Molecular replacement experiences

Michail Isupov
Biosciences, University of Exeter

Outline of this talk

- **Molecular replacement approaches when pipelines fail**
- **Simple steps in conventional MR search (MOLREP)**
- **Phased MR examples**
- **Density modification**
- **Crystal anisotropy**
- **Self-rotation function and examples**
- **Conserved molecular symmetry use example**

A Black Box crystallography pipeline

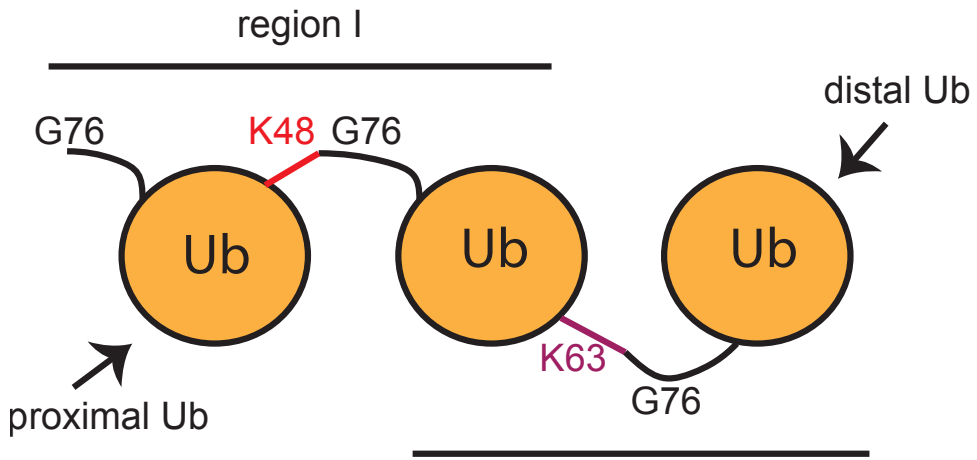
Data,
sequence

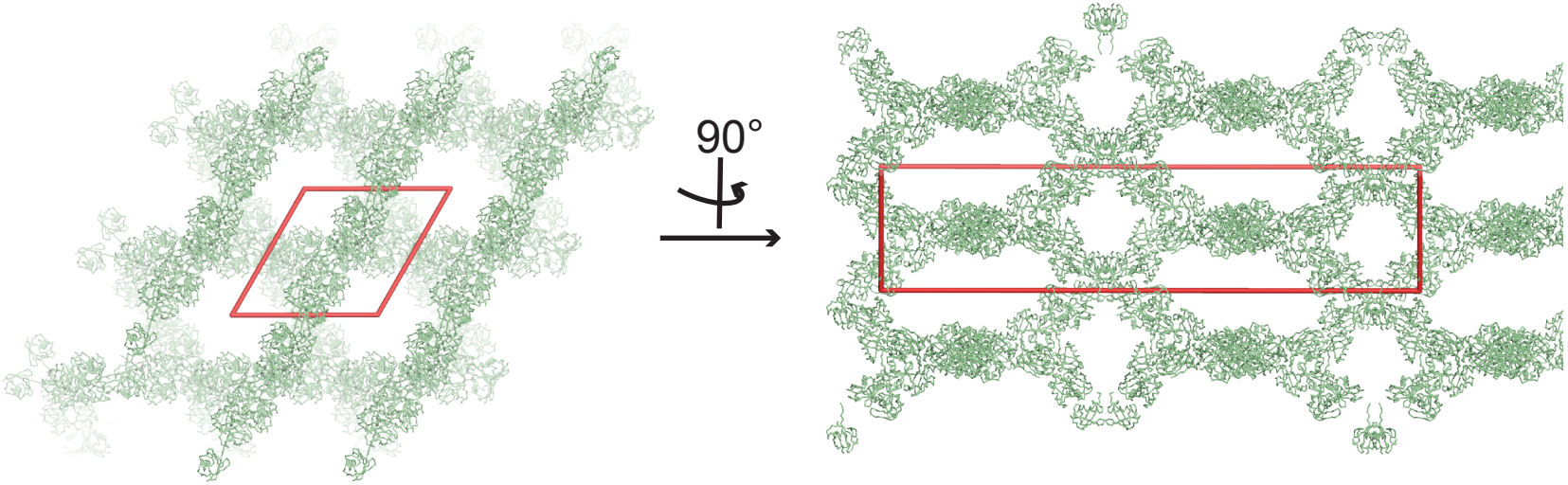
Refined
Model

User does not have to know what is inside

Tri-ubiquitin with Dr Reuven Wiener (number of RF peaks).

Space group $P6_122$
Cell 110.8, 110.8, 417.0 Å
Resolution 3.2 Å
Solvent content 82 %



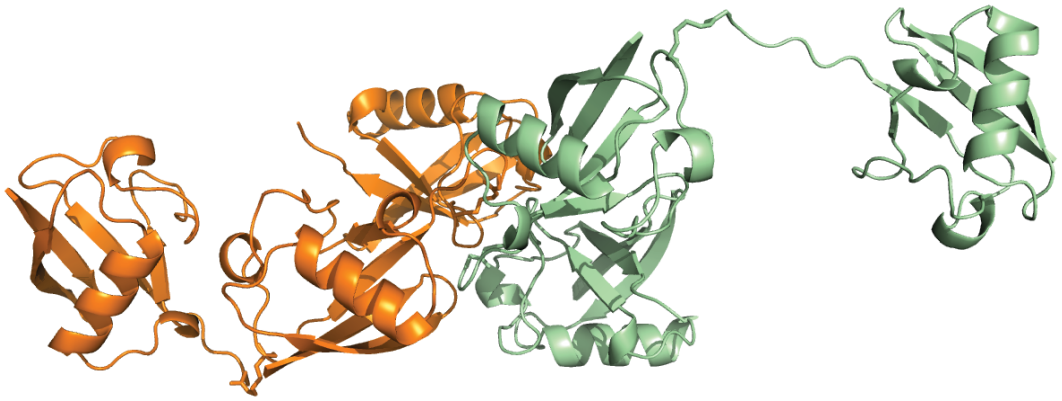


Padala et al. (2017) JMB 429, 3801-3813.

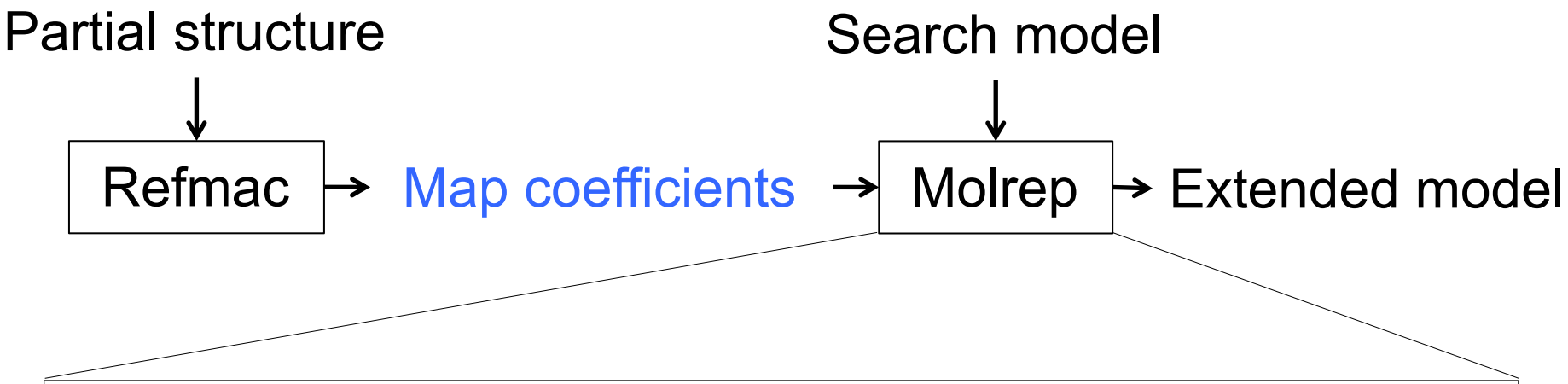
Tri-Ub

MORDA located 3 Ub monomers in $P6_122$ with good contrast, however the partial structure did not refine, probably due to low resolution of the data.

Increasing number of RF peaks to 200 (instead of default 30) in MOLREP gave clear solution for 6 Ub monomers (2 trimers), this model refined to FreeR below 28 %. Two of the correct peaks were in the second hundred of RF peaks list.



Search in the electron density map



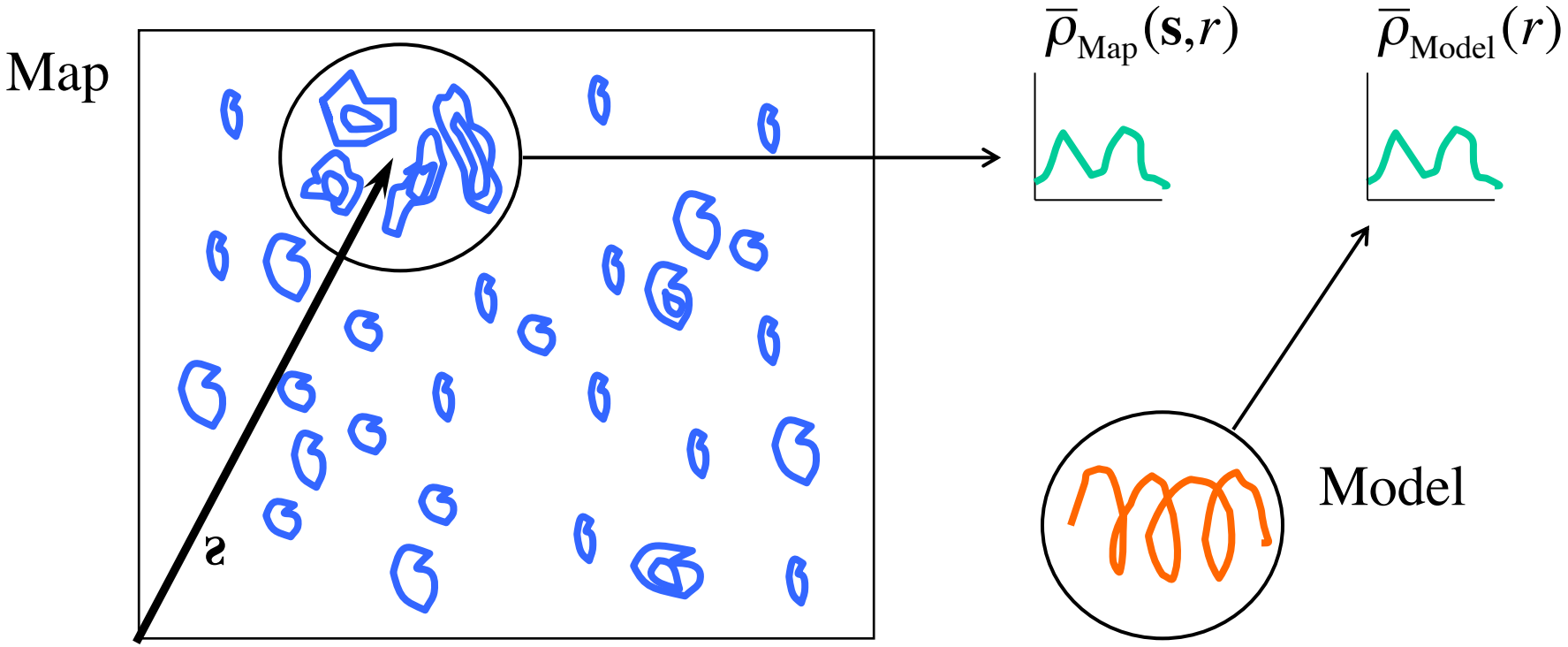
Search in the map

- Calculate 2-1 or 1-1 maps after restrained refinement of partial structure
- Flatten the map which corresponding to the known substructure
- Calculate structure amplitudes from the modified map
- Use these modified amplitudes in Rotation Function
- And finally – Phased TF

Molrep: SAPTF

Spherically Averaged Phased Translation Function
(FFT based algorithm)

$$\mathrm{SAPTF}(\mathbf{s}) = \int \overline{\rho}_{\mathrm{Map}}(\mathbf{s},r)\, \overline{\rho}_{\mathrm{Model}}(r)\, r^2\, \mathrm{d}r$$



Molrep: Search in the map with SAPTF

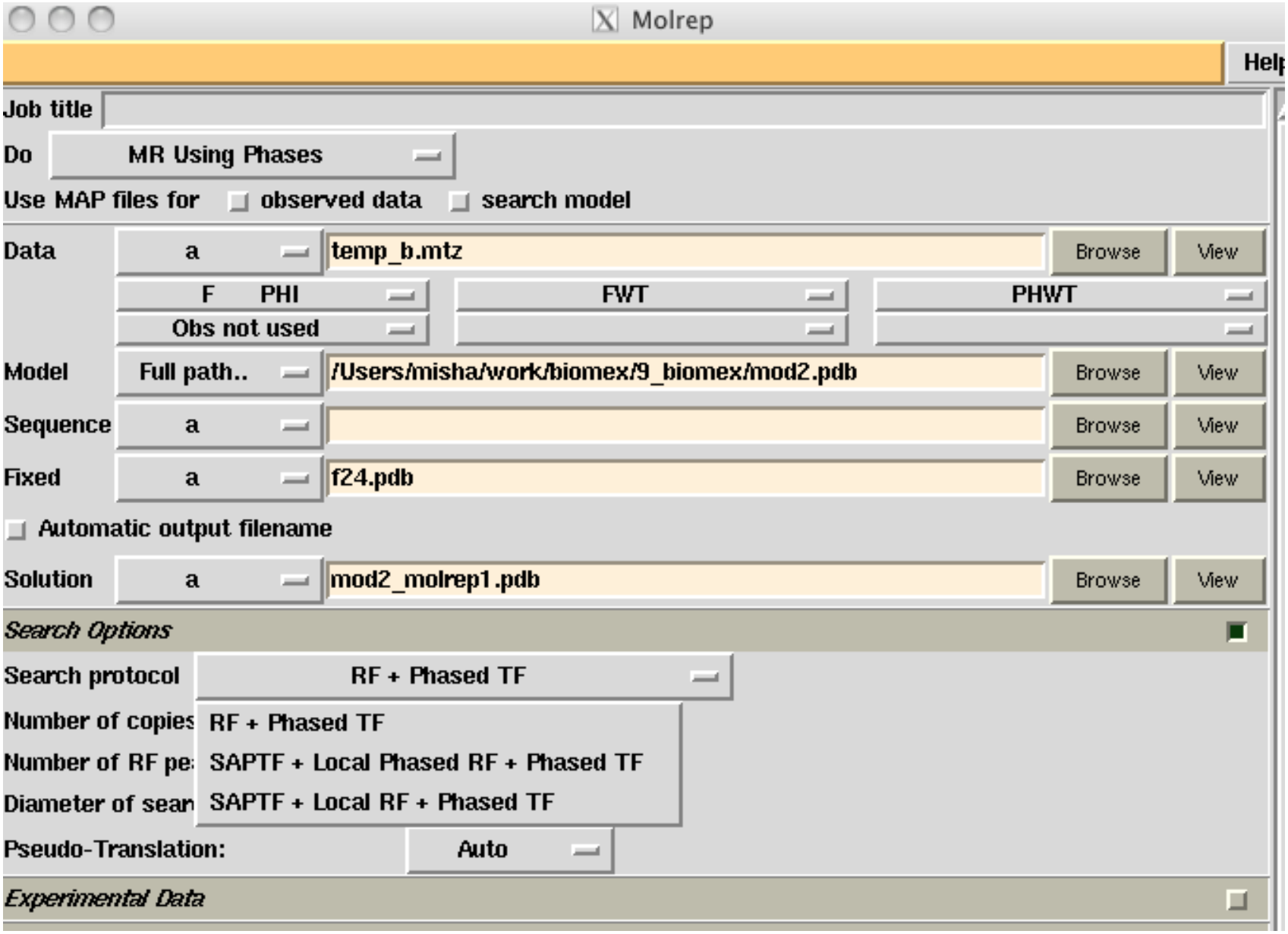
1. Find approximate position:
   Spherically Averaged Phased Translation Function

2. Find orientation:
   Local Rotation Function
   - Structure amplitudes from the density within the SAPTF sphere

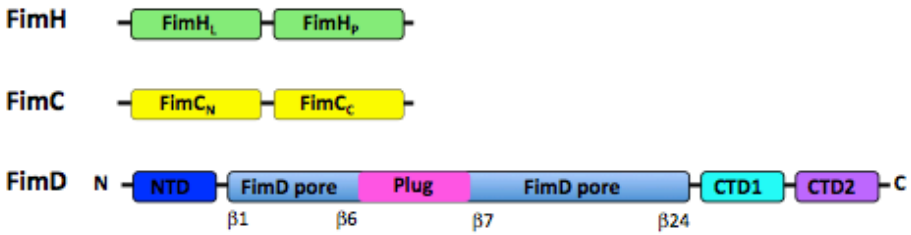
3. Verify and adjust position:
   Phased Translation Function

- Local RF is less sensitive than Phased RF to inaccuracy of the model position

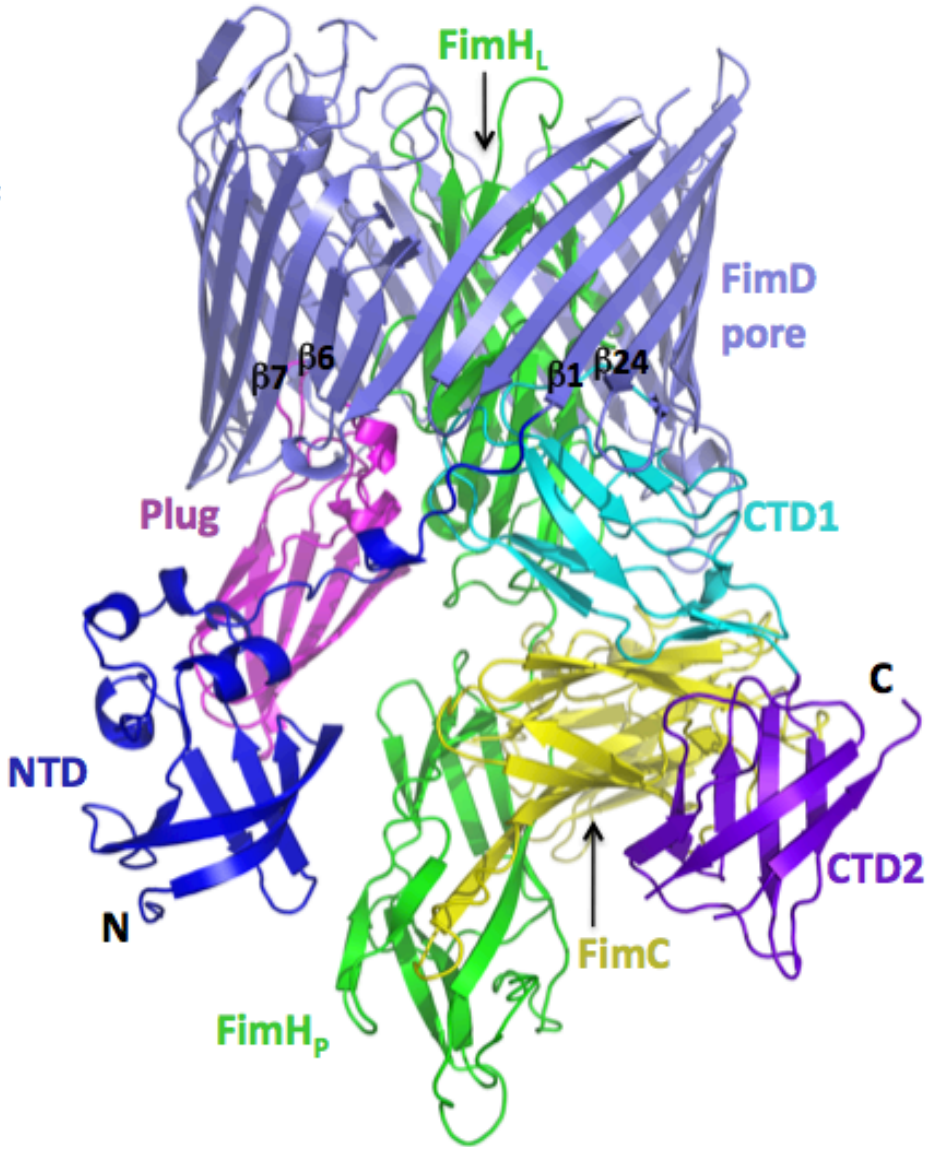
Phased MR options in MOLREP



Usher complex *E. coli*



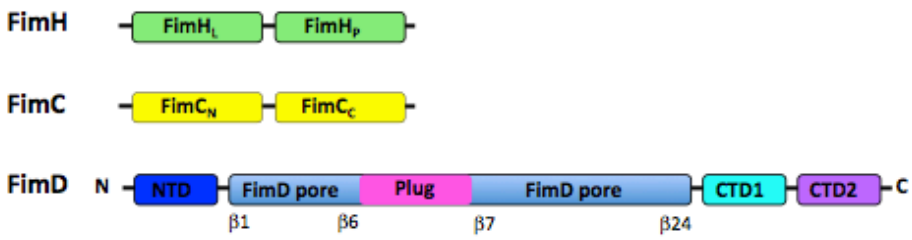
- Asymmetric unit two copies
- Resolution 2.8 Å



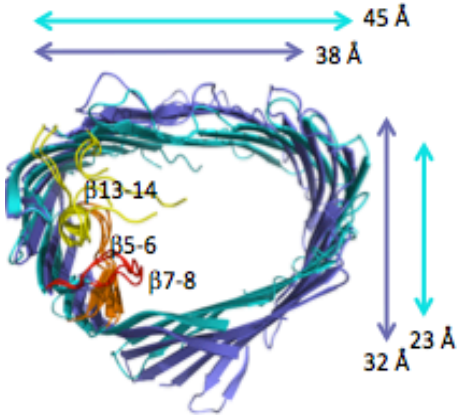
Phan et. al (2011) Nature, 474, 50-53

Usher complex structure solution

1. Conventional MR
   - FimC-N + FimC-C
   - FimH-L + FimH-P
   - FimD-Pore



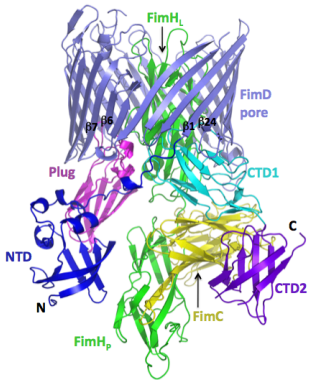
2. Jelly body refinement (Refmac)
   - FimD-Pore

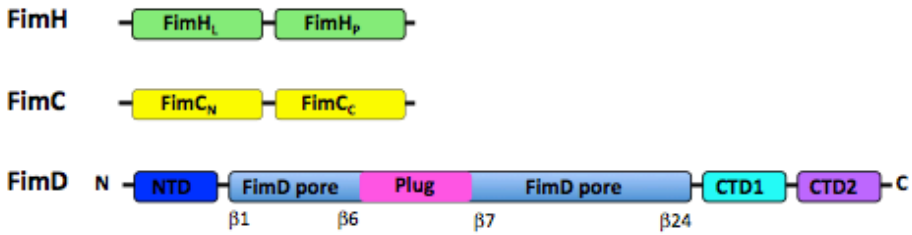


3. <u>Fitting into the electron density</u>
   - FimD-Plug
   - FimD-NTD
   - FimD-CTD-2

4. Manual building
   - FimD-CTD-1

Performance of fitting methods

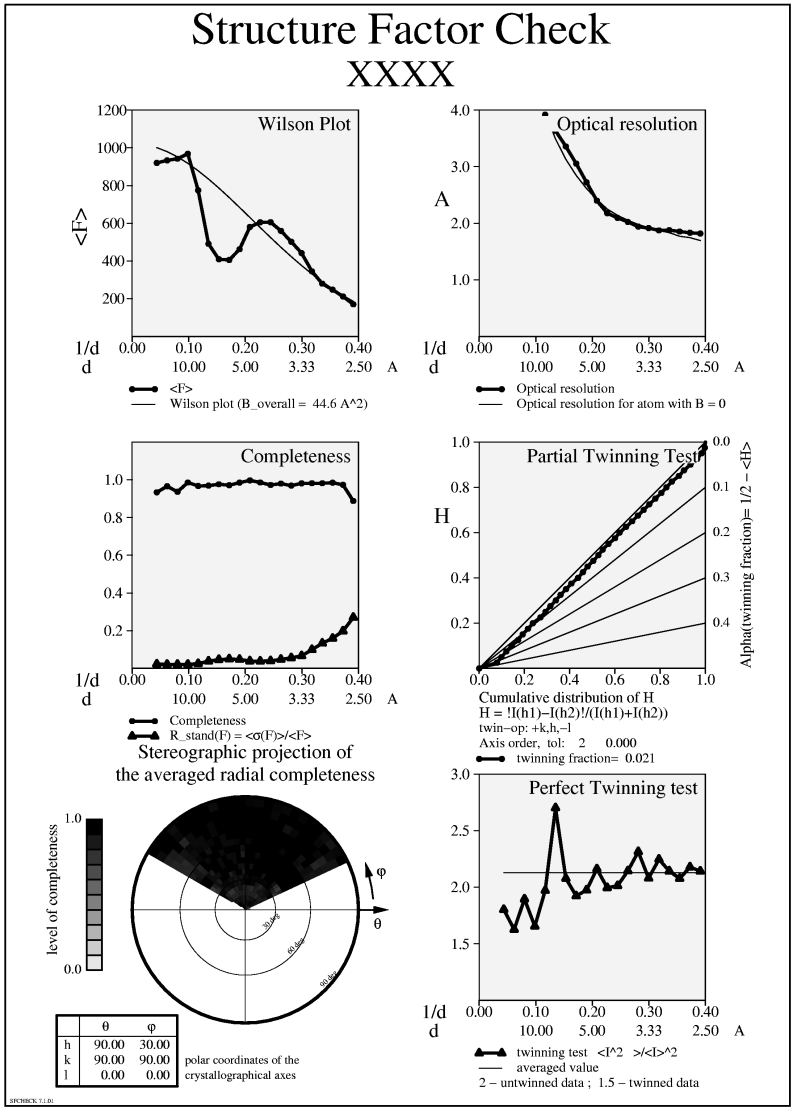




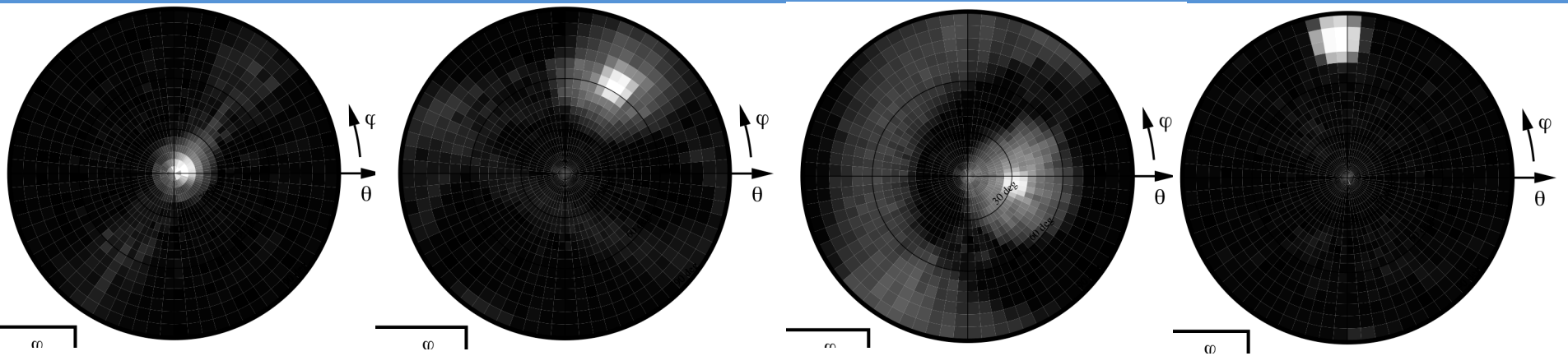
	search model	sequence identity	"Masked" RF PTF	SAPTF PRF PTF	SAPTF Local RF PTF
FimD-Plug	3fip_A	38.5%	2	–	1
FimD-NTD	1ze3_D	100%	2	1	2
FimD-CTD-2	3l48_A	33.3%	–	2	–

Trying several methods is a good practice (also a way of cross-validation)

Sfcheck


Structure Factor Check
XXXX
Wilson Plot
<F>
1/d
d
<F>
Wilson plot (B_overall = 44.6 A^2)
Optical resolution
A
Optical resolution
Optical resolution for atom with B = 0
Completeness
Completeness
R_stand(F) = <σ(F)>/<F>
Partial Twinning Test
H
Alpha(twinning fraction)= 1/2 – <H>
Cumulative distribution of H
H = !I(h1)–I(h2)!/(I(h1)+I(h2))
twin–op: +k,h,–l
Axis order, tol: 2 0.000
twinning fraction= 0.021
Stereographic projection of
the averaged radial completeness
level of completeness
φ
θ
θ φ
h 90.00 30.00
k 90.00 90.00
l 0.00 0.00
polar coordinates of the
crystallographical axes
Perfect Twinning test
twinning test <I^2 >/<I>^2
averaged value
2 – untwinned data ; 1.5 – twinned data

Multicrystal averaging: the same lattice.



In a recent medium resolution (2.5Å) protein complex case I have built most of the model (2 copies of two proteins in a.u.), however a particular interdomain loop (containing a disulphide) in a larger protein remained poorly defined. Inspection of this P1 crystal data statistics suggested good (for P1) completeness of 93% but low redundancy (1.7).

Multicrystal averaging (DMMULTI) using three additional slightly non-isomorphous datasets (2.7-2.9Å) in the same space group (different blind zones), alongside with data of domain structures of the smaller protein in this crystal from pdb and subsequent phased refinement in REFMAC5 produced a map with interpretable density for the elusive loop.

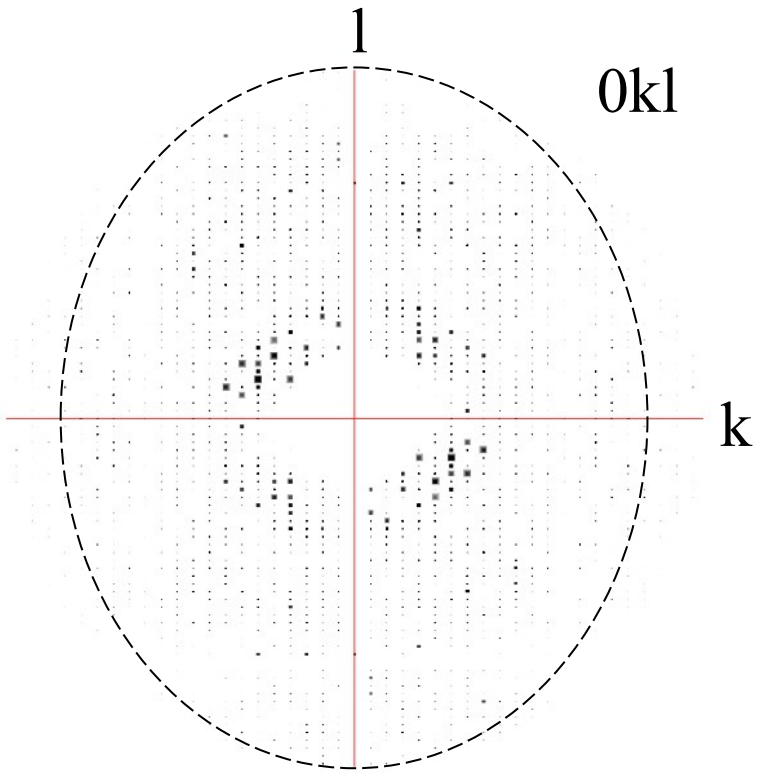
ANISOTROPY CORRECTION OF DATA

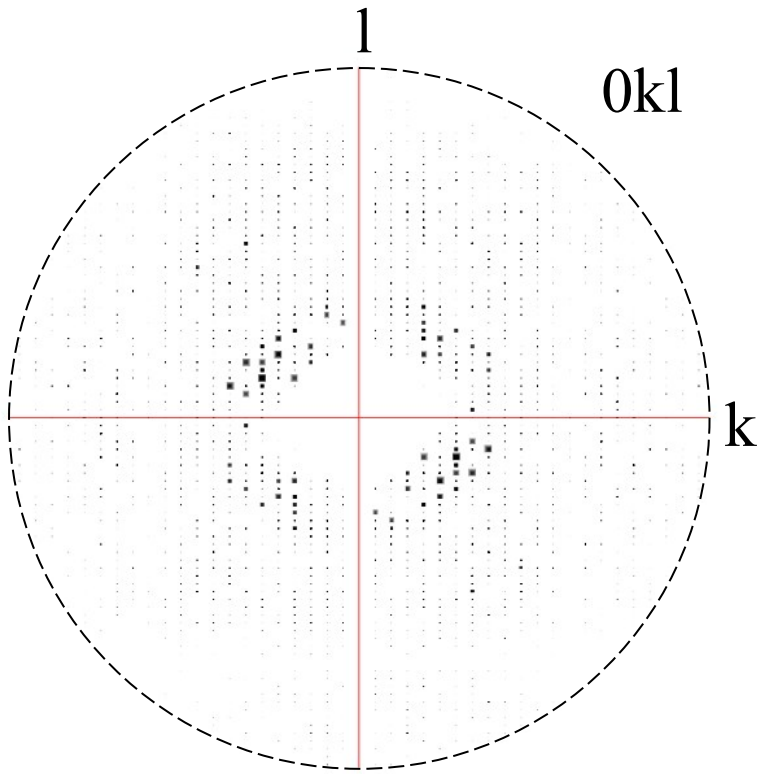
MOLREP does anisotropic corrections using structure factor amplitudes.

This may result in raising the noise level in extreme cases

PHASER uses intensities for correction, which appears to give better results.

STARANISO server uses unmerged intensities to produce elliptically complete data, these appear to give better results in MOLREP MR.





STARANISO corrected data is better in MR phasing
May be better in EP phasing

MR case: F222

RF staraniso

MOLREP

No correct RF peak in first 40

#Sol_		theta	phi	chi	alpha	beta	gamma	Rf	Rf/sigma
Sol_RF	1	34.27	116.59	99.25	70.77	50.81	17.60	0.3181E-01	4.68
Sol_RF	2	146.36	-62.44	142.75	139.60	63.33	84.48	0.2585E-01	3.81
Sol_RF	3	142.88	-54.87	134.42	152.92	67.60	82.66	0.2470E-01	3.64
Sol_RF	4	143.11	-53.26	131.45	156.16	66.36	82.67	0.2453E-01	3.61
Sol_RF	5	149.71	-68.68	153.17	126.76	58.77	84.12	0.2359E-01	3.47
Sol_RF	6	135.07	-72.35	124.14	144.48	77.22	109.17	0.2220E-01	3.27

TF staraniso

High contrast solution

MOLREP

	RF	TF	theta	phi	chi	tx	ty	tz	TF/sg	wRfac	Score
1	1	2	34.27	116.59	99.25	0.224	0.190	0.455	6.38	0.606	0.24150
2	10	14	97.18	130.04	50.83	0.007	0.059	0.385	3.66	0.633	0.18717
3	8	3	30.96	-140.81	79.97	0.123	0.159	0.195	3.73	0.634	0.17748
4	9	13	46.43	58.19	119.28	0.409	0.058	0.093	4.08	0.627	0.17734
5	2	7	146.36	-62.44	142.75	0.380	0.060	0.311	3.60	0.633	0.17621
6	5	3	149.71	-68.68	153.17	0.443	0.333	0.254	4.34	0.634	0.17174
7	7	11	149.69	-95.68	174.41	0.286	0.402	0.302	3.32	0.628	0.16794
8	3	1	142.88	-54.87	134.42	0.334	0.046	0.045	3.90	0.639	0.16778

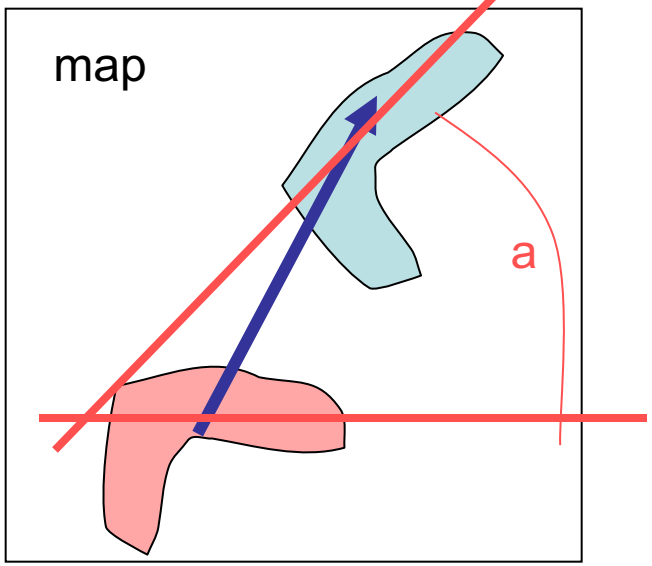
No correction

Correct TF found, very little contrast

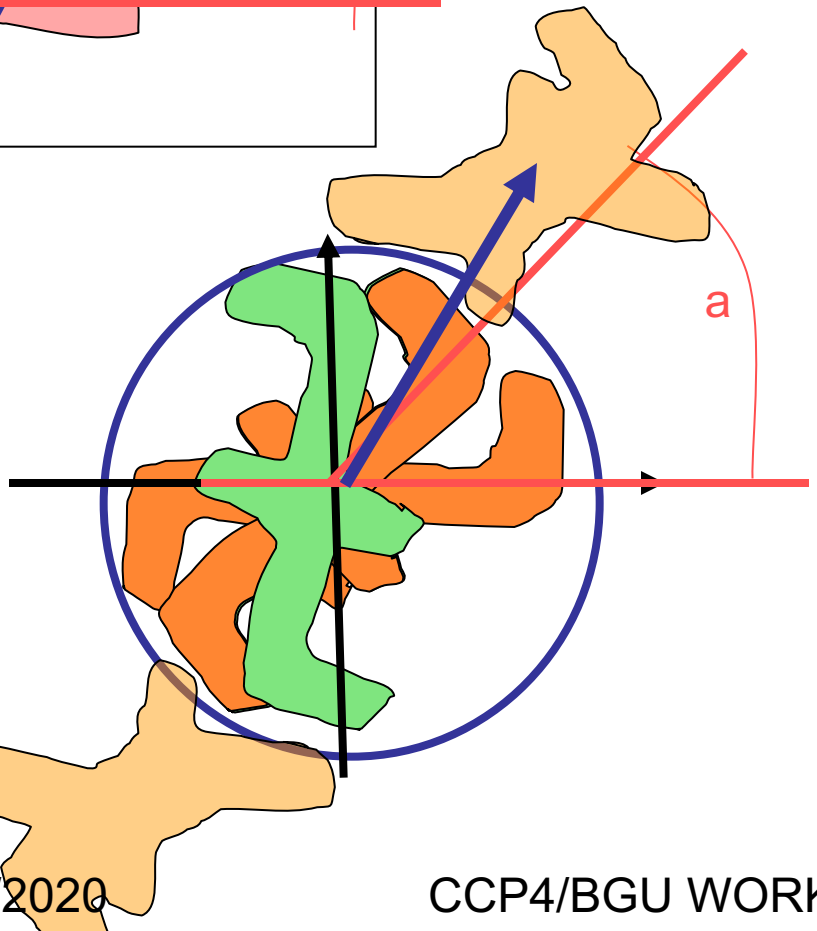
PHASER is less affected

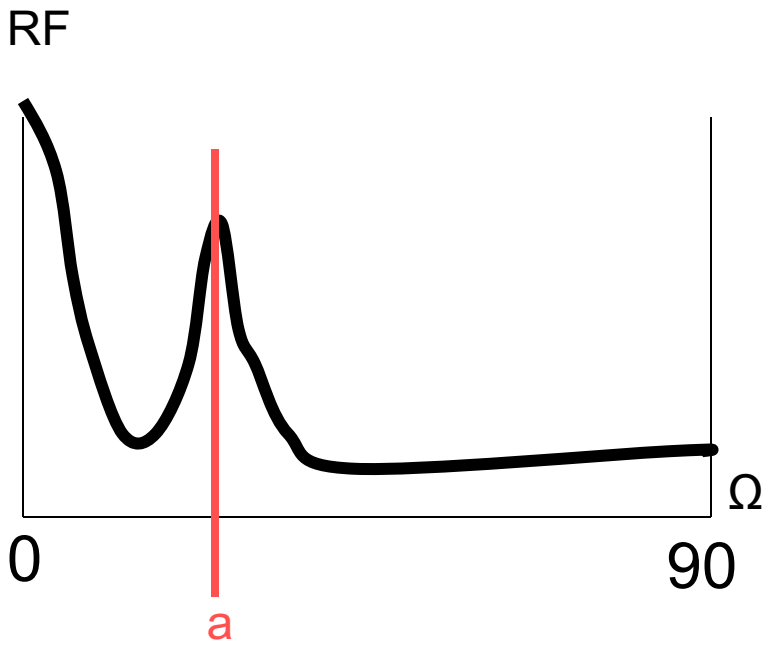
	RF	TF	theta	phi	chi	tx	ty	tz	TF/sg	wRfac	Score
1	1	1	34.27	116.59	99.25	0.475	0.440	0.205	5.15	0.683	0.11021
2	5	13	149.71	-68.68	153.17	0.000	0.147	0.448	4.19	0.687	0.10395
3	7	3	149.69	-95.68	174.41	0.095	0.376	0.204	4.21	0.680	0.10385
4	6	8	135.07	-72.35	124.14	0.018	0.402	0.417	4.79	0.691	0.10316
5	2	5	146.36	-62.44	142.75	0.454	0.422	0.253	4.69	0.688	0.09854
6	9	4	46.43	58.19	119.28	0.252	0.447	0.486	6.46	0.697	0.09835
7	3	14	142.88	-54.87	134.42	0.084	0.296	0.326	4.76	0.701	0.09503
8	4	8	143.11	-53.26	131.45	0.338	0.045	0.066	5.31	0.697	0.09413

Self Rotation Function

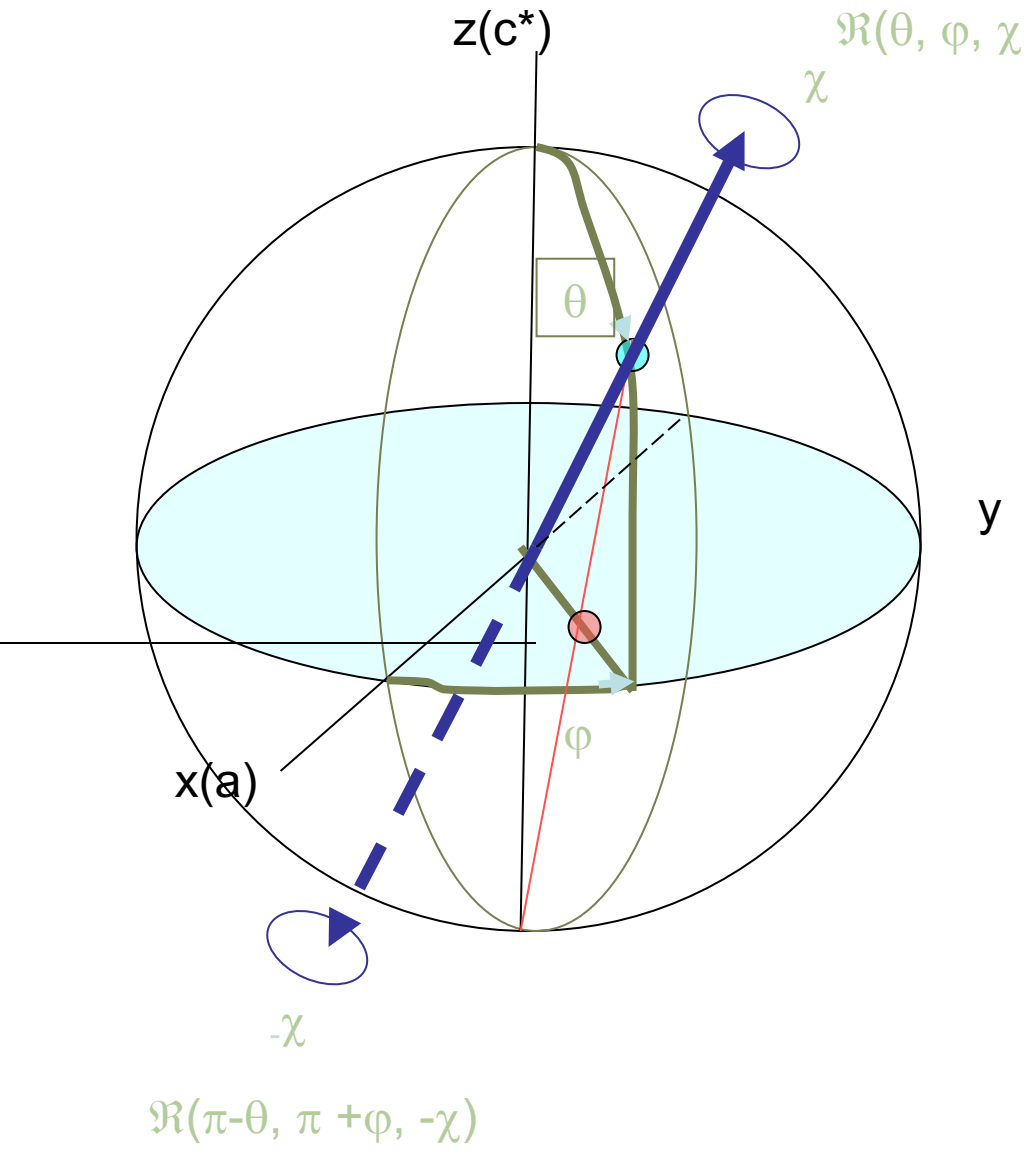


$$RF(\Omega) = \int P_{obs}(r)\ \Re_{\Omega}\{ P_{obs}(r)\}\ dr$$

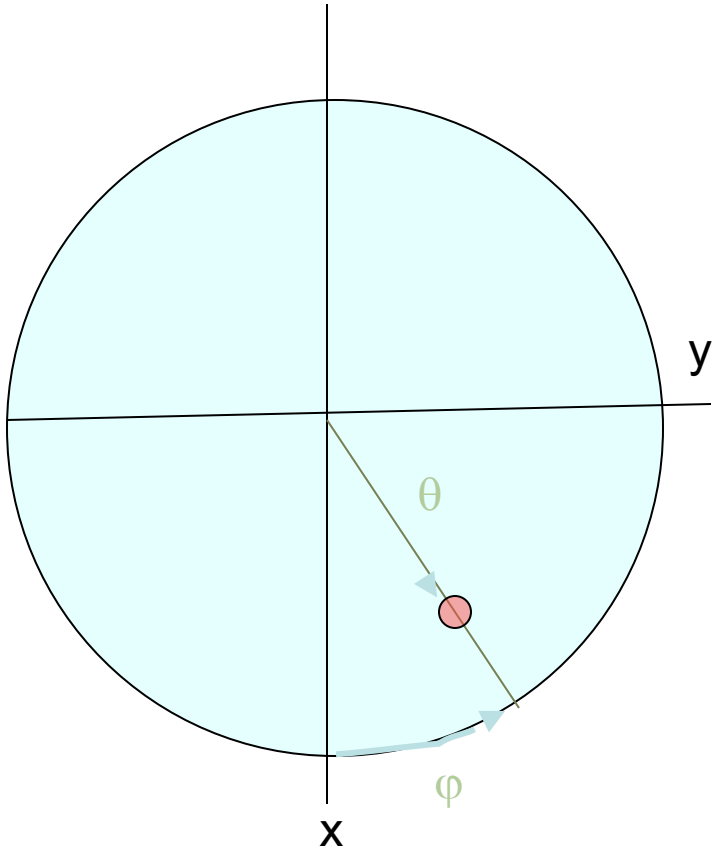




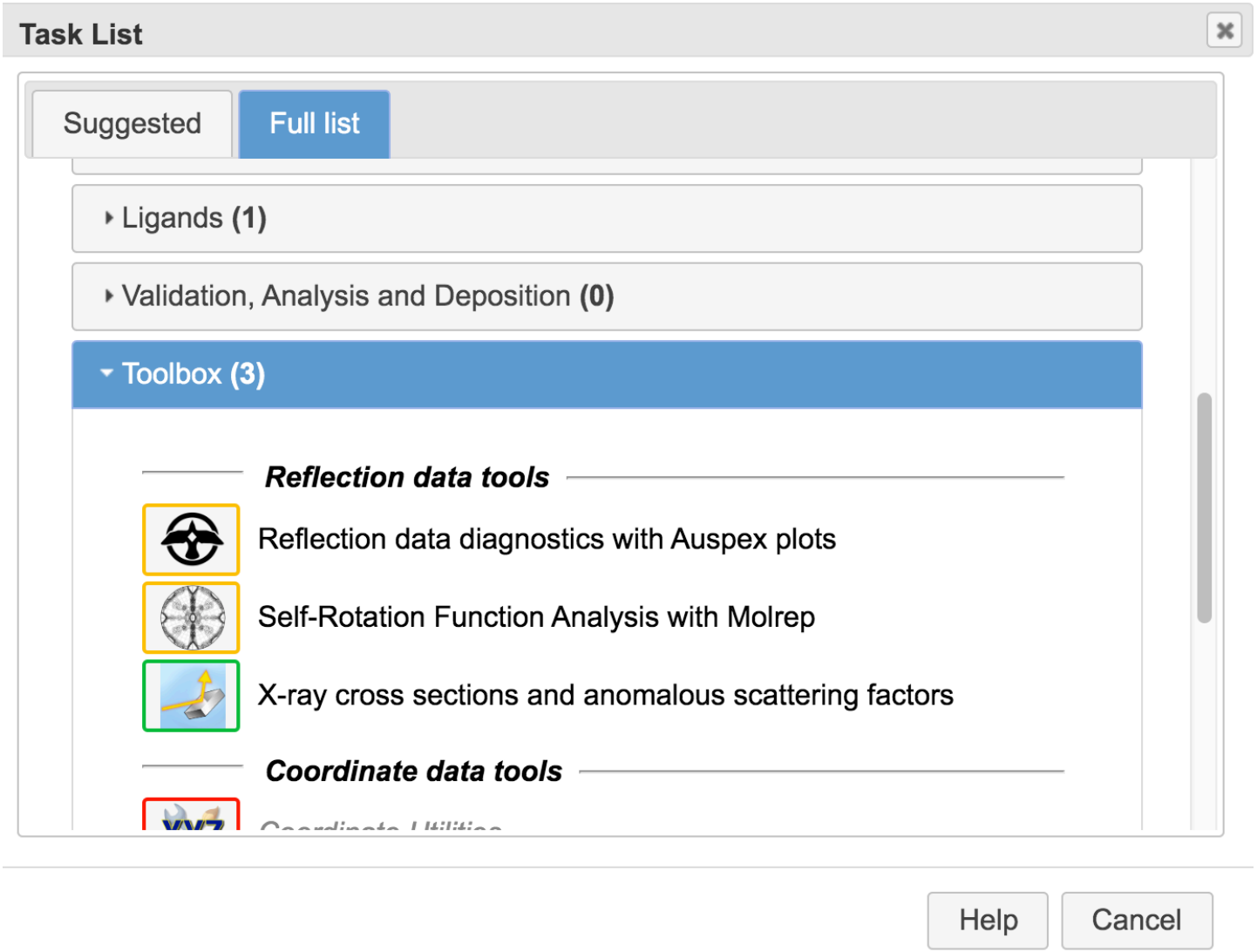
Self Rotation Function



Plot for rotation by χ
(stereographic projection)

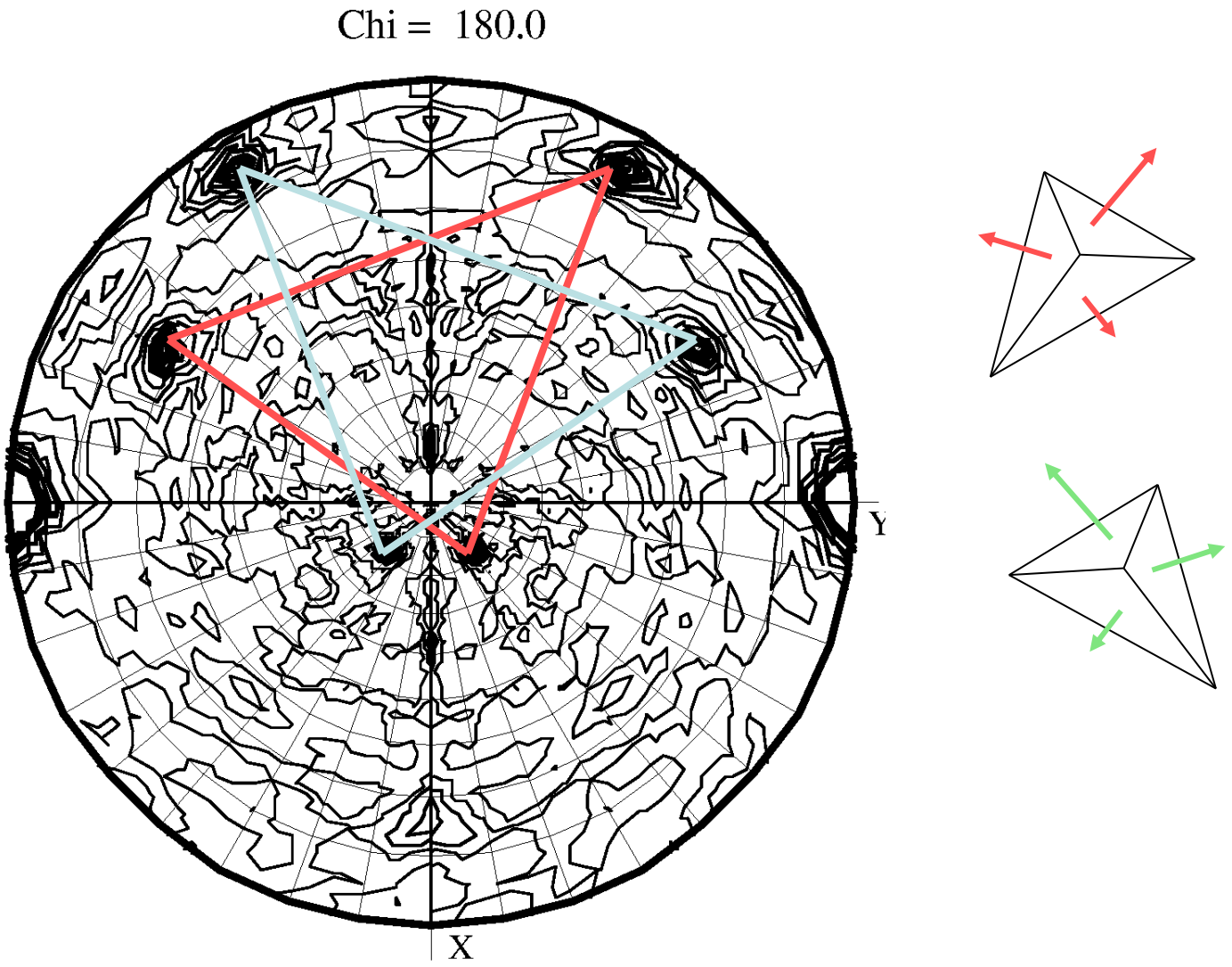


Self Rotation Function



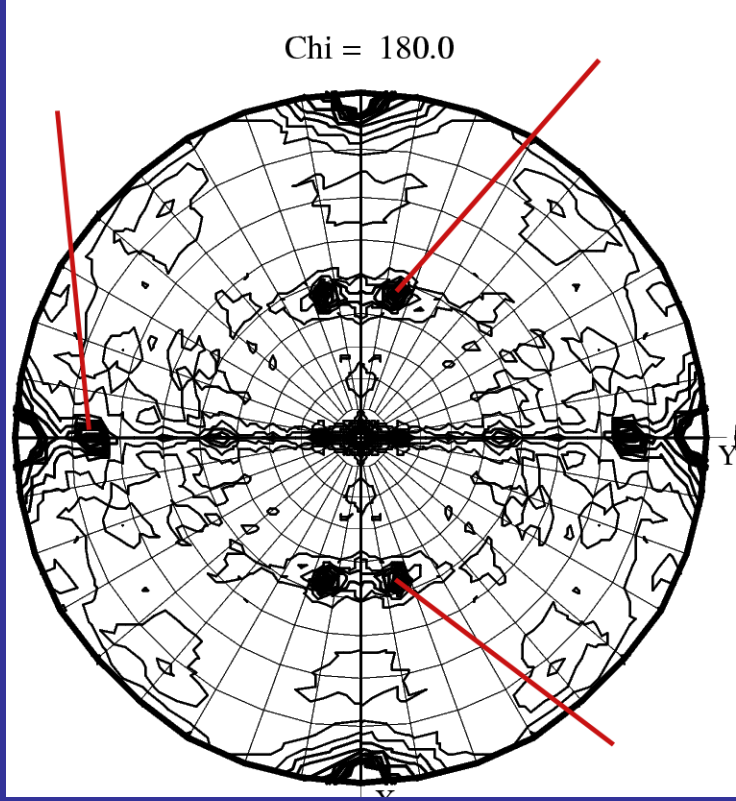
Self Rotation Function

Space group $P2_1$ one tetramer point group symmetry 222 in a.u



SELF-ROTATION FUNCTION.

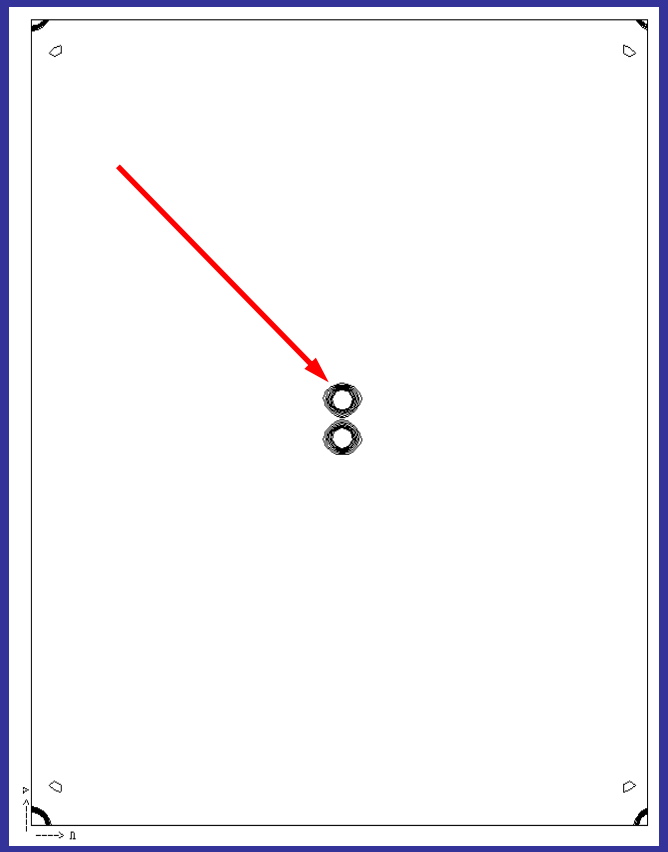
- Can indicate the point group symmetry of the biomolecule.
- Can help to limit the search space for MR.
- SRF calculated using F_{calc} from the putative MR solution should be similar to that of F_{obs}. Otherwise this MR solution is wrong, provided peaks in SRF are not artefacts.

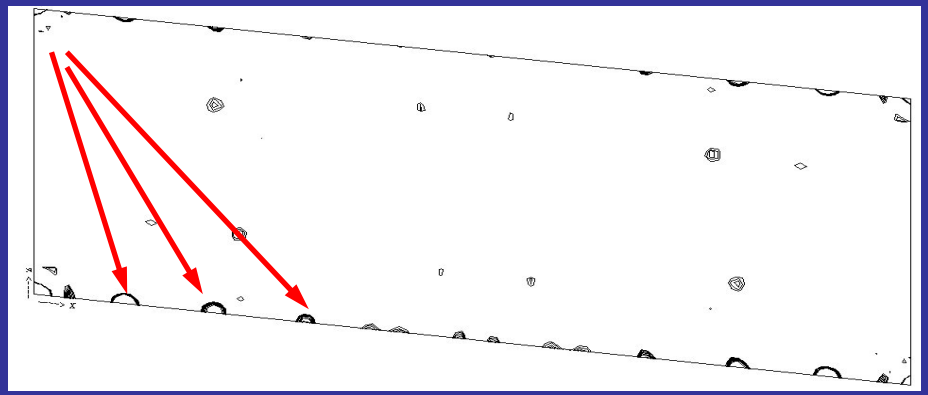


Section Chi=180° of self-rotation function of tetrameric tryptophanase indicates molecular symmetry 222 in the space group $P2_12_12_1$.

NATIVE PATTERSON SYNTHESIS INSPECTION

- Peaks with height comparable with origin (in MOLREP 0.15 of origin) are indicative of pseudo-translation which often complicates MR – can give unreasonably high CC for the wrong solution.
- Also may result in the wrong SG assignment.
- May indicate crystal disorder (OD-structure) and occurs in some types of twinning.





Structure solution of the oxygenating subunit of 3,6-diketocamphane monooxygenase from *Pseudomonas putida*

- 42kDa FMN binding enzyme.
- Dimer from size exclusion chromatography.
- No cloned gene.
- Space group $P2_12_12_1$,
- 50% solvent (native crystal).
- Variable cell parameter c.
- Native data to 2 Å resolution. a= 55.0, b=93.4,c=162.0 Å.
- 1 M NaBr soak. a=54.9, b=93.3, c=140.8 Å.
- 3 wavelength MAD data to 2.5 Å.
- No solution for anomalous substructure.

Isupov et al. (2015) Acta D 71, 2344-2353.

- Homologue – bacterial luciferase from *Vibrio harveyi*.

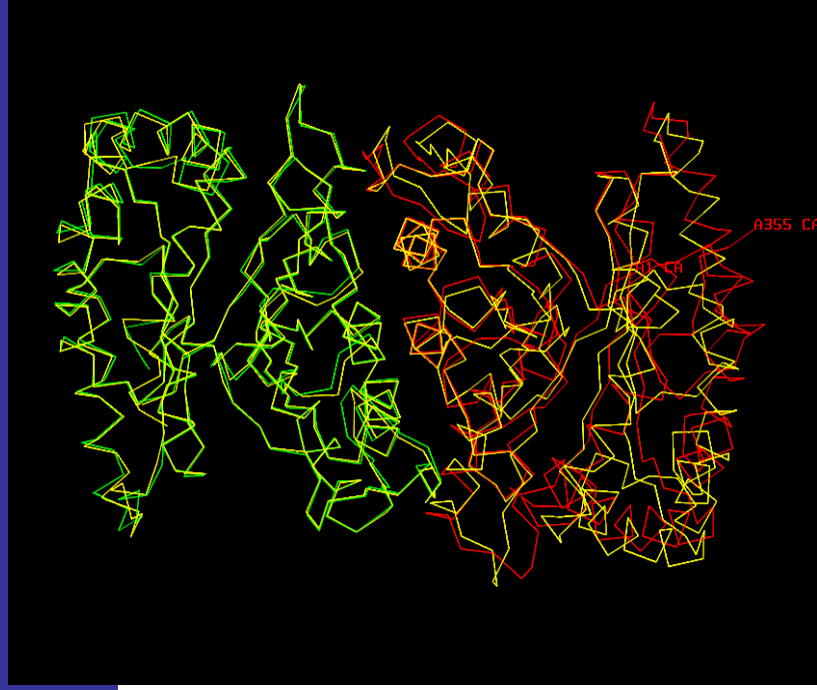
- 17% sequence identity.

- No MR solution for a luciferase monomer.

- $\alpha\beta$, β_2 ---> α_2

- MOLREP with α_2 dimeric model, native data.

- No rotation solution.



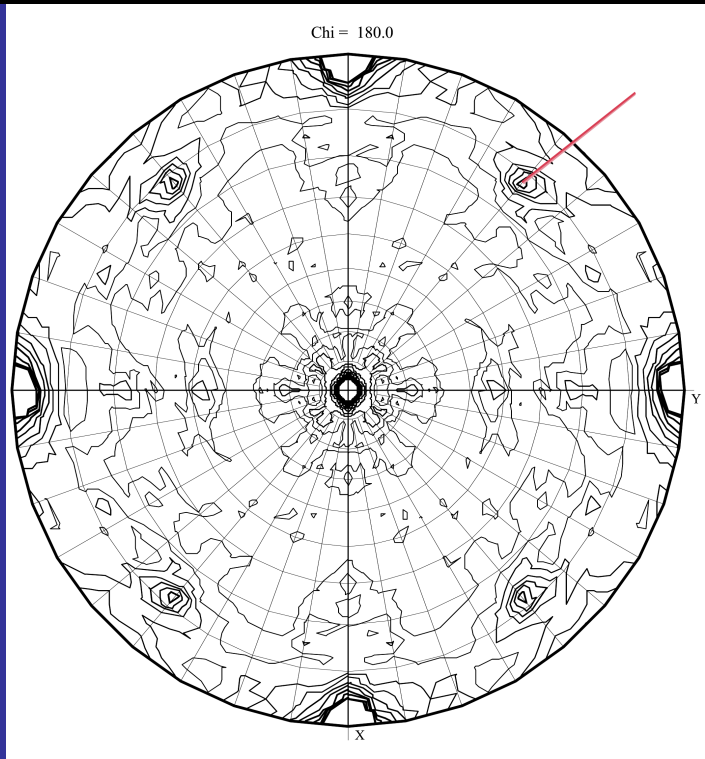


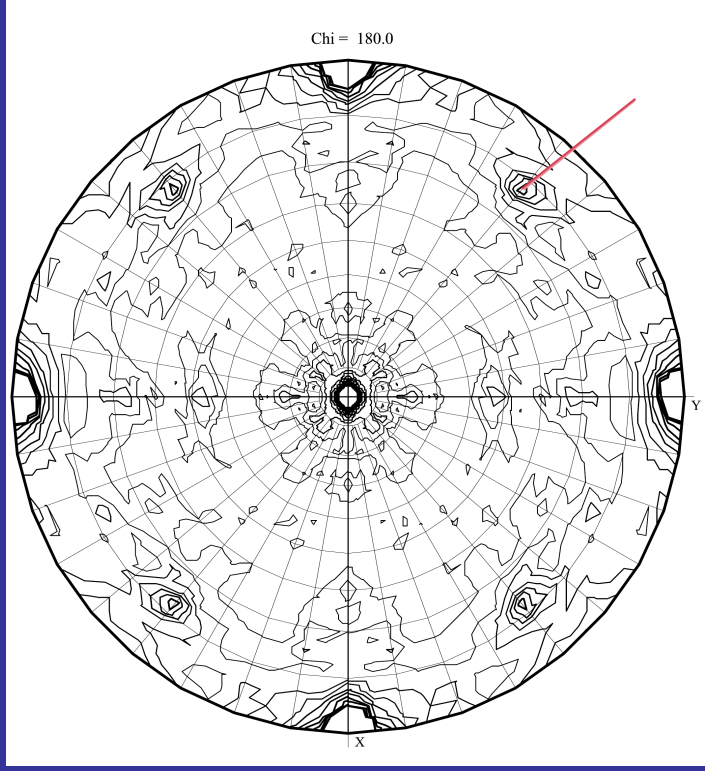
Table of rotation peaks instead of rotation search.

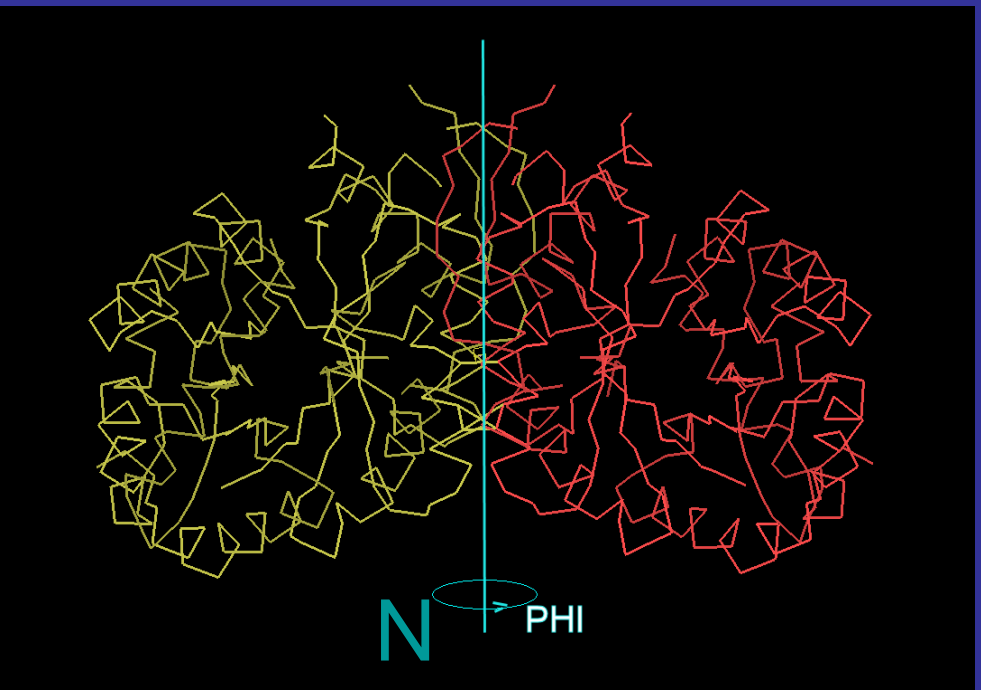
- N-termini top and bottom.
- PHI 0-180° with 2° step – 90x2.

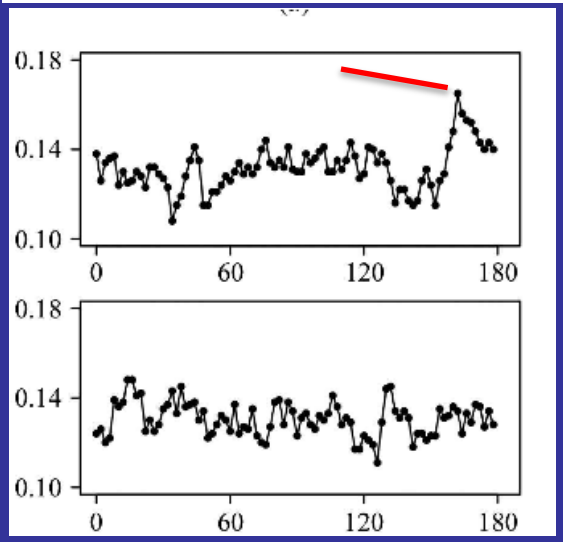
180 TF calculations.

One orientation better for all models and resolutions – no clear translation.

Clear translation solution after rigid body refinement of monomers in P1.







correlation

PHI

Restrained refinement REFMAC, averaging DM, phased refinement – difficult electron density.

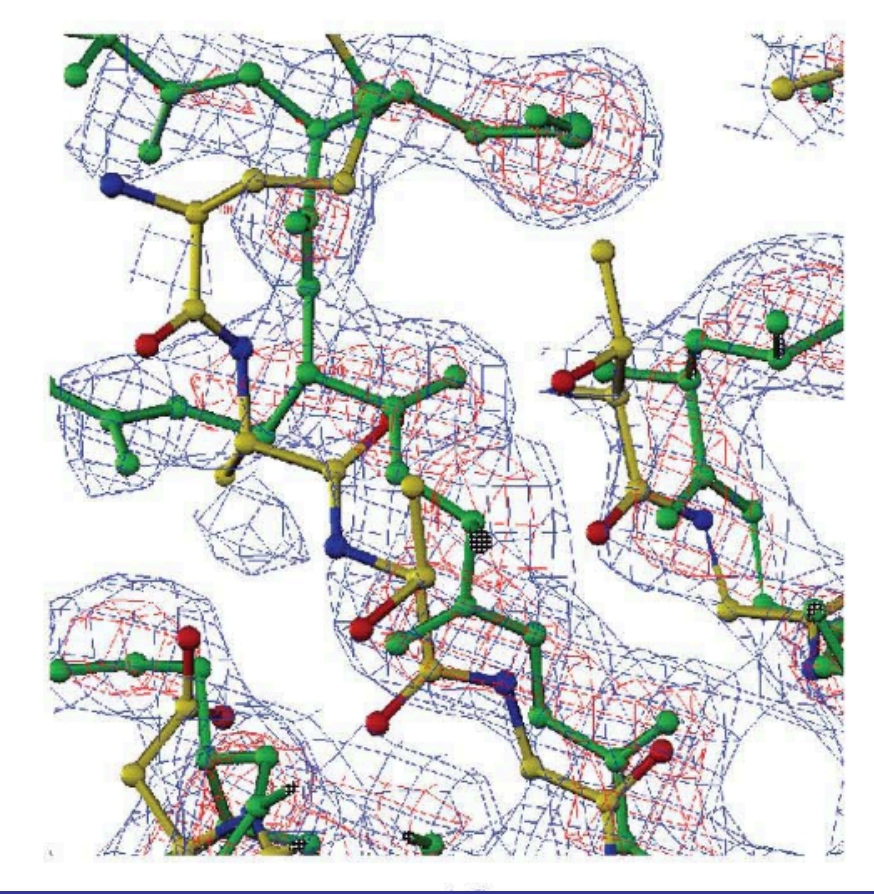
Successful MR against Br data with partially refined model.

Multicrystal averaging DMMULTI.

18 Br sites 14-6s found in anomalous difference Fourier at 10-3 Å (peak wavelength).

3 wavelength MAD phasing MLPHARE FOM 0.26 (2.5 Å) and 0.63 (6Å).

DMMULTI:
Model phases (native xtal),
MAD phases (Br soak).



Electron density after DMMULTI with input of model and MAD phases.

Final R-factor 16.3 %, FreeR 23.3 %.

Solution of the structure of anti-TRAP from *Bacillus licheniformis*.

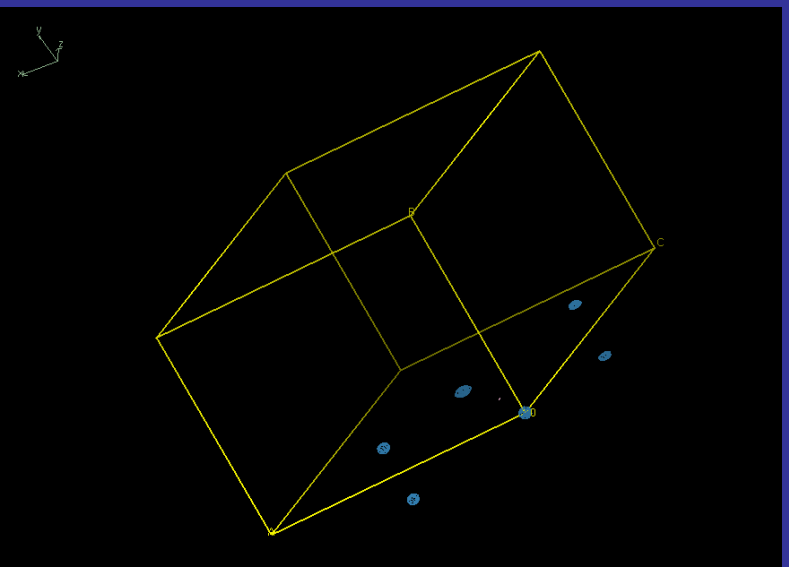
- Anti-TRAP regulates the activity of tryptophan attenuation protein (TRAP) in *Bacilli*.
- 53 amino acids.
- Space group $P2_1$.
- a=118.5, b= 99.8, c= 123.2 Å , β= 117.6°.
- Data to 2.2 Å.

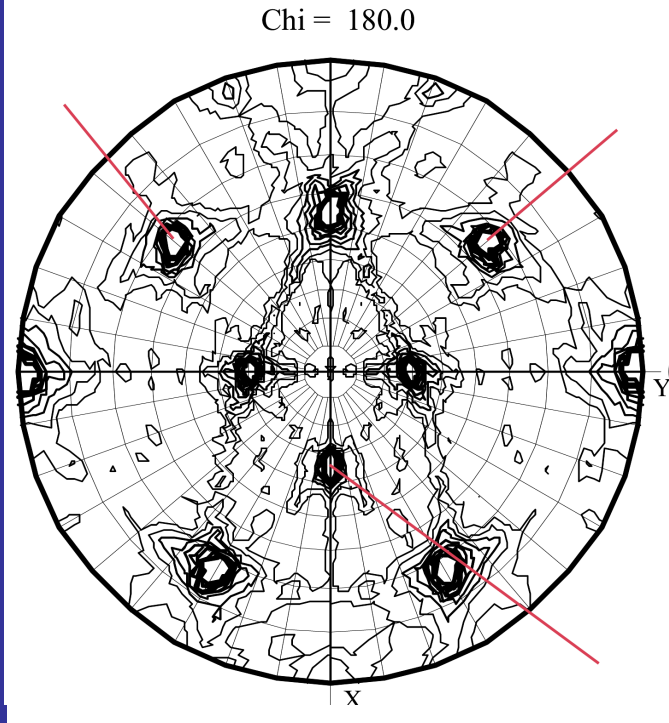
B. subtilis dodecamer has cubic 23 symmetry.

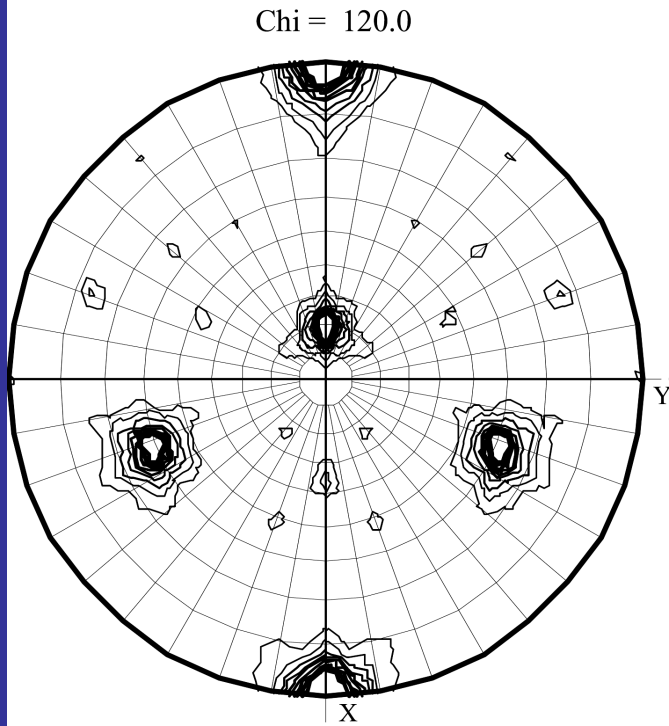
Model

- *Bacillus subtilis* anti-TRAP is a 12-mer with 23 point group symmetry.
- 64 % sequence identity .

- Native Patterson peaks.
- (0.5 0.13 0.0) 0.4 origin
- (0.5 0.0 0.5) 0.4 origin
- (0.0 0.13 0.5) 0.16 origin
- SRF – 23 point group symmetry, apparent 432 due to special orientation of 12-mer in relation to crystal dyad.
- A.u. contains 3 or 4 12-mers with 23 point group symmetry related by pseudotranslation.







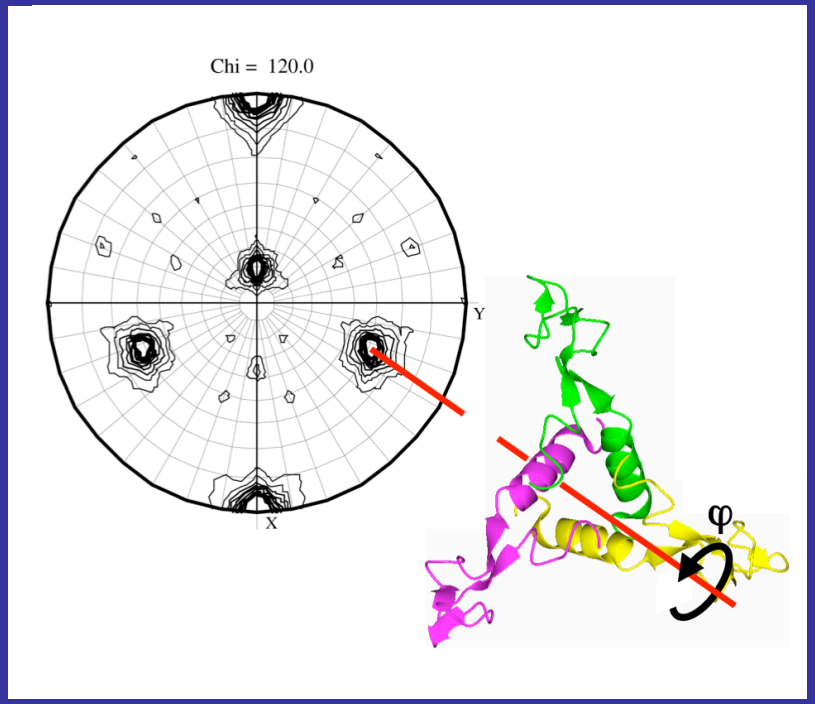
- MOLREP.
- No solution for a dodecamer model.
- No solution for a monomer.

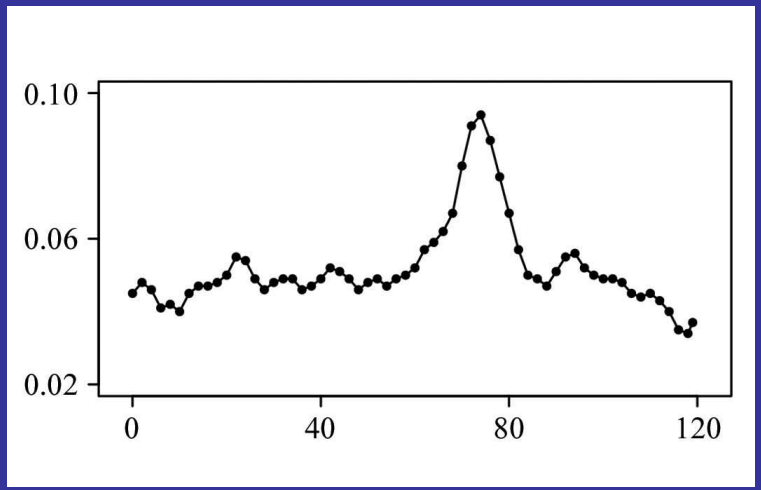
- Evidence of trimeric species from analytical ultracentrifugation and SEC.

- No rotation solution for a trimer.

Trimeric model.

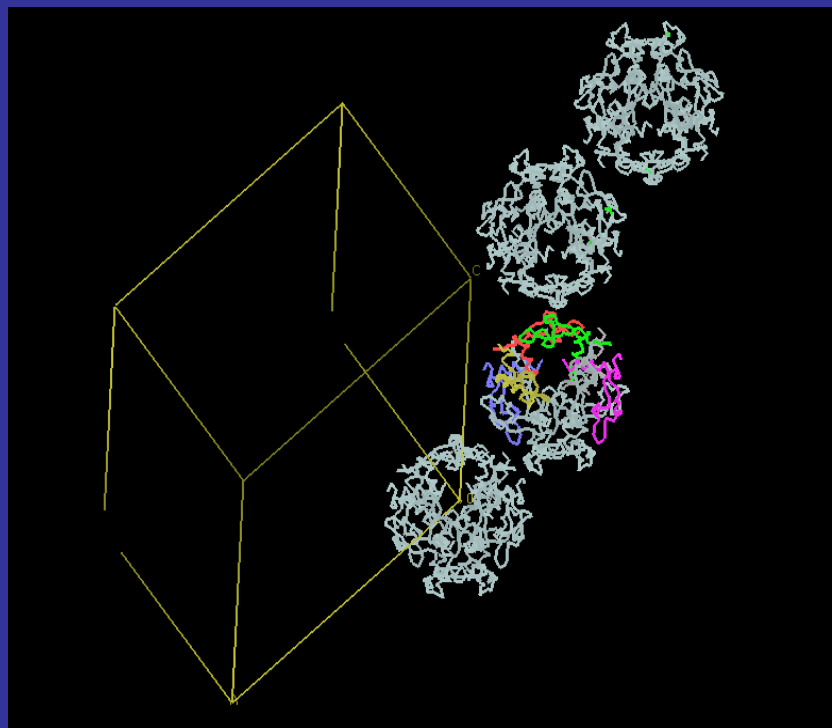
- Table of rotation peaks instead of rotation search.
- An increment of 2 degrees in the range 0-120 degrees.
  60 runs of translation function.
- Due to special orientation of the dodecamer in relation to crystallographic 2_1 axis the trimer does not need to be turned over.
- Found solution was fixed and the search repeated for another NCS three-fold until one full dodecamer was build.
- Resulting dodecamer was used as a model for MR.
- Clear translation peaks.



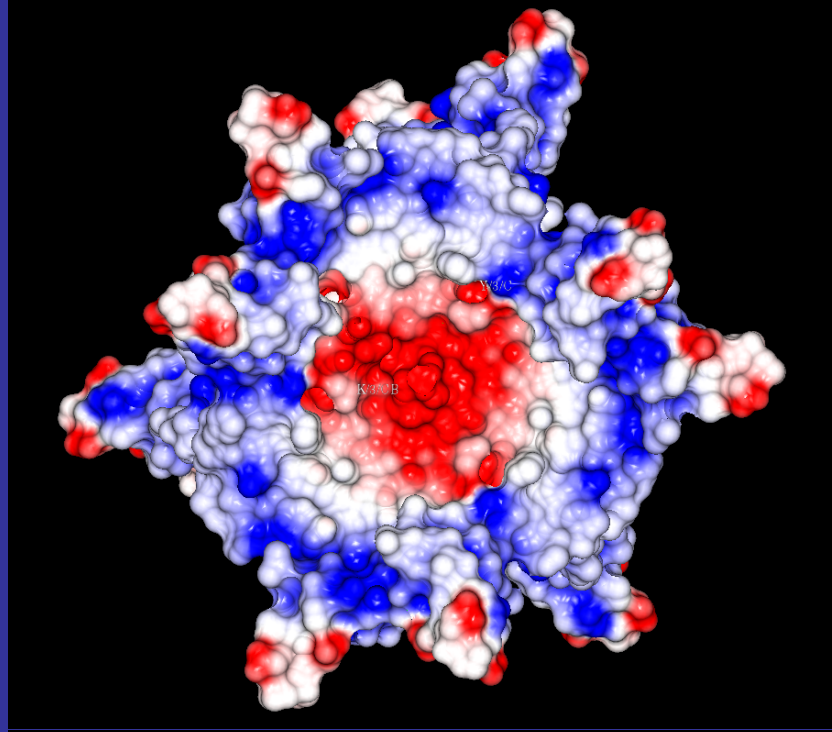


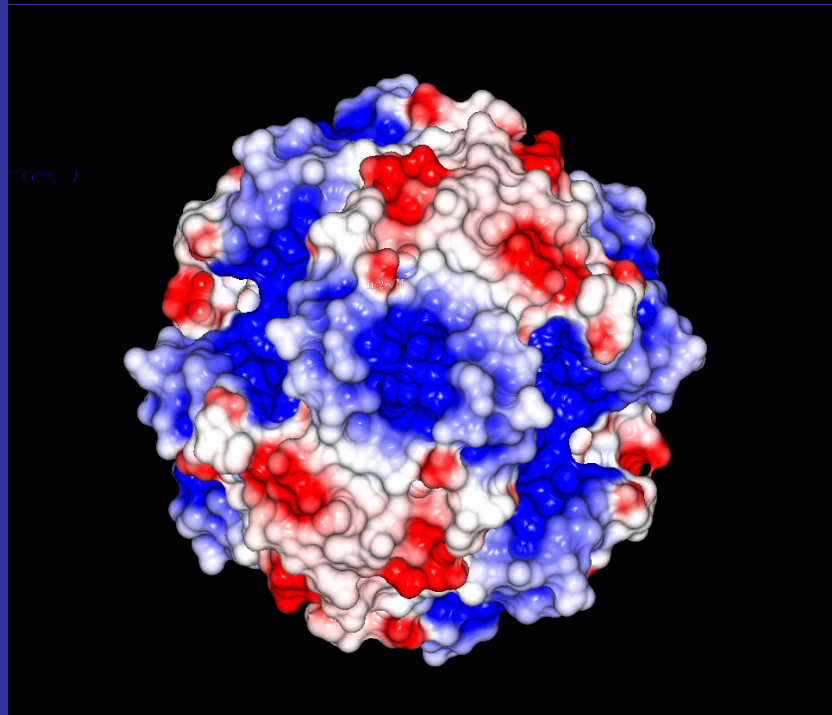
- Final model contained 4 dodecamers in asymmetric unit, related by pseudotranslation.
- It was subjected to restrained refinement in REFMAC.
- However FreeR did not decrease lower than 43 %.
- Main chain breaks in 2Fo-Fc map.
- One dodecamer with fewer main chain breaks in the density was resubmitted to MR. The resulting structure easily refined to R_{cryst} 19.7%, FreeR 25.4 %.

- ZANUDA later revealed origin assignment – related error in the original TF.



- The resulting 12-mer is different from the original one.
- Later, another crystal form was obtained which contained *B. licheniformis* 12-mer of *B. subtilis* type (Antson and Shevtsov, private communication).
- Depending on yet unknown environmental conditions the two dodecamers appear to interconvert.





Shevtsov et al. (2010) J Struct Biol 170, 127–133

Solution of the structure of peroxiredoxin 2 from human erythrocytes

- Antioxidant enzyme, PRX.
- 22 kDa monomer.

- Purified from old blood packs.

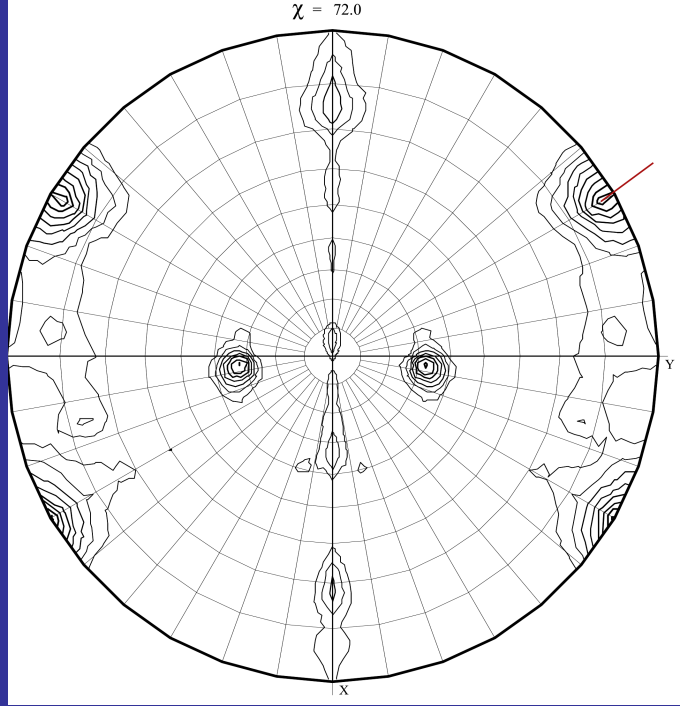
- Symmetry $P2_1$.
  a=88.9, b=107.0, c=119.5Å, b=110.9˚ .
- Native data to 1.7 Å.

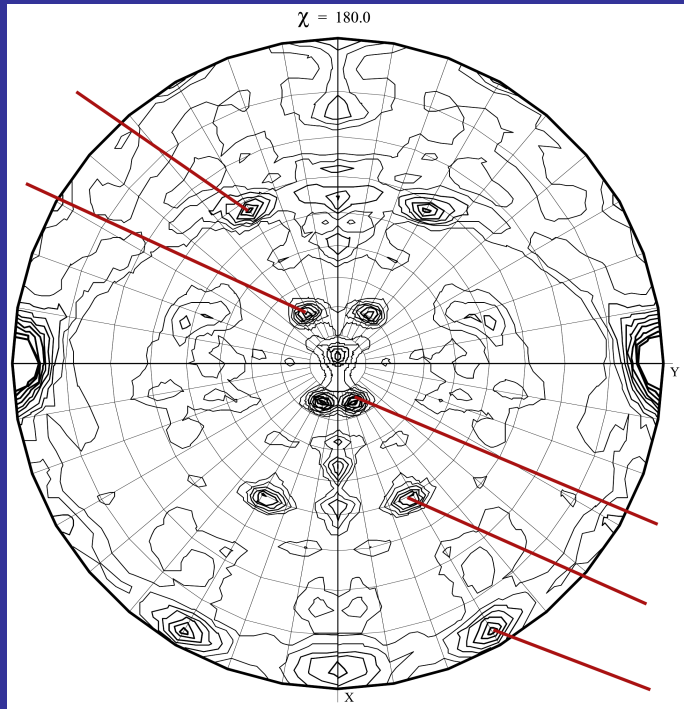
- Poor native crystals isomorphism prevented heavy atom location
- Model - dimeric PRX6.
- 30% sequence identity.
- No MR solution for dimeric model.

PRX6 dimer

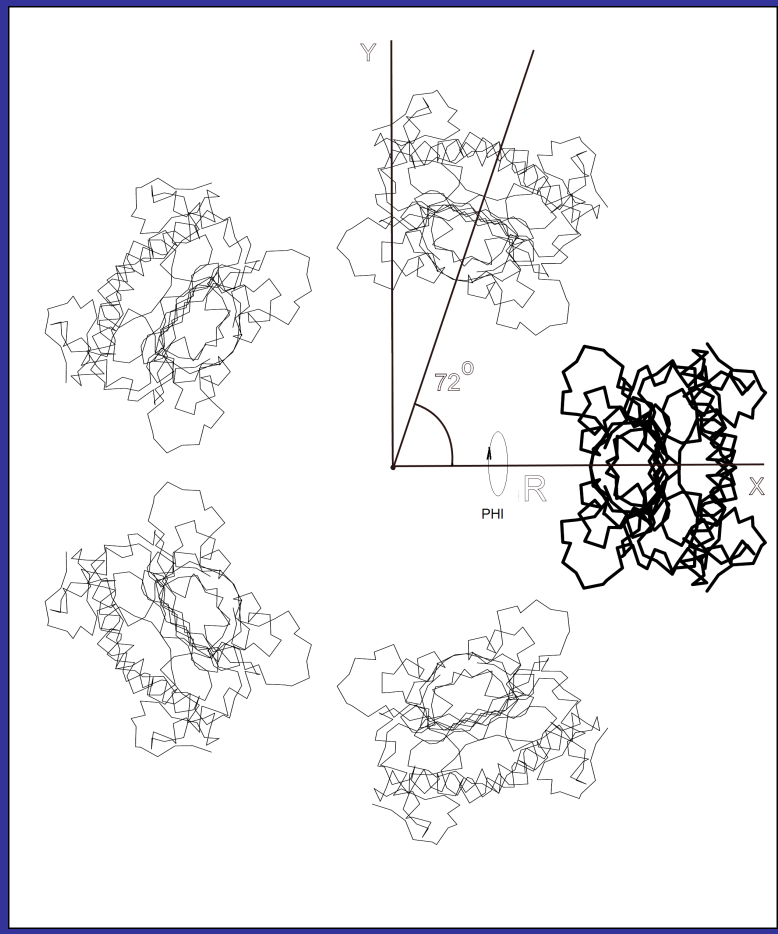
- Analytical ultracentrifugation, size exclusion chromatography and crystal packing results inconclusive: 6-14 subunits.

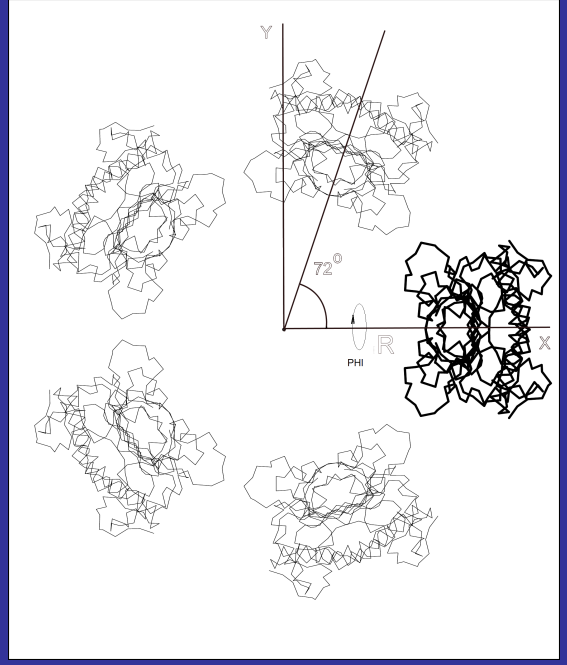
- Self-rotation (MOLREP) revealed that PRX is a decamer with molecular symmetry 52.

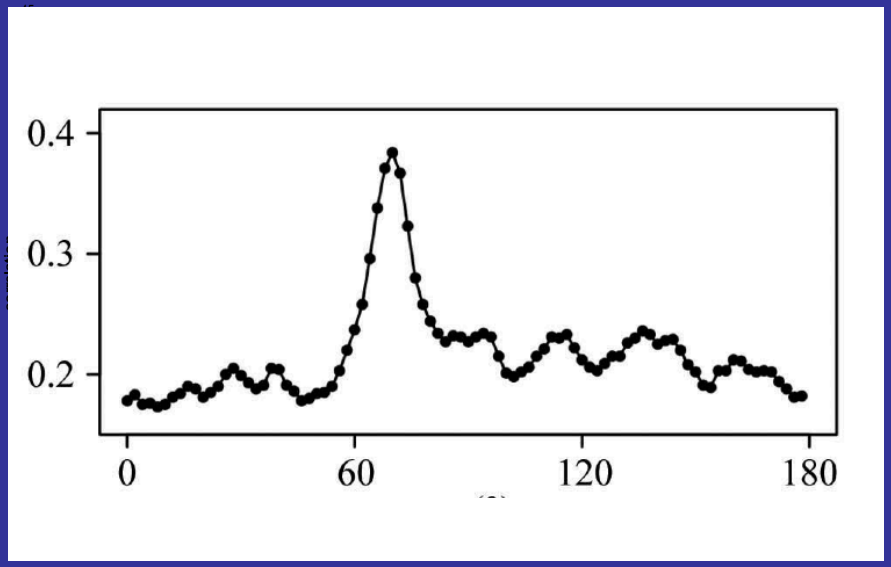




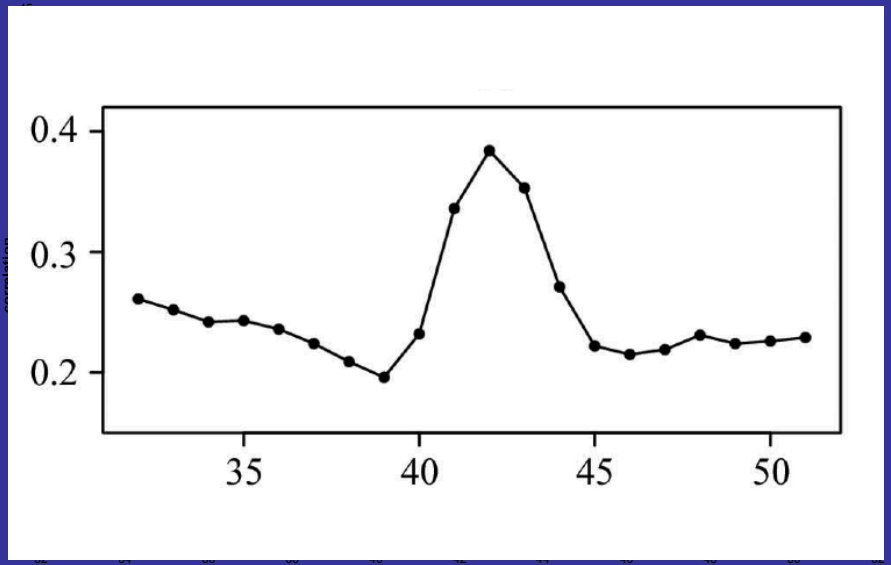
- Polyalanine dimer.
- Amino acids 1-189 out of 224 to cut off poorly conserved domain.
- All possible decamers with point group symmetry 52 were generated using this dimer.
- Radius limits (32-52) and (-52 - -32).
- Angle limits 0-180˚
- 3600 models.







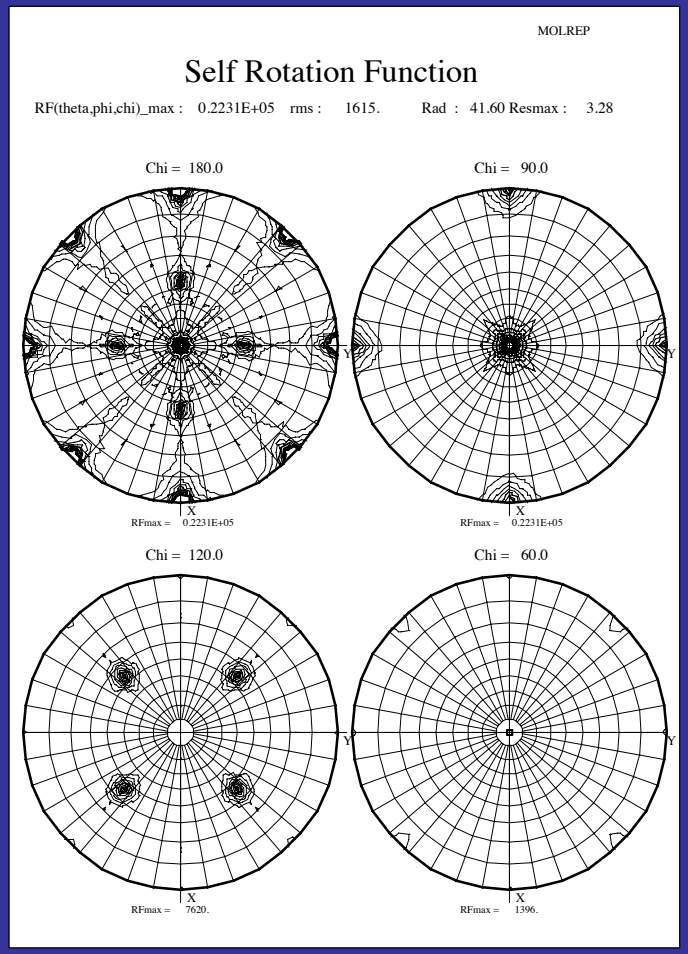
- Alignment of decamer dyads to NCS dyads.
- Translation search for 3600 models.
- AMORE
- 10-fold averaging by DM.
- Refinement (REFMAC) using external (averaged) phases.



Final Rcryst 19.2 %, FreeR 25.6 %.

Schroder et al. (2000). *Structure with Folding & Design*, **8**, 605-615.

Warning: Self-rotation function can be misleading



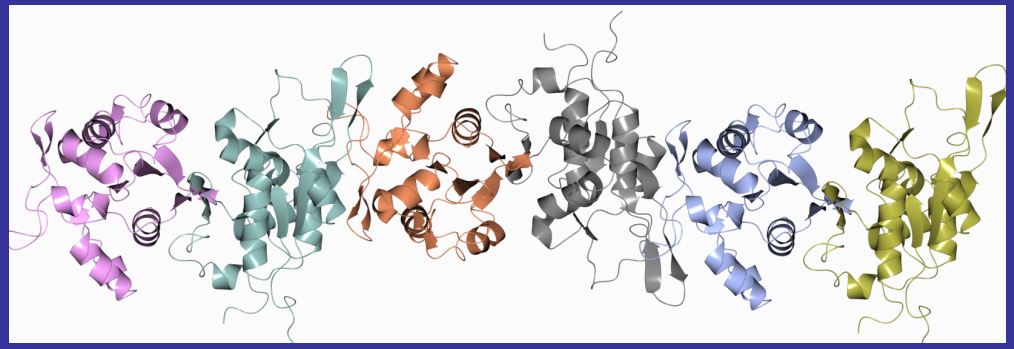
Solvent boundaries can introduce artifacts, also special arrangement of molecules

Space group $P4_12_12$

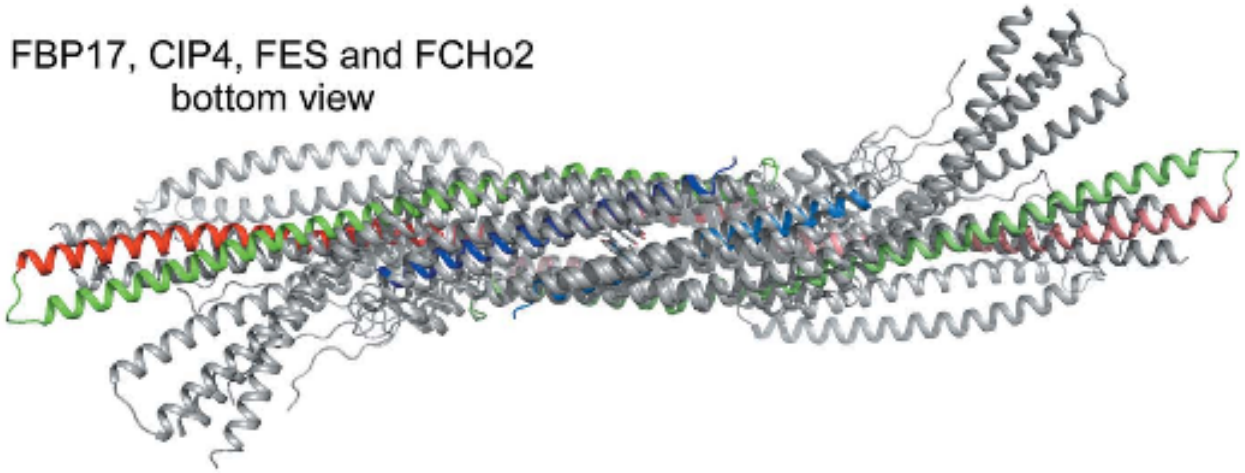
Apparent NCS symmetry 432

Impossible due to packing.

2 monomers per a.u. are related by 90° rotation and ¼ fractional translation creating 4_1 screw columns with symmetry molecules along a and b crystal axes.



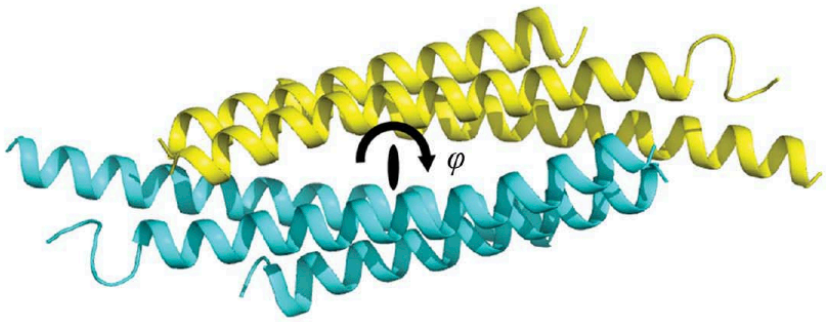
Human SrGAP2 with Dr Yarden Opatowsky
Molecular symmetry without NCS

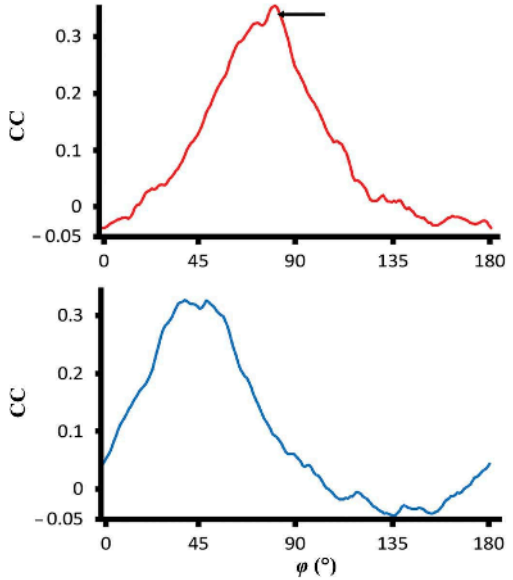


- Space group C2, one monomer 484 aa per au.
- Coiled coil protein
- Conserved central six-helical core of the dimer.
- Centre of mass fixed, unknown rotation around crystallographic dyad.
- Nearest homologue 19% sequence identity over 60% of length

Sporny et al. (2017) *Mol Biol Evol* **34**, 1463-1478.

Exhaustive search on a crystallographic dyad.





- Easy MR, but difficult refinement
- Space group C2, one monomer per a.u.
- Conserved central six-helical core of the dimer.
- Centre of mass fixed, unknown rotation around crystallographic dyad.
- One-parametric exhaustive search with minimal model using TF as score function (packing function switched off in MOLREP)

Importance of weak anisotropic data

Problem with model rebuilding and refinement starting with partial polyalanine model (171 aa out of 484). Starting mean phase error of 87.6° to 3Å (84.6° to 5.5Å).

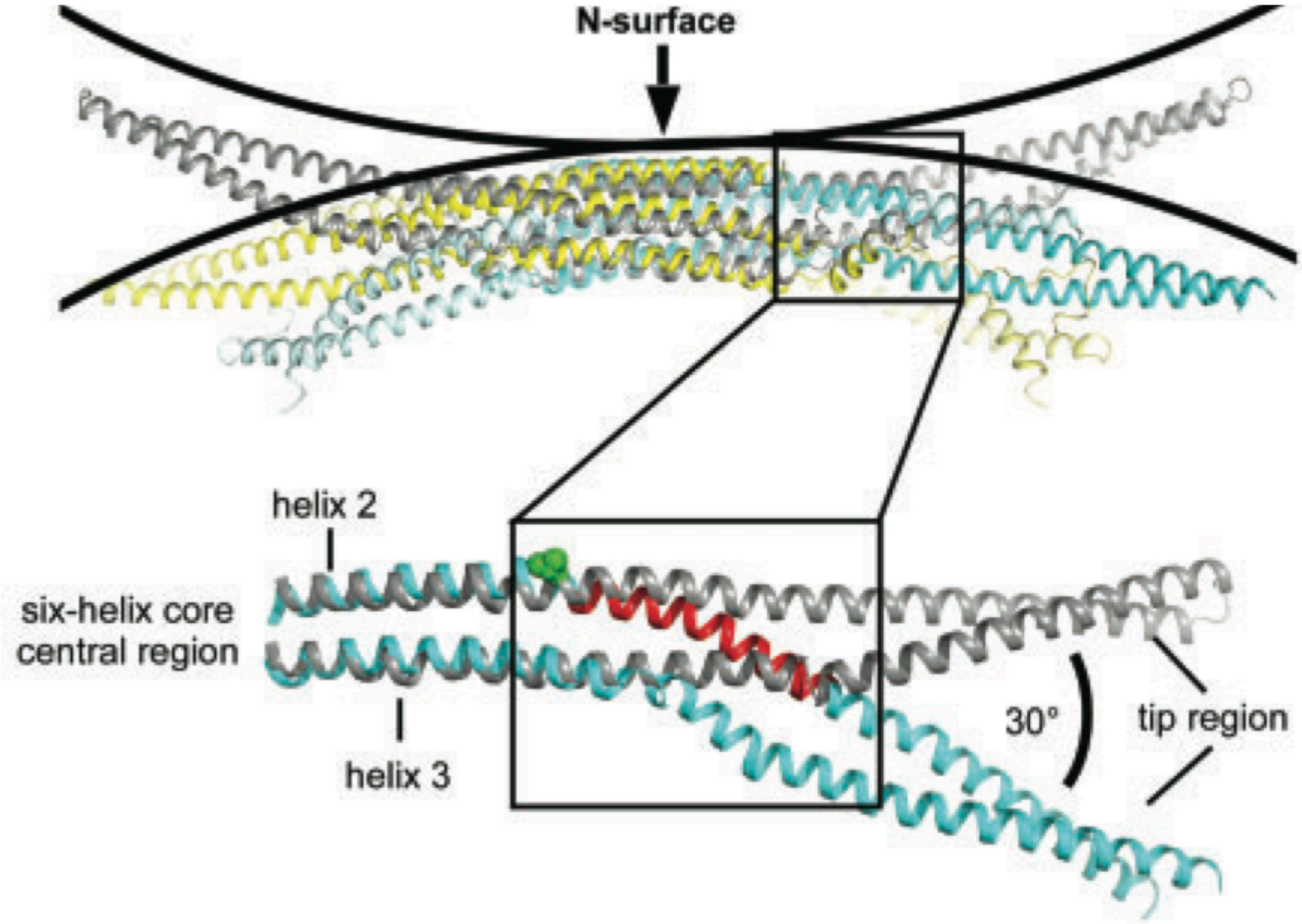
Multicrystal averaging and phased refinement in REFMAC5.
Data from two poorly isomorphous crystals in space group C2
1) a = 203.8, b = 29.9, c = 95.0Å, β= 91.9 to 2.2 Å ($CC_{1/2}$=0.3)
or 2.7 Å ($\langle I \rangle / \langle \sigma(I) \rangle = 2$)
2) a = 216.9, b = 29.6, c = 94.7Å, β= 92.0 to 2.9 Å

Two more helical stretches added - mean phase error 82.4° to 3Å , after refinement and helix idealization - 75.4° to 3Å 6 cycles of SHELXE model autotracing at 2.2Å improved phases to 66° , no improvement was observed for the same run at 2.7Å

Hierarchy of phase improvement approaches in SRGAP2 case (weak 2.2 Å) good 2.7 Å

- Multicrystal averaging/ phased refinement (DMMULTI/REFMAC) of a partial model is improving phases at any stage, provided initial phases were correct
- SHELXE model autotracing is not as powerful at lower resolution in the beginning of refinement/model improvement, and phase improvement requires better initial phases
- ARP/wARP and Buccaneer procedures are extremely useful e.g. for sequence assignment but require significantly better starting phases at medium resolution

The structure is unlikely to be solved by a full model.



Acknowledgements

- Fred Antson York, UK
- Garib Murshudov LMB, UK
- Ronan Keegan CCP4, UK
- Andrey Lebedev CCP4, UK
- Jennifer Littlechild Exeter, UK
- Yarden Opatowsky Bar-Ilan University, Ramat Gan
- Alexei Vagin York, UK
- David Waterman CCP4, UK
- Reuven Wiener Hebrew University, Jerusalem