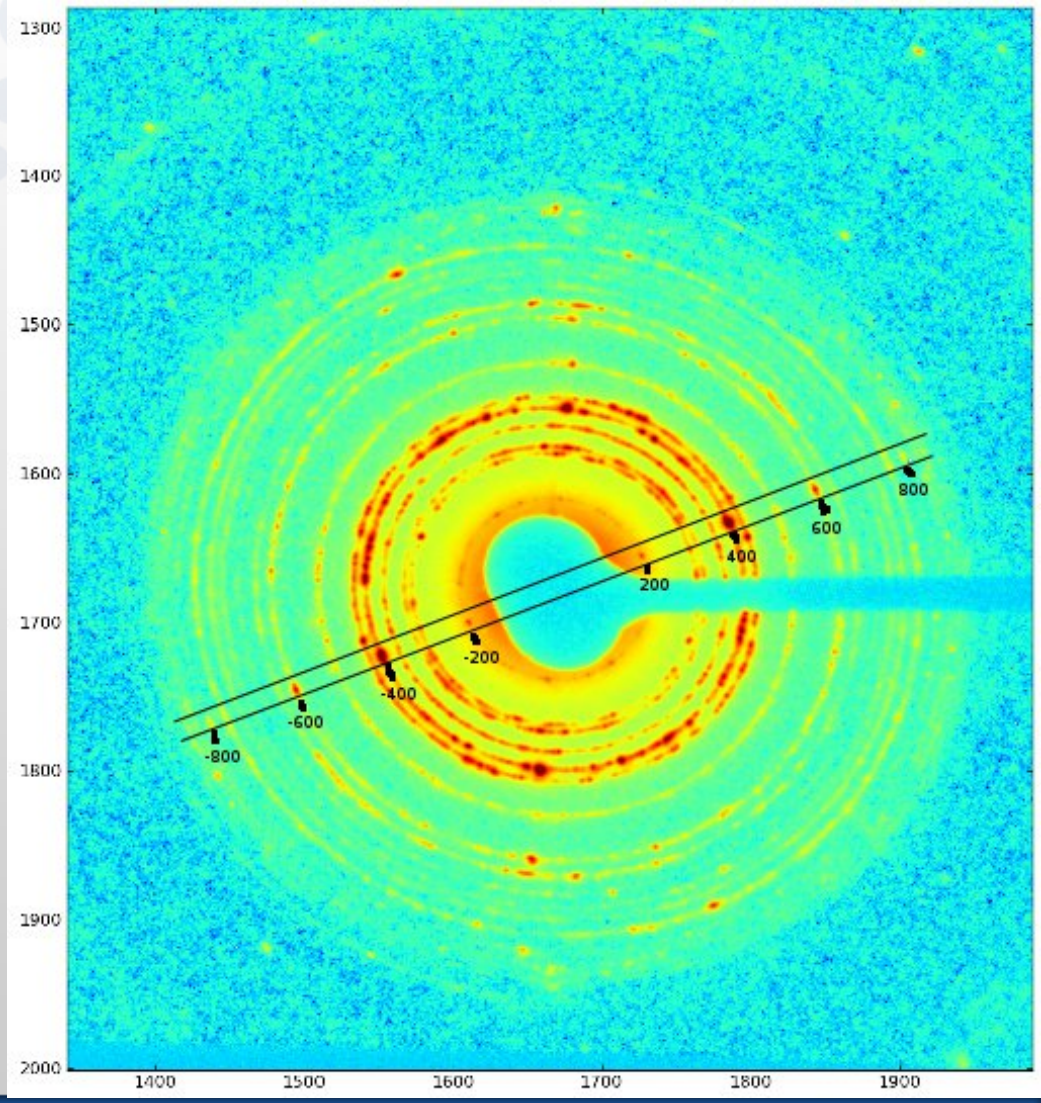
- If a model is fitting noise, the calculated values for these excluded peaks is likely to get worse.

- For overlapped data:
  - Excluded reflections are free variables
  - “observed” values are those which best fit the data
  - Compare as before

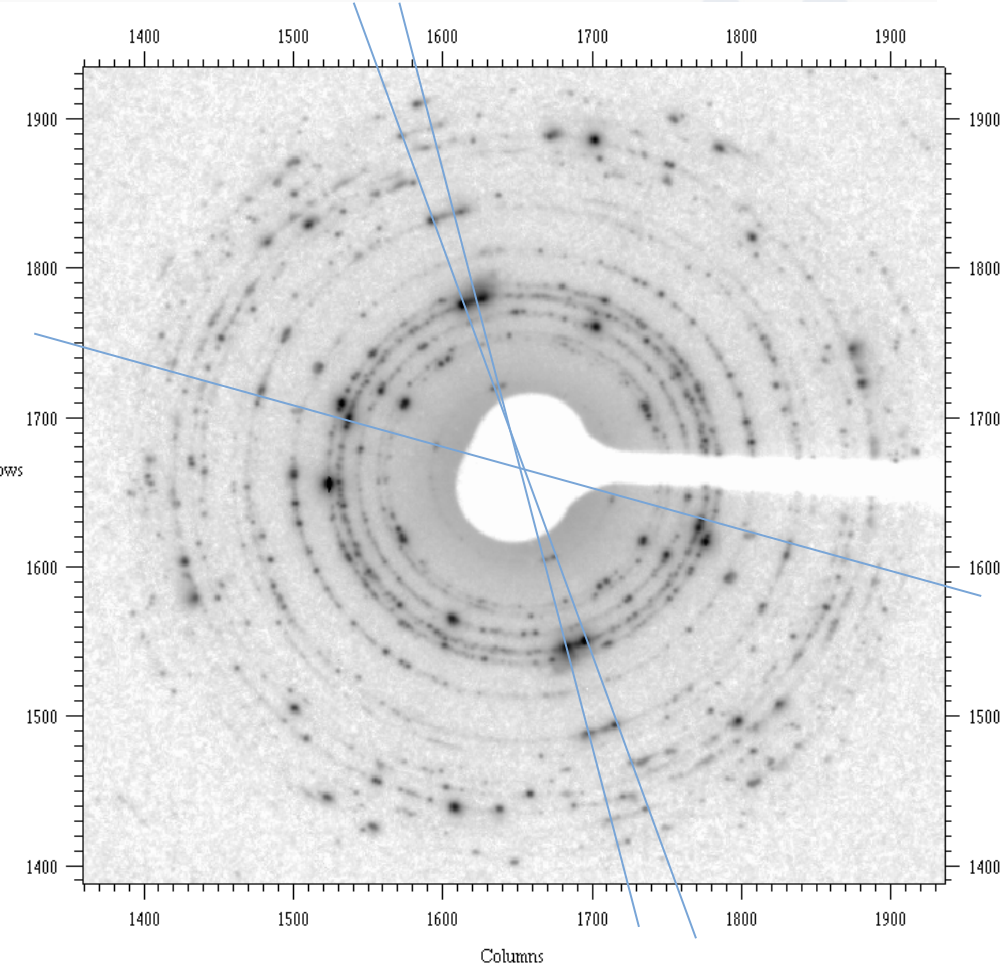
Rfree + powder refinements with weight matrices – Wright, Z. Krist ’04
Used python + cctbx (R. Grosse Kunstleve tomorrow)

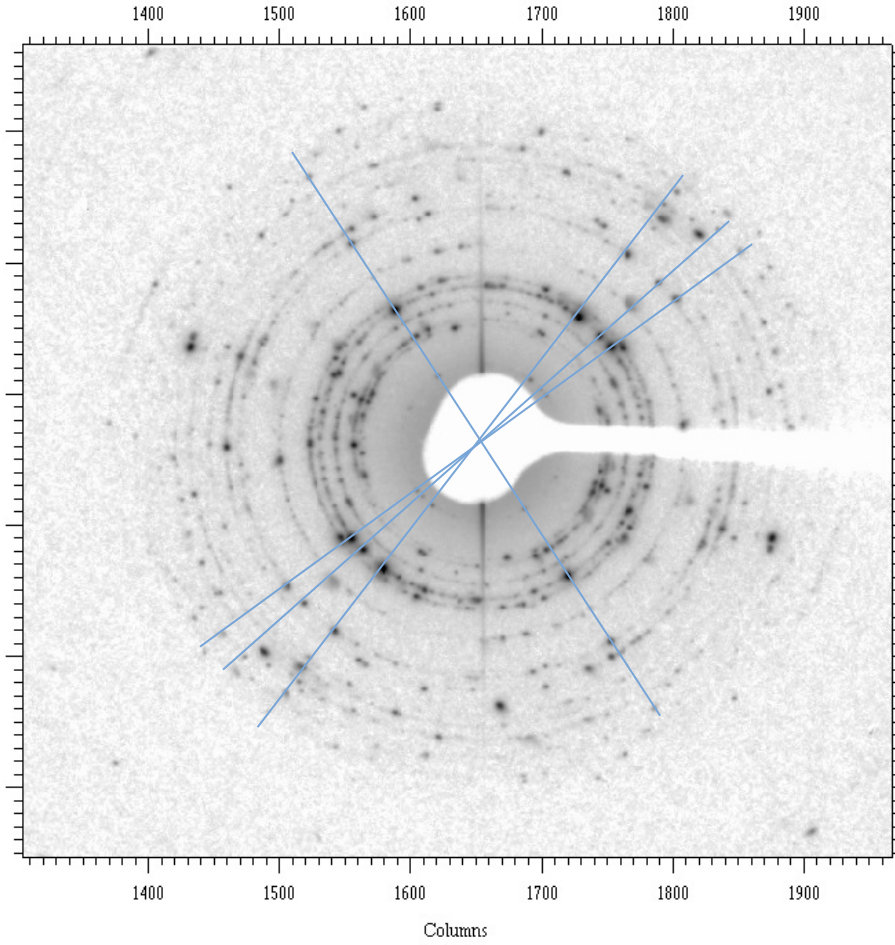
Space group absences overlapped?

- 2_1 absences are overlapped in powder diagram
- Since 2-fold is the long 337 Å axis
- Systematic problem as next axis is 1/3 length

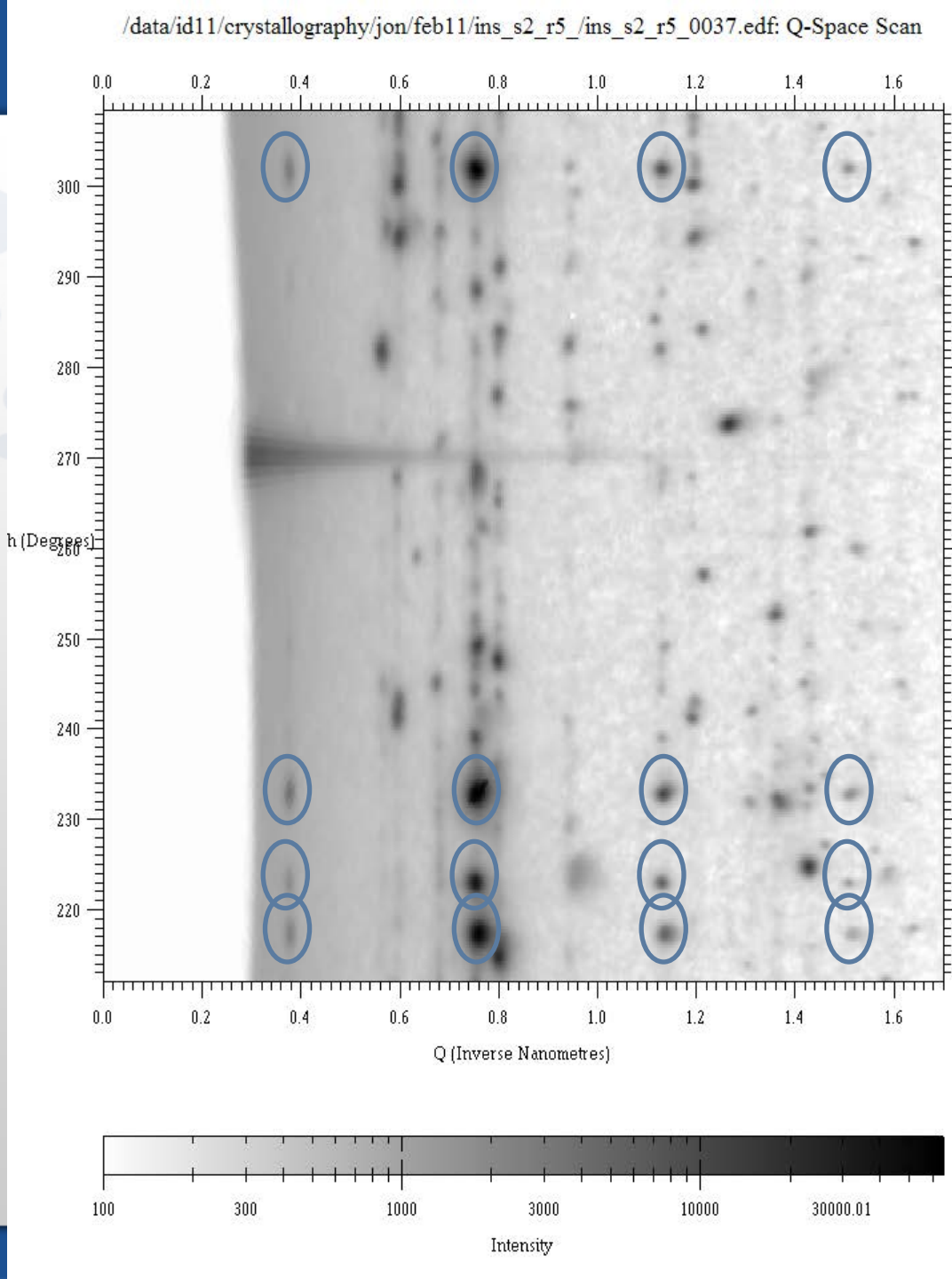


Newer sample with crystallites…





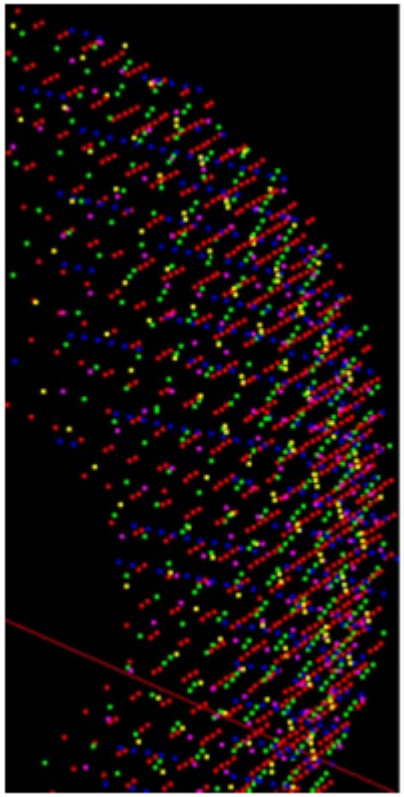
- Cake image (radial transform)
- Ewald sphere is roughly flat
- We never see spots for the k=odd positions
- Assume the symmetry is $P2_1$.



Multigrain crystallography

Z. Kristallogr. **227** (2012) 63–78 / **DOI** 10.1524/zkri.2012.1438

Henning O. Sørensen[I,1,*], Søren Schmidt[I], Jonathan P. Wright[II], Gavin B. M. Vaughan[II], Simone Techert[III], Elspeth F. Garman[IV], Jette Oddershede[I], Jav Davaasambu[III], Karthik S. Paithankar[IV,2], Carsten Gundlach[II,3] and Henning F. Poulsen[I,*]

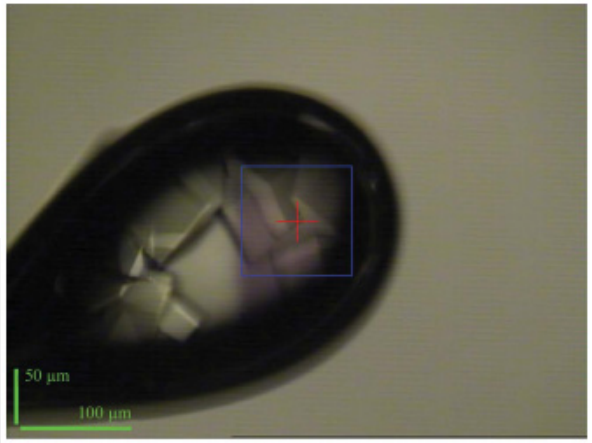


Finding orientation matrices for multiple single crystals

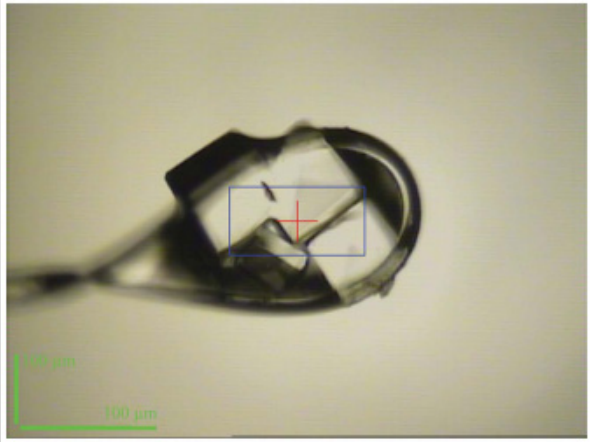
EU project "TotalCrystallography"

Large software investment

http://fable.sourceforge.net



(*a*)



(*b*)

Acta Cryst. (2011). D**67**, 608–618

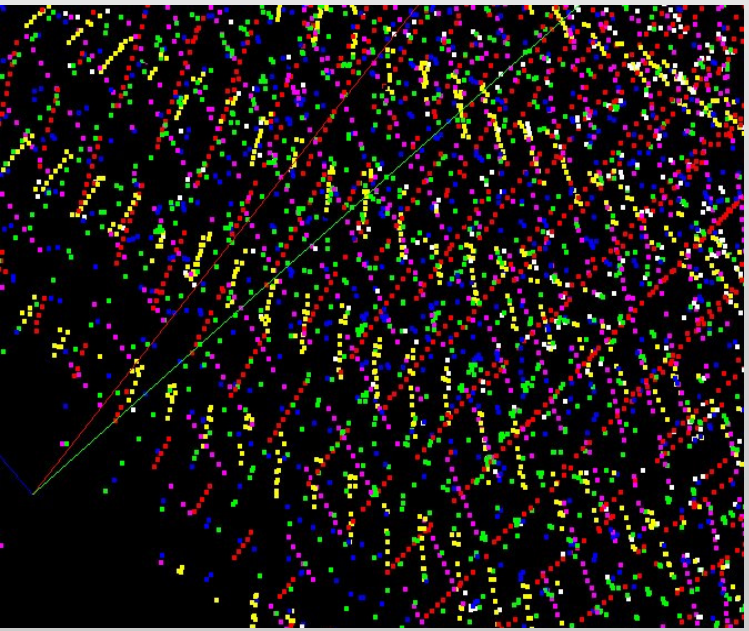
Simultaneous X-ray diffraction from multiple single crystals of macromolecules

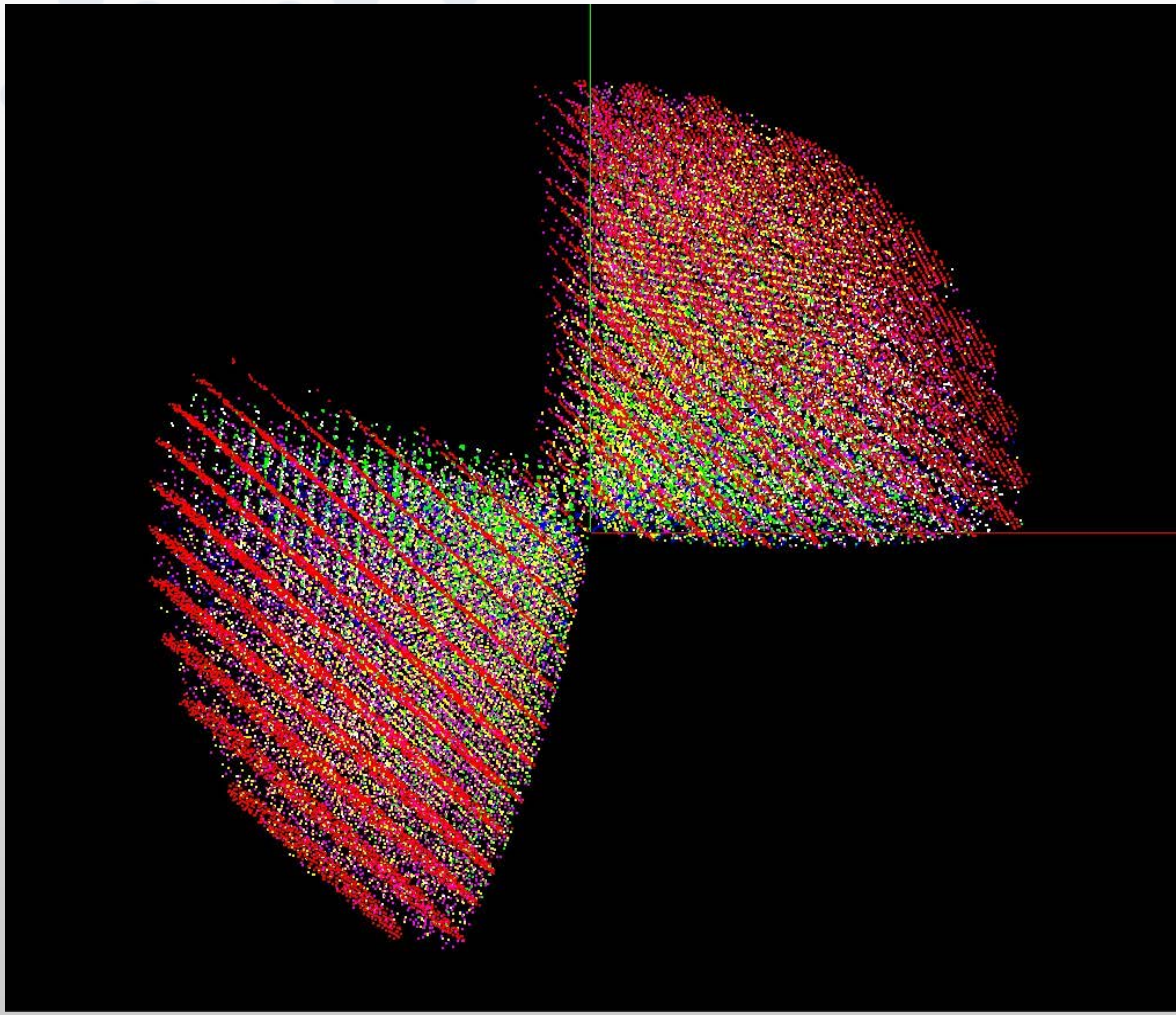
Karthik S. Paithankar,[a]‡ Henning O. Sørensen,[b]§ Jonathan P. Wright,[c] Søren Schmidt,[b] Henning F. Poulsen[b] and Elspeth F. Garman[a]*

Karthik S. Paithankar,[a]‡
Henning O. Sørensen,[b]§
Jonathan P. Wright,[c] Søren
Schmidt,[b] Henning F. Poulsen[b]
and Elspeth F. Garman[a]*

5 Crystals of lysozyme, ID14 data

- index_unknown.py
- FFT based method





Fundamental size limit (protein)

Structure, Vol. 11, 13–19, January, 2003, ©2003 Elsevier Science Ltd. A

How does Radiation Damage in Protein Crystals Depend on X-Ray Dose?

Piotr Sliz,[1] Stephen C. Harrison,[1,3] and Gerd Rosenbaum[2]

Is radiation damage to cryopreserved protein crystals strictly proportional to accumulated dose at the high-flux density of beams from undulators at third-generation synchrotron sources? The answer is "yes," for overall damage to several different kinds of protein crystals at flux densities up to 10^{15} ph/sec/mm^2 (APS beamline 19-ID). We find that, at 12 keV (1 Å wavelength), about ten absorbed photons are sufficient to "kill" a unit cell. As this corresponds to about one elastically scattered photon, each unit cell can contribute only about one photon to total Bragg diffraction. The smallest crystal that can yield a full data set to 3.5 Å resolution has a diameter of about 20 μm (100 Å unit cell).

- Comes from radiation dose tolerance for a single crystal
- 20 micron crystal **required** for 100 Angstrom protein unit cell
- Smaller crystals?
  - powder methods

LETTER

doi:10.1038/nature09750

Femtosecond X-ray protein nanocrystallography

Henry N. Chapman[1,2], Petra Fromme[3], Anton Barty[1], Thomas A. White[1], Richard A. Kirian[4], Andrew Aquila[1], Mark S. Hunter[3], Joachim Schulz[1], Daniel P. DePonte[1], Uwe Weierstall[4], R. Bruce Doak[4], Filipe R. N. C. Maia[5], Andrew V. Martin[1], Ilme Schlichting[6,7], Lukas Lomb[7], Nicola Coppola[1]†, Robert L. Shoeman[7], Sascha W. Epp[6,8], Robert Hartmann[9], Daniel Rolles[6,7], Artem Rudenko[6,8], Lutz Foucar[6,7], Nils Kimmel[10], Georg Weidenspointner[11,10], Peter Holl[9], Mengning Liang[1], Miriam Barthelmess[12], Carl Caleman[1], Sébastien Boutet[13], Michael J. Bogan[14], Jacek Krzywinski[13], Christoph Bostedt[13], Saša Bajt[12], Lars Gumprecht[1], Benedikt Rudek[6,8], Benjamin Erk[6,8], Carlo Schmidt[6,8], André Hömke[6,8], Christian Reich[9], Daniel Pietschner[10], Lothar Strüder[6,10], Günter Hauser[10], Hubert Gorke[15], Joachim Ullrich[6,8], Sven Herrmann[10], Gerhard Schaller[10], c Messerschmidt[13], John D. Bozek[13], Stefan P. Hau-Riege[16], rra[14], Dmitri Starodub[14], Garth J. Williams[13], Janos Hajdu[5], n[5], Andrea Rocker[5], Olof Jönsson[5], Martin Svenda[5], Stephan Stern[1], , Faton Krasniqi[6,7], Mario Bott[7], Kevin E. Schmidt[4], Xiaoyu Wang[4], ds[7], Richard Neutze[18], Stefano Marchesini[17], Raimund Fromme[3], Gorkhover[19], Inger Andersson[20], Helmut Hirsemann[12], John C. H. Spence[4]

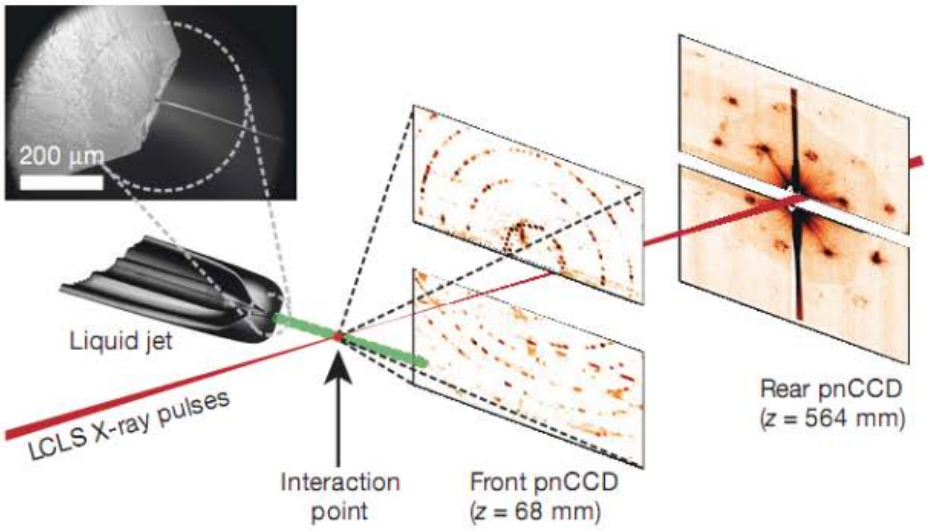


Figure 1 | Femtosecond nanocrystallography. Nanocrystals flow in their buffer solution in a gas-focused, 4-µm-diameter jet at a velocity of 10 m s^{-1} perpendicular to the pulsed X-ray FEL beam that is focused on the jet. Inset, environmental scanning electron micrograph of the nozzle, flowing jet and focusing gas[30]. Two pairs of high-frame-rate pnCCD detectors[12] record low- and high-angle diffraction from single X-ray FEL pulses, at the FEL repetition rate of 30 Hz. Crystals arrive at random times and orientations in the beam, and the probability of hitting one is proportional to the crystal concentration.

Single shot per crystal in a random orientation

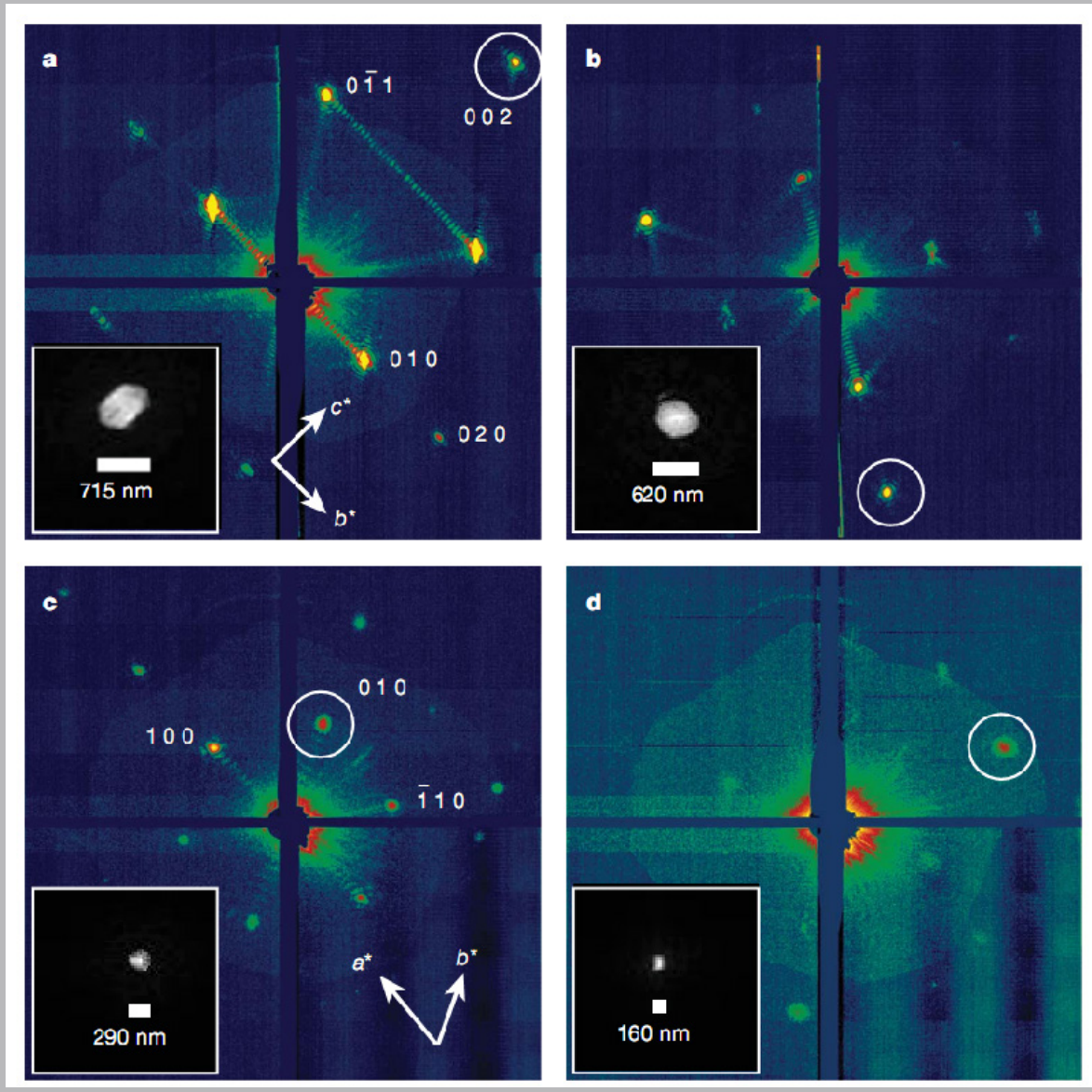
Crystals less than 2 microns.

8.5 Angstrom data for photosystem II.
P63: 281x281x165 (would be12 A for powder)

Millions of crystals averaged, one at a time...

- Nature 470,
- 73–77 (2011)

- Submicron crystal shapes from coherent diffraction



High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography

Sébastien Boutet,[1*] Lukas Lomb,[2,3] Garth J. Williams,[1] Thomas R. M. Barends,[2,3] Andrew Aquila,[4] R. Bruce Doak,[5] Uwe Weierstall,[5] Daniel P. DePonte,[4] Jan Steinbrener,[2,3] Robert L. Shoeman,[2,3] Marc Messerschmidt,[1] Anton Barty,[4] Thomas A. White,[4] Stephan Kassemeyer,[2,3] Richard A. Kirian,[5] M. Marvin Seibert,[1] Paul A. Montanez,[1] Chris Kenney,[6] Ryan Herbst,[6] Philip Hart,[6] Jack Pines,[6] Gunther Haller,[6] Sol M. Gruner,[7,8] Hugh T. Philipp,[7] Mark W. Tate,[7] Marianne Hromalik,[9] Lucas J. Koerner,[10] Niels van Bakel,[11] John Morse,[12] Wilfred Ghonsalves,[1] David Arnlund,[13] Michael J. Bogan,[14] Carl Caleman,[4] Raimund Fromme,[15] Christina Y. Hampton,[14] Mark S. Hunter,[15] Linda C. Johansson,[13] Gergely Katona,[13] Christopher Kupitz,[15] Mengning Liang,[4] Andrew V. Martin,[4] Karol Nass,[16] Lars Redecke,[17,18] Francesco Stellato,[4] Nicusor Timneanu,[19] Dingjie Wang,[5] Nadia A. Zatsepin,[5] Donald Schafer,[1] James Defever,[1] Richard Neutze,[13] Petra Fromme,[15] John C. H. Spence,[5] Henry N. Chapman,[4,16] Ilme Schlichting[2,3]

Structure determination of proteins and other macromolecules has historically required the growth of high-quality crystals sufficiently large to diffract x-rays efficiently while withstanding radiation damage. We applied serial femtosecond crystallography (SFX) using an x-ray free-electron laser (XFEL) to obtain high-resolution structural information from microcrystals (less than 1 micrometer by 1 micrometer by 3 micrometers) of the well-characterized model protein lysozyme. The agreement with synchrotron data demonstrates the immediate relevance of SFX for analyzing the structure of the large group of difficult-to-crystallize molecules.

We collected about 1.5 million individual "snapshot" diffraction patterns for 40-fs duration pulses at the LCLS repetition rate of 120 Hz using the CSPAD. About 4.5% of the patterns were classified as crystal hits, 18.4% of which were indexed and integrated with the CrystFEL software (*14*) showing excellent statistics to 1.9 Å resolution (Table 1 and table S1). In addition, 2 million diffraction patterns were collected by using x-ray pulses of 5-fs duration, with a 2.0% hit rate and a 26.3% indexing rate, yielding 10,575 indexed patterns. The structure, partially shown in Fig. 2A,

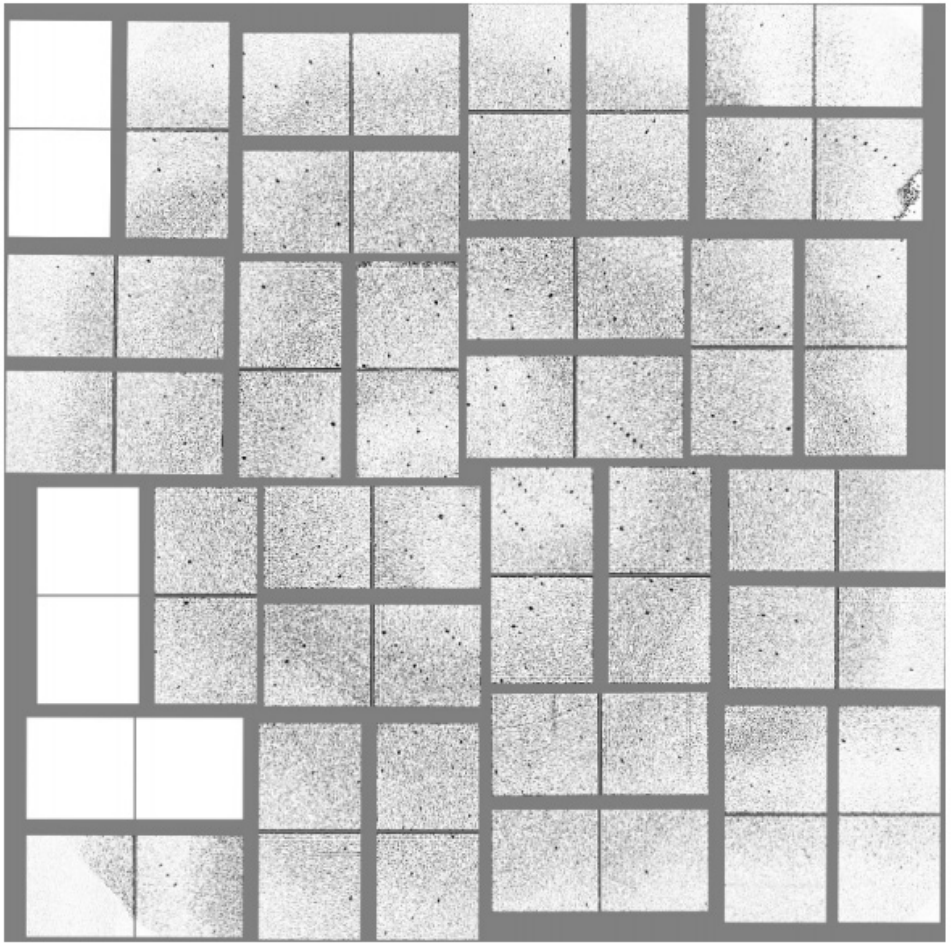


Fig. S1.
A typical diffraction pattern using a single 40 fs pulse showing Bragg peaks to the edge of the CSPAD detector. Note that some of the tiles of the CSPAD were not functional at the time of the measurement and appear completely white.

Summary

- Control sample preparation to get reproducible data
- Get the best data you can
  - High resolution
  - Multi-dataset
- Structures can be “solved” if approximate molecule is available
- Refinements of small proteins are possible
- Structural detail is ~1/10 cell parameter

Thank you for listening

- ESRF Proteins
  - Irene Margiolaki
  - Andy Fitch
  - Yves Watier
  - Ines Collings
  - Sotonye Dagogo
  - Lisa Knight
  - Mark Jenner
  - Lucy Saunders
- ESRF Matsci
  - Gavin Vaughan
  - Alex Bytchkov
  - Loredana Erra
  - Caroline Curfs
  - Carsten Gundlach
  - Gaelle Suchet
  - Henri Gleyzolle
  - Denis Van Brussel
  - Jean Michel Chaize
  - Andy Goetz
  - Ricardo Hino
  - Michel Rossat
  - Patras
  - Irene Margiolaki
  - Fotini Karavassili
  - Anastasia E. Giannopoulou
  - Eleni Kotsiliti
- Marseille
  - Nicolas Papageorgiou
  - Bruno Coutard
  - Violane Lantez
  - Ernest Gould
  - Marion Giffard
  - Francois Bonneté
- EPFL
  - Marc Schiltz
  - Celine Besnard
  - Sebastian Basso
- Novo Nordisk
  - Matthias Normann
  - Gerd Schlukebier
  - Annette Frost Jensen
  - Anders Norlander
- University of Oxford
  - Elspeth Garman
  - Karthik Paithankar
- EMBL Hamburg
  - Mattias Wilmanns
  - Nikos Pinotsis
- Argonne National Laboratory
  - Bob Von Dreele