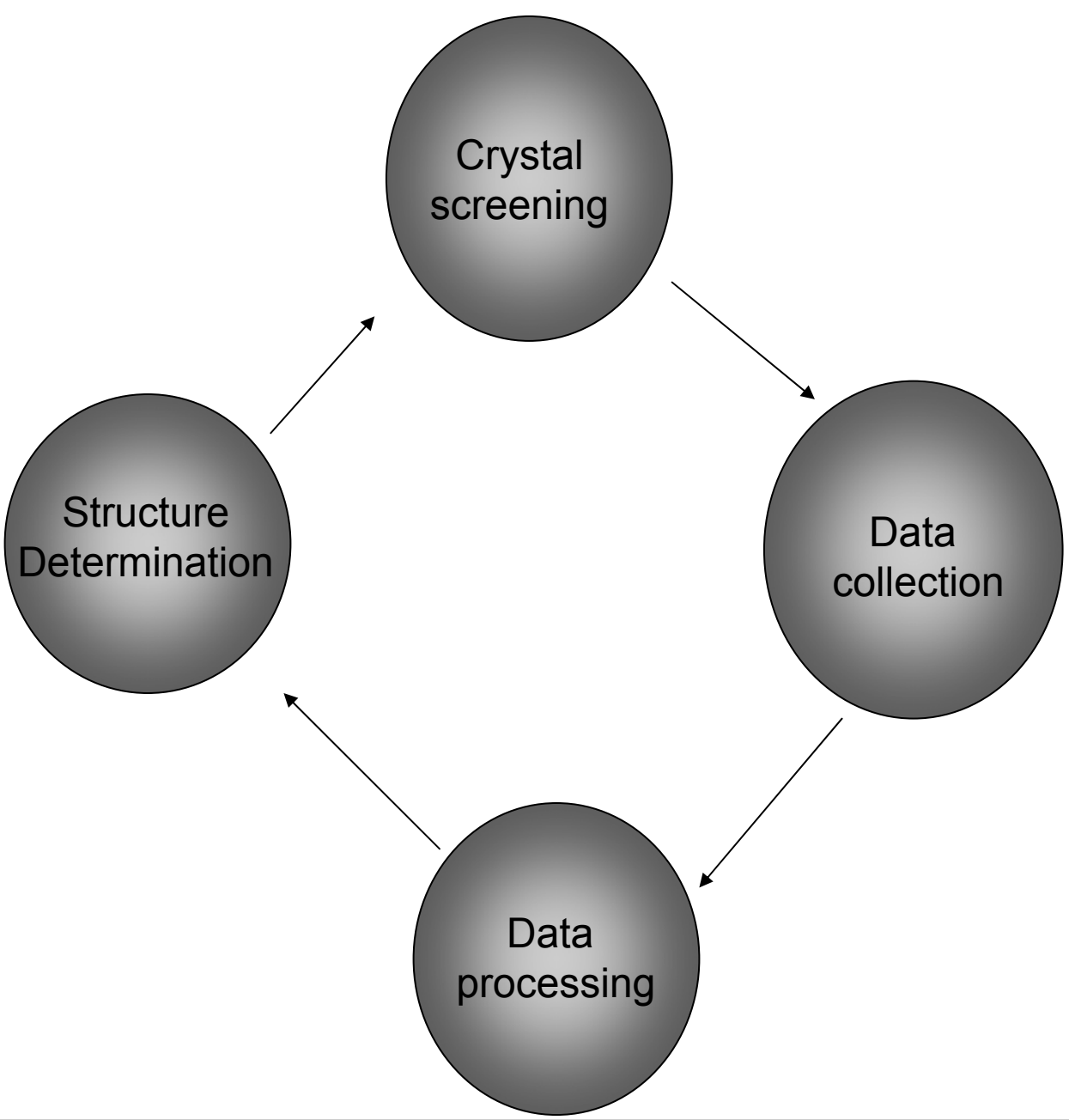
Has my experiment worked? MR

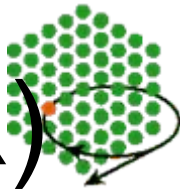
Santosh Panjikar

EMBL-Hamburg Outstation


UNERA
UNERA

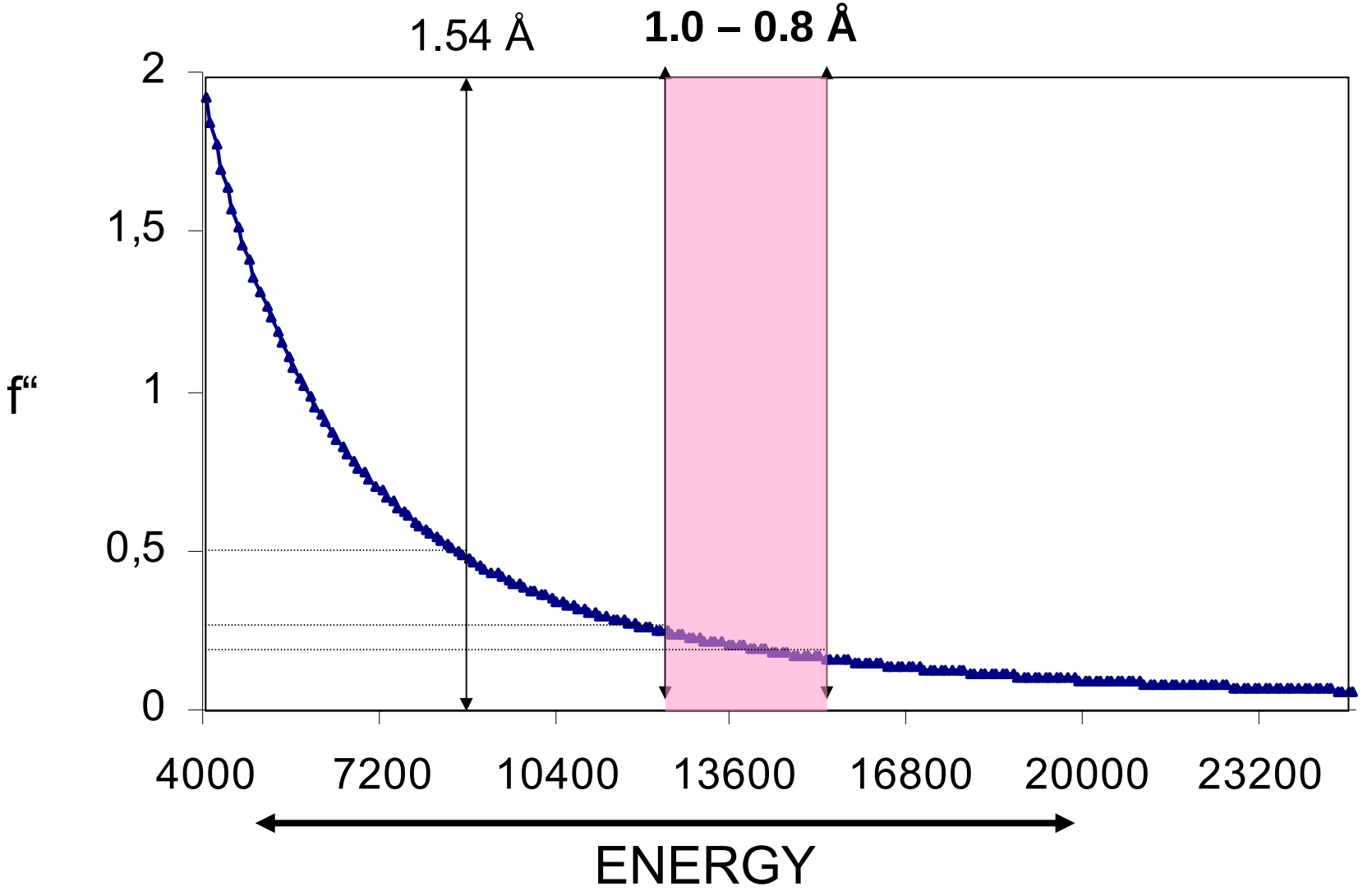

Crystal screening
Data collection
Data processing
Structure Determination

Molecular Replacement (MR)

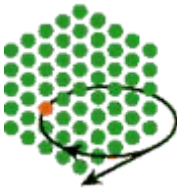


- Require Native dataset
- The coordinates of a homologous protein structure to serve as a search model
- A computer program that determines the orientation and position of the search model(s) in the asymmetric unit.

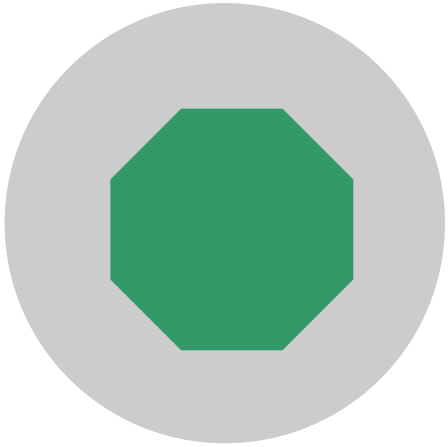
Native data collection



The search model

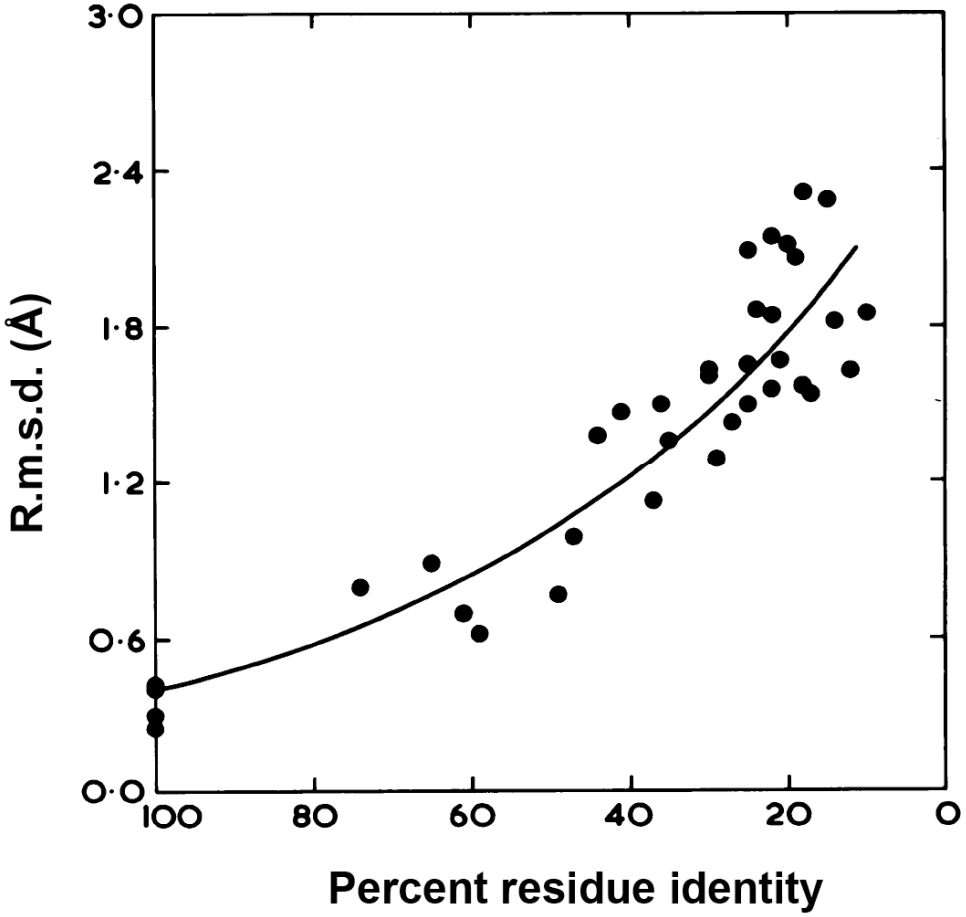


•The higher the sequence identity the easier MR.



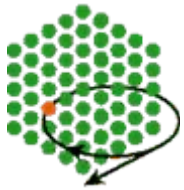
It is not possible to give an minimum value of identity at which MR will work. Below 30% it is usually difficult but it can be challenging with a 100% identical molecule.

Rmsd increases with low sequence identity



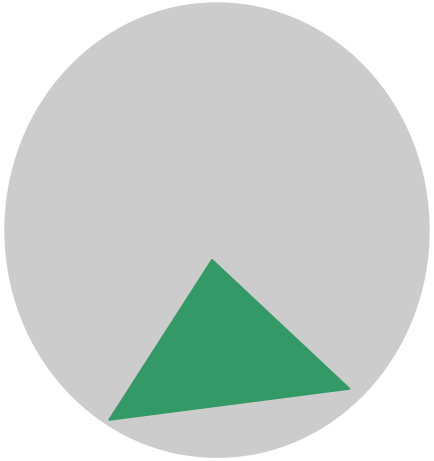
Chothia & Lesk, 1986

The search model



COMPLETENESS

- The search model should represent a significant fraction of the unknown

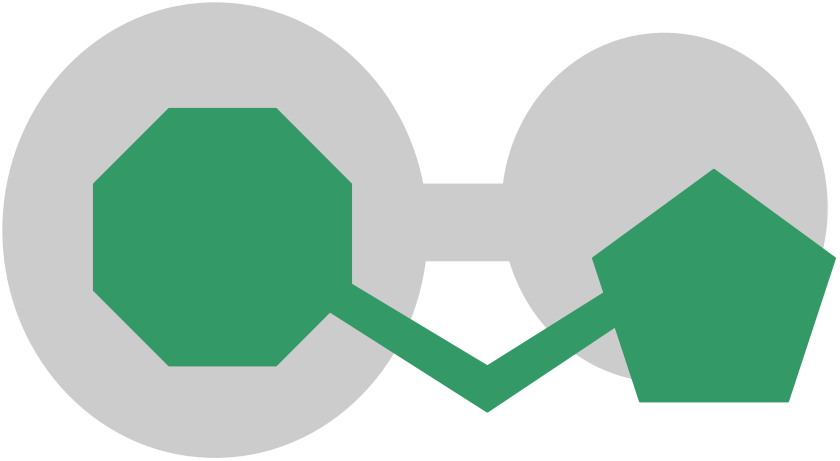


If it only represents 10-20% of the unknown structure MR will be difficult.

The search model

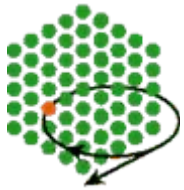
FLEXIBILITY

- Some proteins (for instance antibodies) have domains that can adopt different relative orientations with respect to each other.



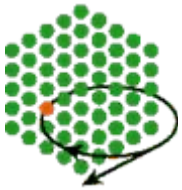
MR will be difficult if the search model does not have the same relative orientation of these domains

Success of MR



- The success of the MR method for solving protein structures from experimental diffraction data depends on the availability of the search model and its accuracy and completeness.

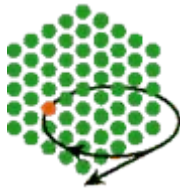
The MR programs and pipelines



- AMORE (Navaza,1994)
- MOLREP (Vagin & Teplyakov, 1997)
- PHASER (Read., 2001)
    - MRBUMP (Keegan & Winn, 2008)
    - BALBES (Long *et al.*, 2008)

The software pipeline makes several decisions concerning (i) truncation of the model in uncertain parts; (ii) the actual protocol for sequence alignment and homology modelling; and (iii) the choice of the MR software, the consensus approach is to derive a variety of models and try MR for all of them one by one

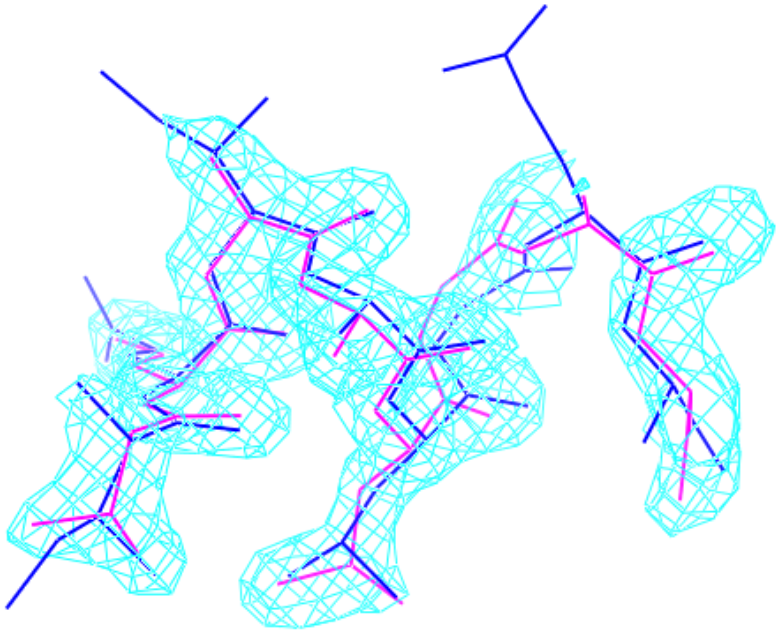
Molecular Replacement



- In principle, accurate phases can be obtained in a matter of hours or even minutes.
- In practice, however, sometimes structure determination by MR is not straightforward.

Why ?

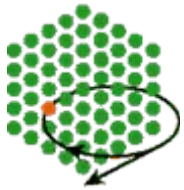
- One major problem is that maps calculated from MR phases inevitably retain some memory of the search model (model bias)



σ_A -weighted

• The more similar the phasing model to the unknown is, the less biased the calculated map will be.

Reduction of model bias

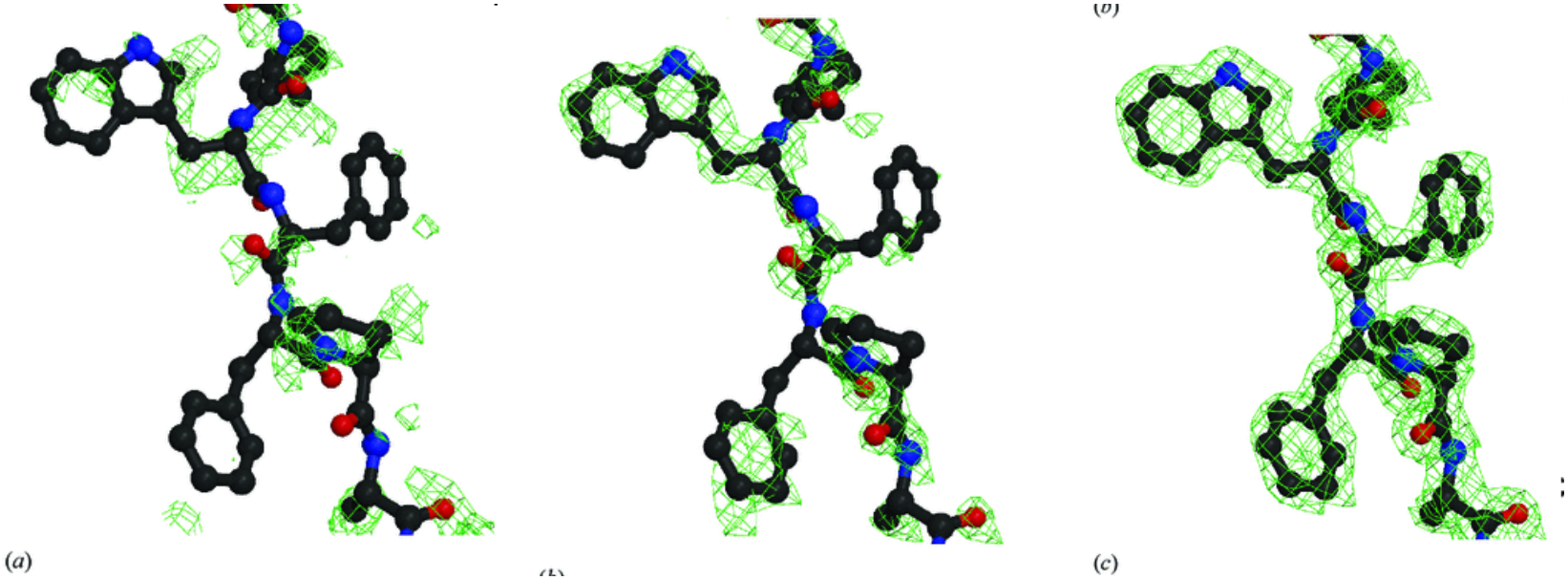


- Repeated cycling of real-space and reciprocal-space refinement
- Density modification and NCS-averaging
- Free atom modelling, refinement and model building as implemented in ARP/wARP (Perrakis et al., 1999)
- SHAKE and wARP (Reddy et al., 2003)
- An ML-based reciprocal-space density-modification method (Prime & Switch) (Terwilliger et al., 2004)

- Omission of parts of the model (Bhat et al, 1988)
- Simulated-annealing OMIT maps (Bruenger 1998)
- Iterative-build OMIT maps (Terwilliger et al., 2008)

Free atom refinement and autobuilding

Final model for *B. subtilis* dUTPase with the electron-density maps



(*a*) After MR and rigid-body refinement,

Resolution 1.7 Å

(b) After DM and NCS averaging

After and automated model building in *ARP/wARP*

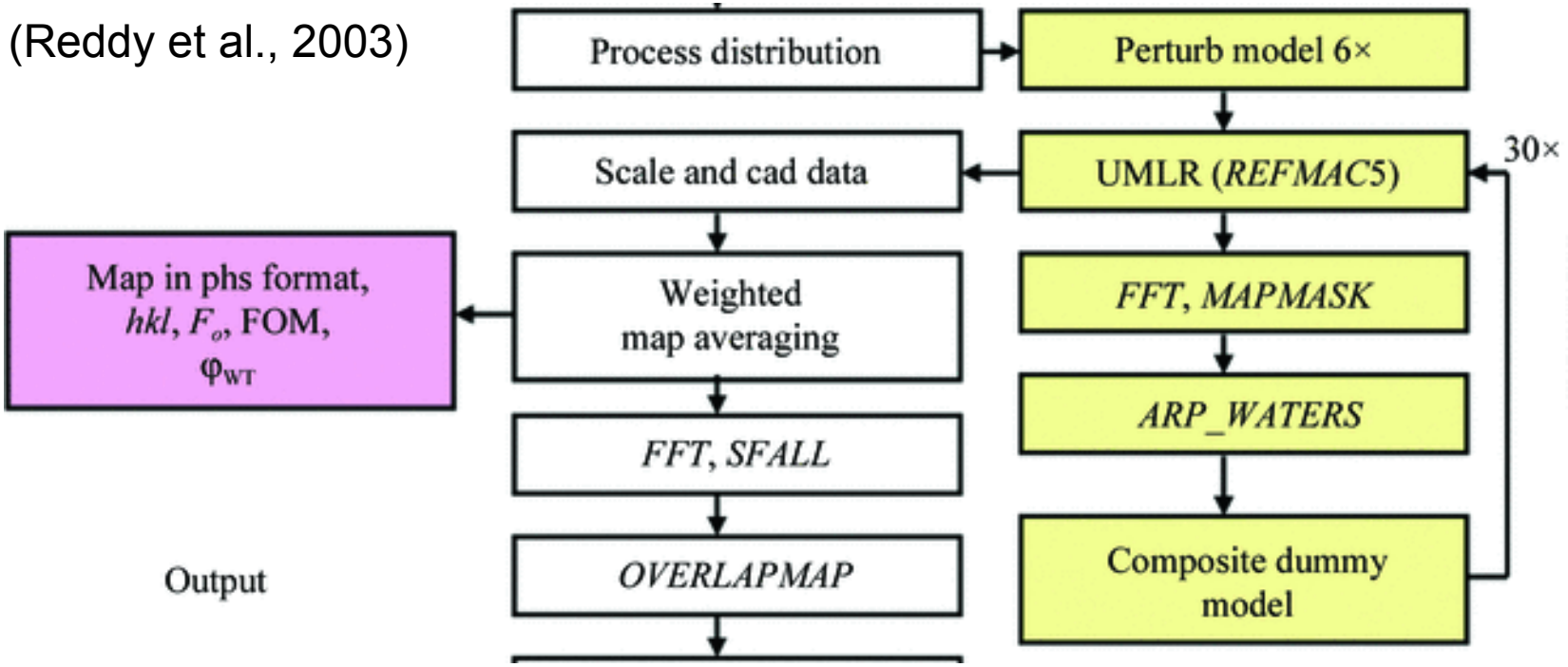
Perrakis et al., 1999

SHAKE & WARP

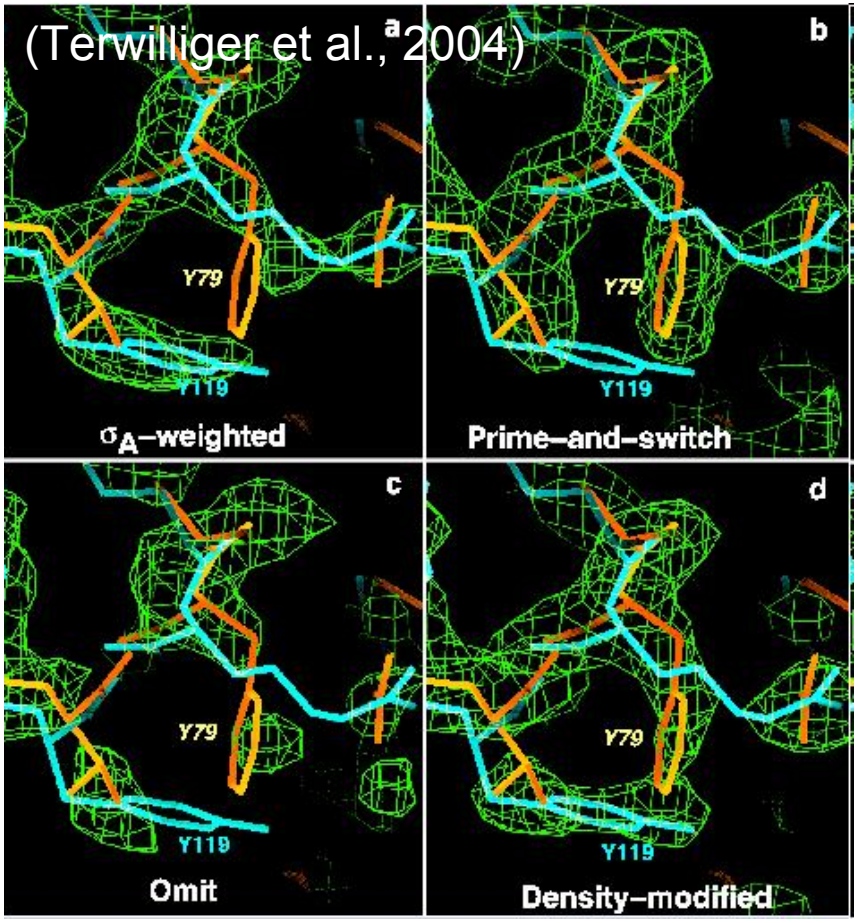
Bias removal server: http://tuna.tamu.edu/

This map improvement server returns a bias minimized, 6-fold averaged map generated from a model and diffraction data (with optional preceding Molecular Replacement). It does not build or repair the model

(Reddy et al., 2003)



Prime and Switch



Initial phase is calculated from MR solution model.

An initial electron-density map is calculated from these phases and measured amplitudes using A-weighting (Read, 1986).

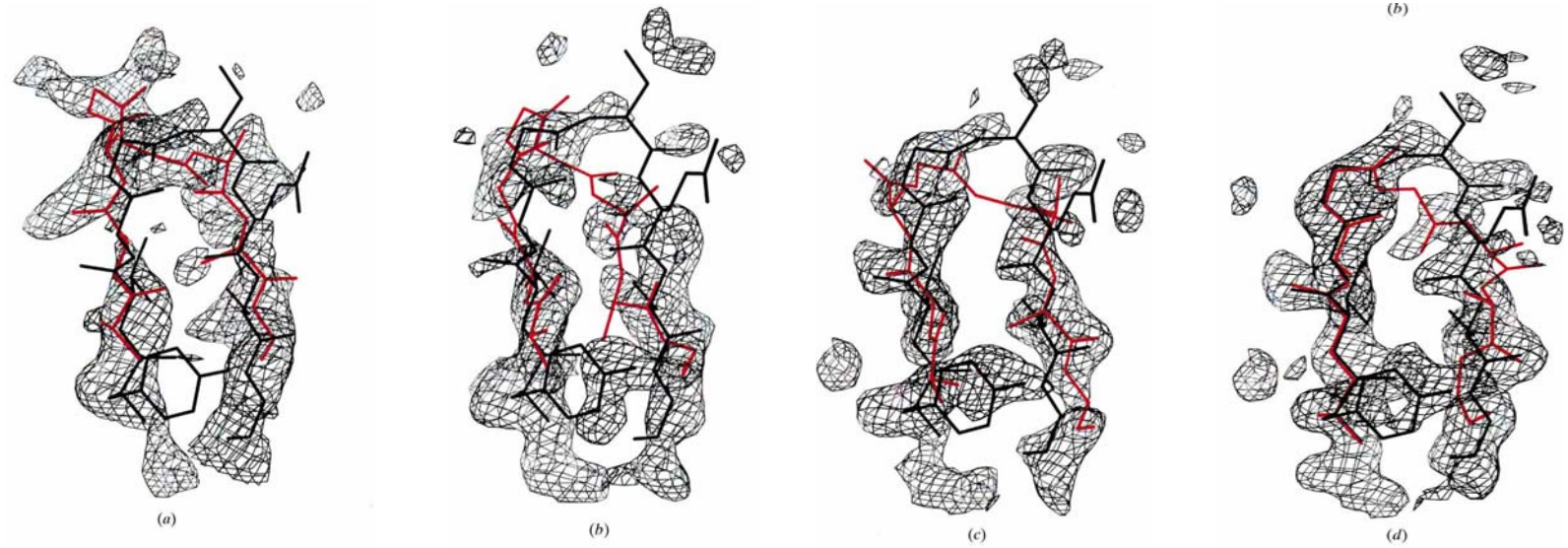
Regions of the map containing solvent, macromolecule or other features are identified

Plausible values of electron density in each such region are obtained from histograms of their values in this map.

The probability that the electron-density map corresponds to a macromolecule is estimated along with first and second derivatives of the probability with respect to each crystallographic phase.

Finally, crystallographic phases are iteratively adjusted so as to maximize the overall probability that the electron-density map corresponds to a macromolecule

Simulated annealing



Initial electron-density map prior to refinement

after refinement with the LSQ target

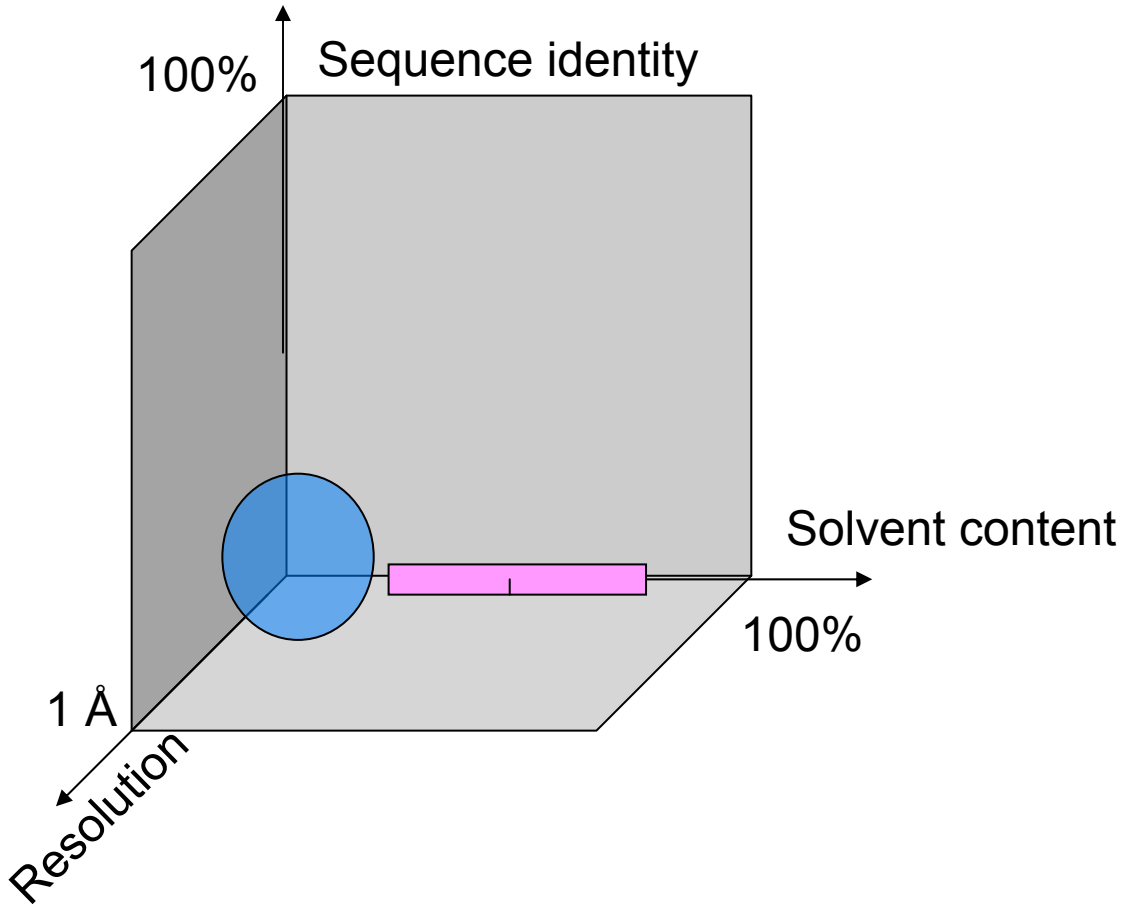
after refinement with the MLF target

after refinement with the MLHL target

Adams et al., 1999

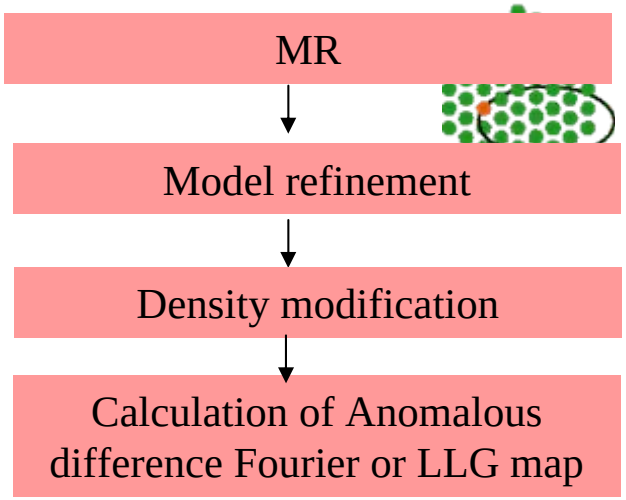
Quality of the model

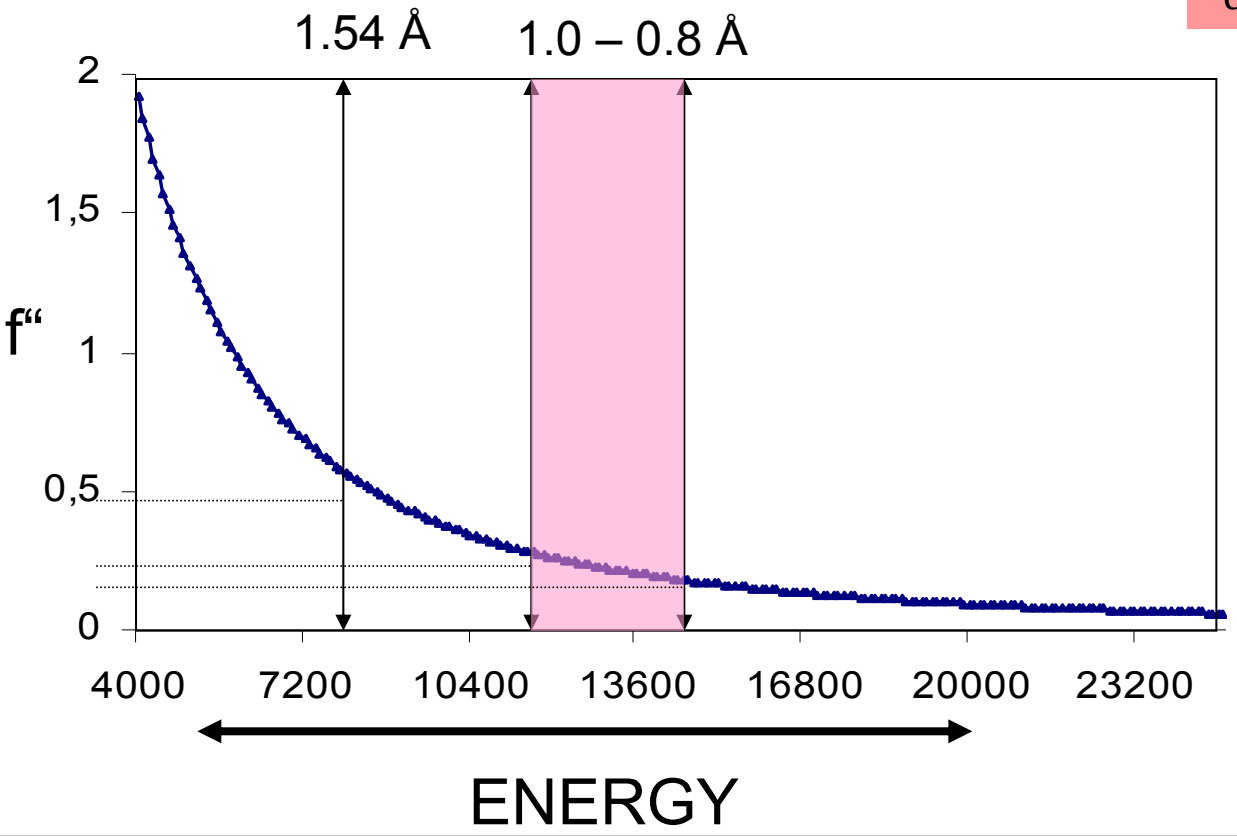
Data quality
Model quality



Validation of MR

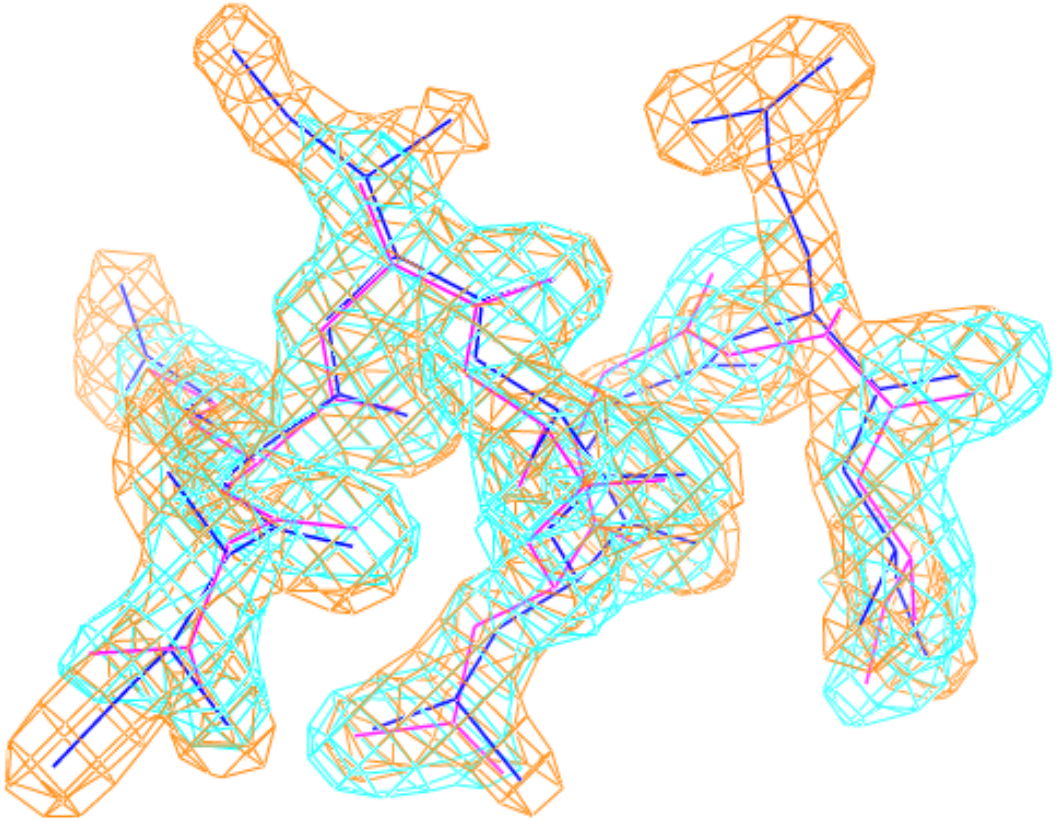
- Collection of Anomalous data





1pjk

Use of anomalous peaks in phase improvement



Calculation of Anomalous difference Fourier or LLG map

↓

Refinement of anomalous peaks (above 5σ) at max resolution

↓

Dual fragment phasing using the partially refined/built model and anomalous peaks

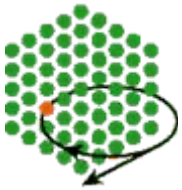
↓

Phase combination of resultant phases, density modification and/or NCS-averaging

↓

Model building

MRSAD

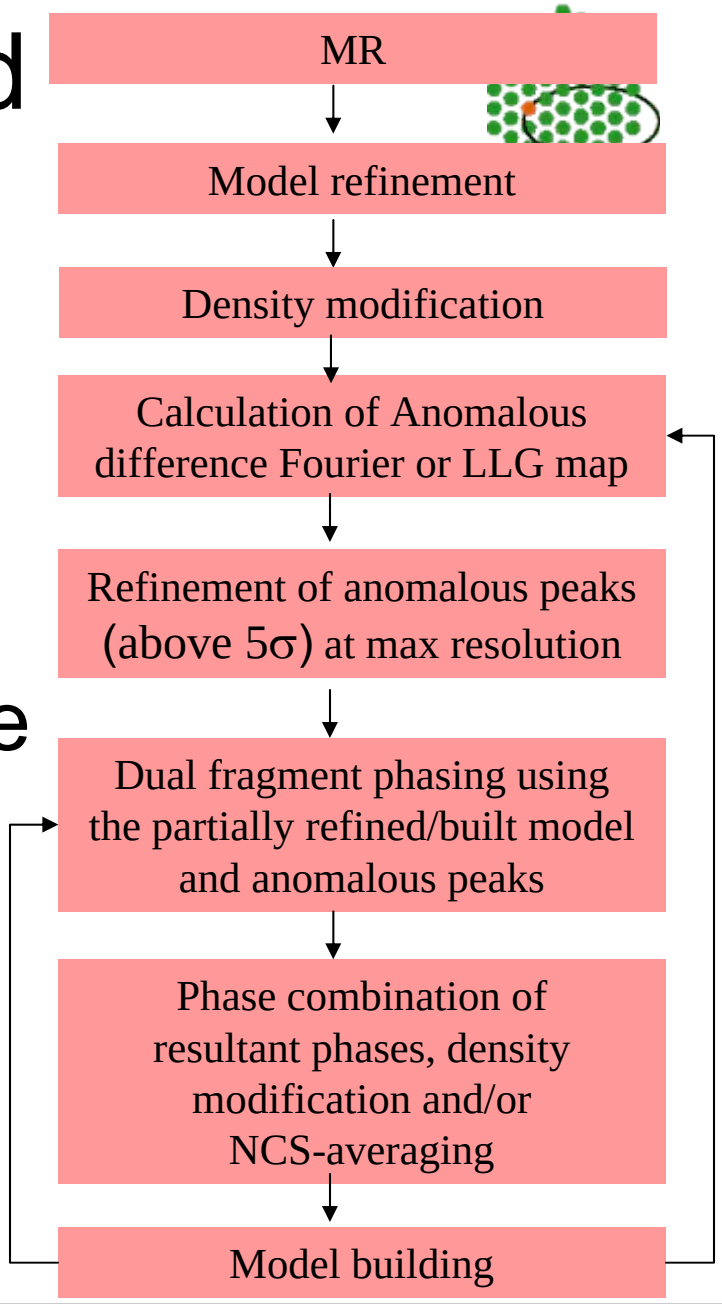


- An attractive feature of MRSAD strategy is that the anomalous scatterer substructure can be quickly determined from the MR solution.

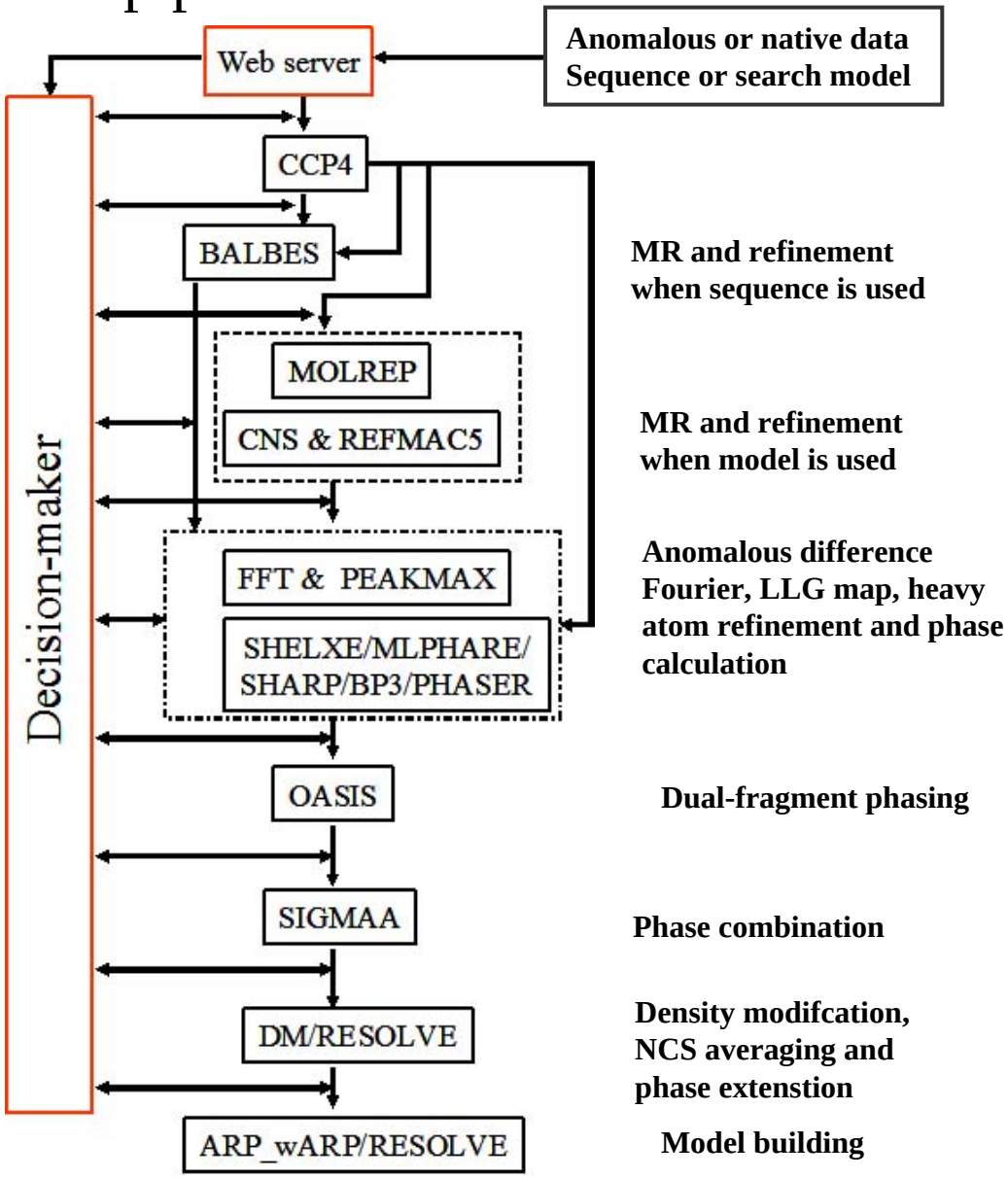
- SAD phases are virtually independent of the MR phases, MRSAD provides an experimental method to overcome the model bias inherent to MR

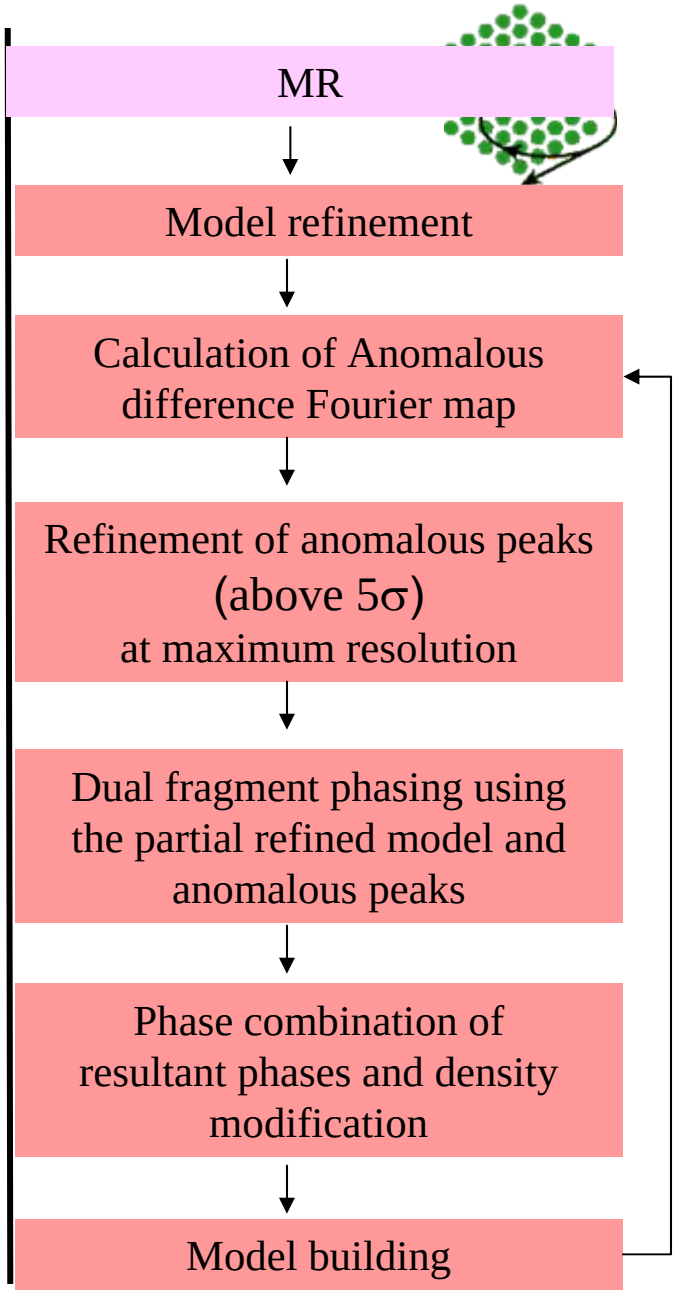
Combination of MR and SAD phasing

- Validation of MR
- Model bias reduction
- Phase improvement of model or experimental phase
- Iterative model completion



Software pipeline for MRSAD



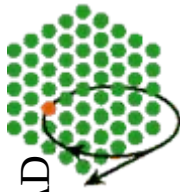


Long wavelength (2.0 Å) dataset

Longwavelength dataset with PDB code
Total residue per molecule = 110 - 755
Number of mol. in asu = 1 - 4
Total number of Se in asu = 2 - 32
Resolution =3.00 - 1.80 Å
Space group

Data quality

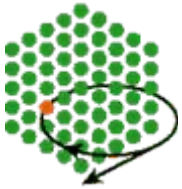
R_{anom}=0.013 - 0.036
R_{merge}=0.031 - 0.141
Multiplicity= 3 - 38
R_{anom}/R_{pim}= 0.83 - 2.00



											SAD	MRSAD
2G4L	hydroxynitrile lyase	257	1	18	1.84	$C222_1$	0.022	0.067	12	0.83	-----	
2G4J	Glucose isomerase	386	1	11	1.85	I222	0.036	0.141	13	0.86	-----	1.80
2G4O	LeuB	355	4	32	2.00	$P2_12_12_1$	0.019	0.06	12	0.90	-----	
2G4R	MogA	156	3	03	1.92	$P2_1$	0.019	0.07	06	0.91	-----	
2G4W	Ribonuclease A (C2)	128	2	18	1.84	C2	0.022	0.063	06	0.97	-----	1.00
2G4I	Concanavalin A	237	1	07	2.40	I222	0.037	0.104	11	1.05	-----	1.60
2G4K	hARH3	256	1	19	1.82	$C222_1$	0.019	0.069	12	1.06	-----	2.20
2G51	Trypsin (p1)	240	1	10	1.84	P1	0.02	0.06	03	1.09	-----	
2G4X	Ribonuclease A ($P3_221$)	128	1	18	1.95	$P3_221$	0.019	0.053	18	1.10	-----	1.40
2G4N	α-lactalbumin	122	6	54	2.30	$P2_12_12$	0.03	0.071	12	1.12	-----	
2G4Y	Thaumatin	207	1	16	1.98	$P4_12_12$	0.013	0.064	25	1.12	0.682	
2G4T	PPE (Na)	240	1	13	1.84	$P2_12_12_1$	0.015	0.043	12	1.15	0.549	
2G4V	Proteinase K	279	1	13	2.14	$P4_32_12$	0.018	0.045	26	1.18	0.559	
2G4S	NBR1PB1	85	1	05	2.15	$P6_322$	0.013	0.053	34	1.18	0.558	
2G52	Trypsin ($P2_1$)	240	1	12	1.84	$P2_1$	0.033	0.076	06	1.22	-----	1.80
2G4U	PPE (Ca)	240	1	13	1.84	$P2_12_12_1$	0.033	0.123	13	1.38	0.667	
2G55	Trypsin (P3121)	240	1	10	1.82	$P3_121$	0.019	0.086	10	1.46	0.708	
2G4Z	Thermolysin	316	1	06	1.98	$P6_122$	0.018	0.067	38	1.64	0.655	
2G4Q	Lysozyme at pH 8.0	129	1	15	1.84	$P4_32_12$	0.024	0.031	23	1.73	0.675	
2G4P	Lysozyme at pH 4.5	129	1	17	1.84	$P4_32_12$	0.022	0.04	24	2.00	0.645	
2G4M	Insulin	51	1	06	1.80	$I2_13$	0.019	0.057	35	2.00	0.788	

Mueller-Dieckmann et al., 2007

MRSAD examples on JCSG SeMet data



Dataset

JCSG dataset with PDB code
Total residue per molecule = 110 - 755
Number of mol. in asu = 1 - 4
Total number of Se in asu = 2 - 32
Resolution =3.00 - 1.80 Å
Space group

Search model quality

Search model with PDB code
sequence identity= 35 - 51%
rmsd after optimal overlap = 1.2 - 2.4 Å

Data quality

R_{anom}=0.032 - 0.107
R_{merge}=0.064 - 0.134
Multiplicity=3.4 - 7.5
R_{anom}/R_{pim}=0.67 - 2.40

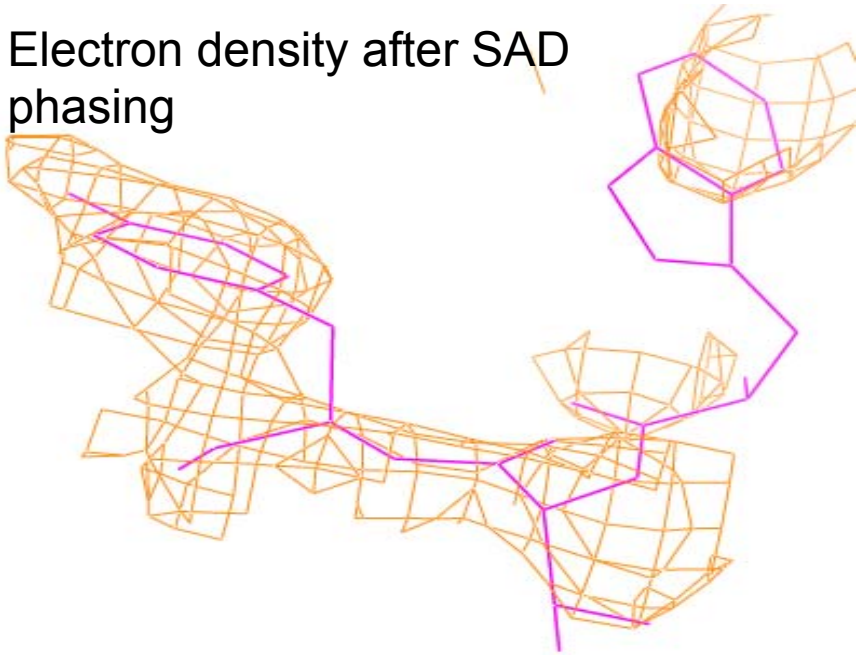
													MRSAD	SAD
2hh6	116	1	06	2.04	$P6_522$	2o4t	0.44	1.04/072	0.056	0.10	7.5	1.43	N	Y
1vky	563	1	06	3.00	I222	1yy3	0.49	1.84/247	0.086	0.126	3.5	1.10	Y	N
2gi3	452	1	13	1.80	$P3_221$	2g5i	0.50	1.34/403	0.058	0.112	6.3	1.19	Y	N
2hxv	354	1	05	2.80	I222	2d5n	0.41	2.32/316	0.057	0.078	4.0	1.27	Y	N
2o08	193	2	08	2.90	$C222_1$	2ogi	0.42	1.30/180	0.058	0.07	3.7	1.36	Y	Y
1vmf	407	1	05	1.90	$P2_12_12_1$	2cu5	0.42	1.09/124	0.107	0.134	4.1	1.41	Y	N
1zbt	320	1	08	2.40	P43212	2b3t	0.49	2.43/224	0.043	0.071	6.7	1.45	Y	Y
1vmi	329	1	08	2.32	$P6_322$	1xco	0.42	1.76/311	0.046	0.062	5.0	1.48	Y	Y
2f4l	276	4	28	2.50	$P2_12_12_1$	2ii1	0.36	1.03/274	0.075	0.084	3.8	1.49	Y	N
2fvg	311	1	06	2.50	$P2_12_12_1$	1vhe	0.37	1.70/227	0.073	0.081	3.8	1.51	Y	Y
1vjo	755	1	08	2.00	$P2_12_12_1$	2ch2	0.43	1.15/371	0.061	0.064	4.1	1.68	Y	Y
1vjr	269	1	06	2.40	$P4_12_12$	1zjj	0.40	1.55/240	0.076	0.072	4.6	2.00	Y	Y
1vkn	338	4	32	2.45	$P2_1$	1xyg	0.46	1.18/316	0.069	0.065	6.1	2.40	Y	Y

Initial partial model built
from SAD phasing protocol

20 – 40%

Residue per monomer: 261
Molecules in ASU: 4
Max Resolution: 2.3 Å
Space group: P21

Electron density after SAD phasing



Part of the complete structure, shown in experimental map

MR

Model refinement

Density modification

Calculation of Anomalous difference Fourier or LLG map

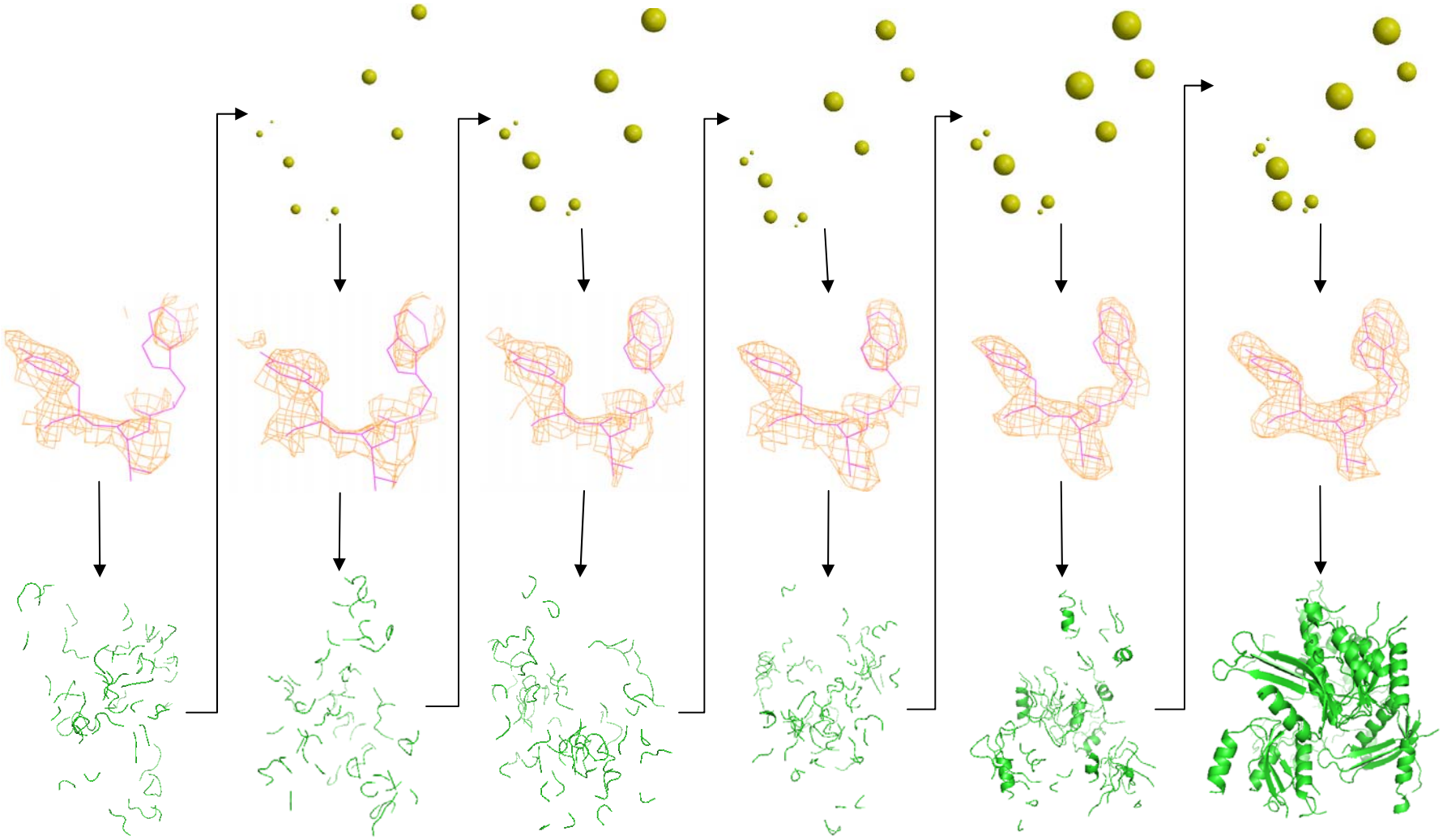
Refinement of anomalous peaks (above 5σ) at max resolution

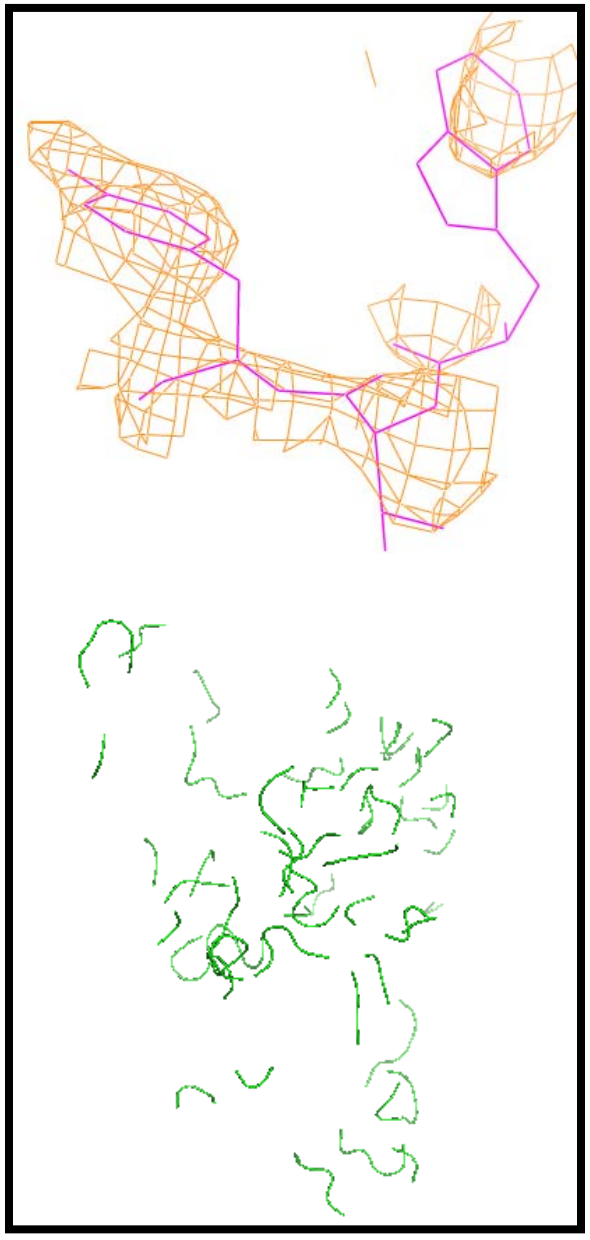
Dual fragment phasing using the partially refined/built model and anomalous peaks

Phase combination of resultant phases, density modification and/or NCS-averaging

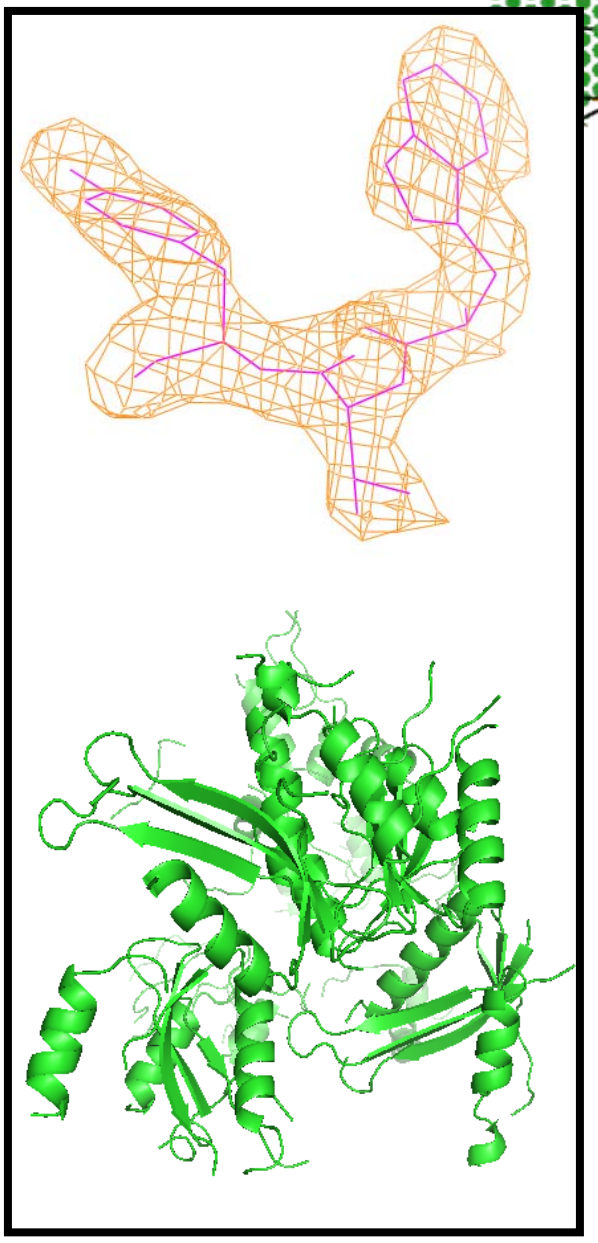
Model building

Evolution of substructure, electron density and model



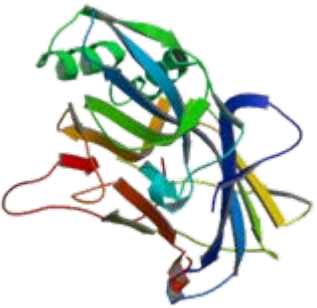


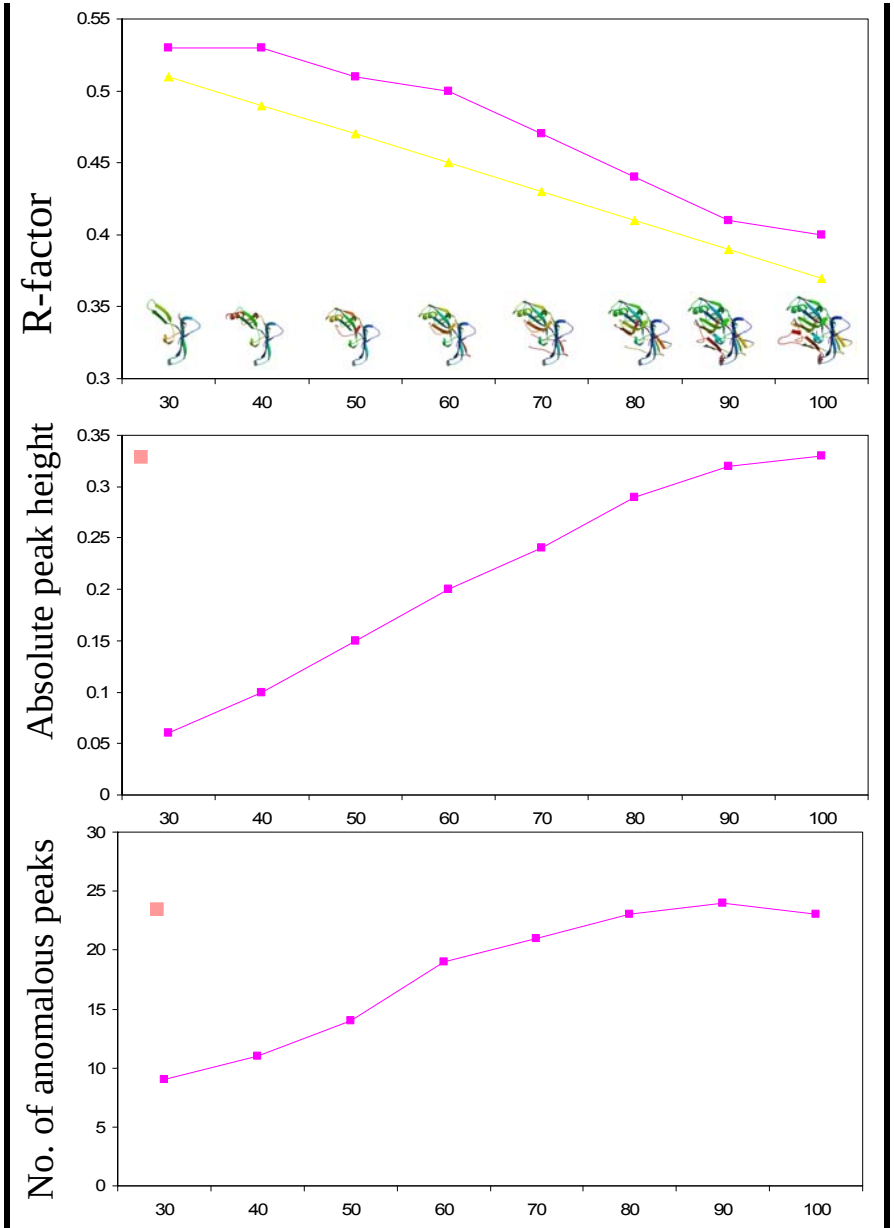
SAD to MRSAD



A means of completing your model

Br atoms 24
Resolution 1.8 Å
R_{merge} 10.0 (43.6)
Space group C2
Redundancy: 7
Residues : 263





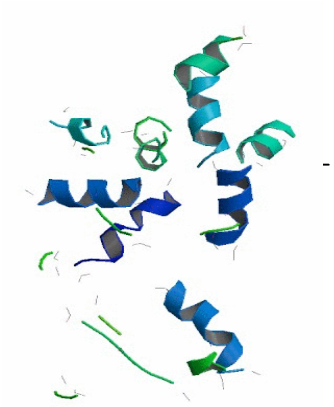
Cycle 1	34	263	257	262	262	260	262	262
Cycle 2	262							

Final model built →

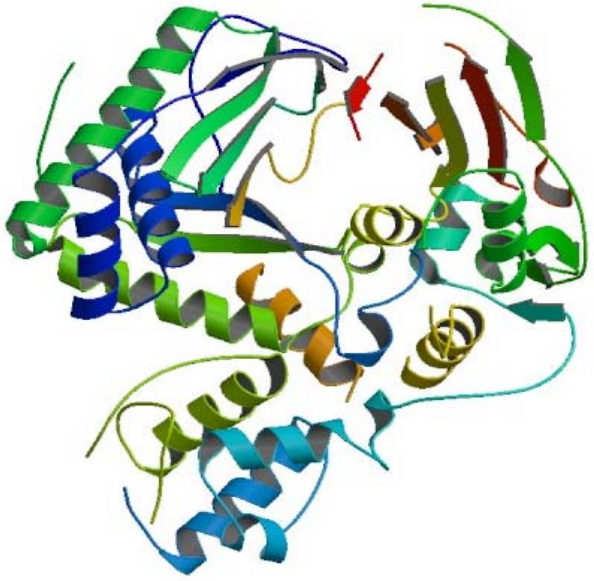
MR
↓
Model refinement
↓
Calculation of Anomalous difference Fourier map
↓
Refinement of anomalous peaks (above 5σ) at maximum resolution
↓
Dual fragment phasing using the partial refined model and anomalous peaks
↓
Phase combination of resultant phases and density modification
↓
Model building

Buliding structure from partial model

Monomer in ASU
547 residues/monomer
13 selenium/monomer
Space group $P4_32_12$
Maximum resolution 2.5 Å



After MRSAD



MR

Model refinement

Density modification

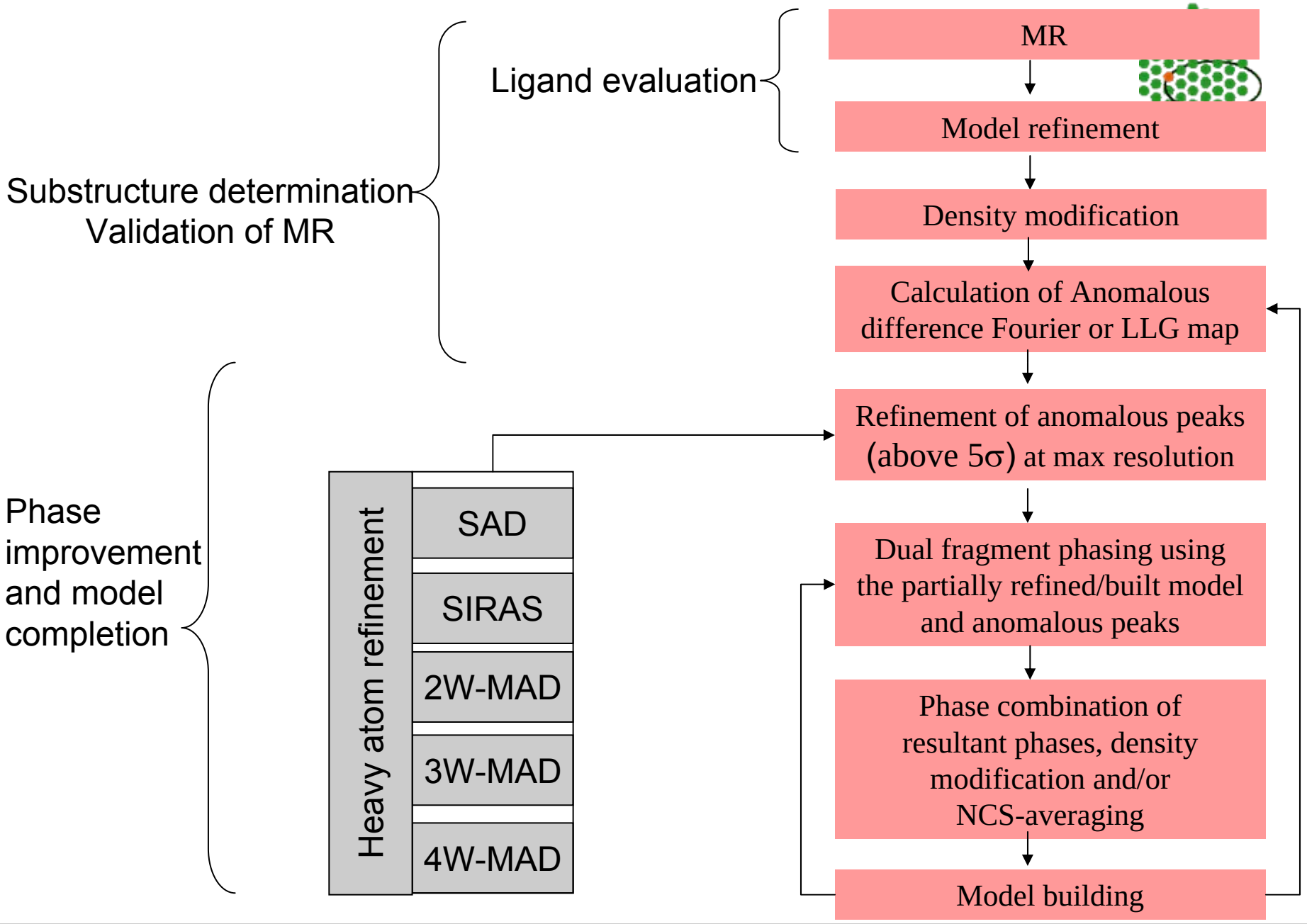
Calculation of Anomalous difference Fourier or LLG map

Refinement of anomalous peaks (above 5σ) at max resolution

Dual fragment phasing using the partially refined/built model and anomalous peaks

Phase combination of resultant phases, density modification and/or NCS-averaging

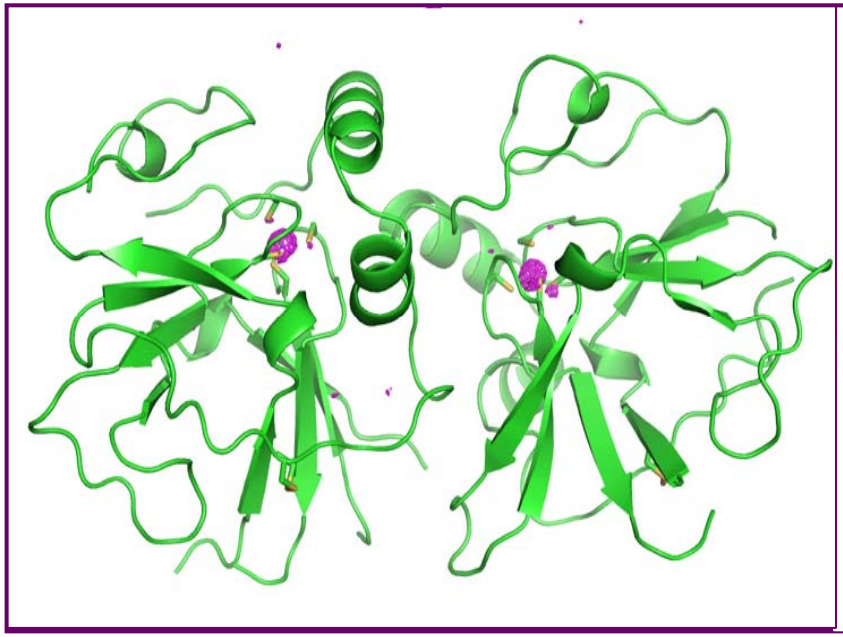
Model building

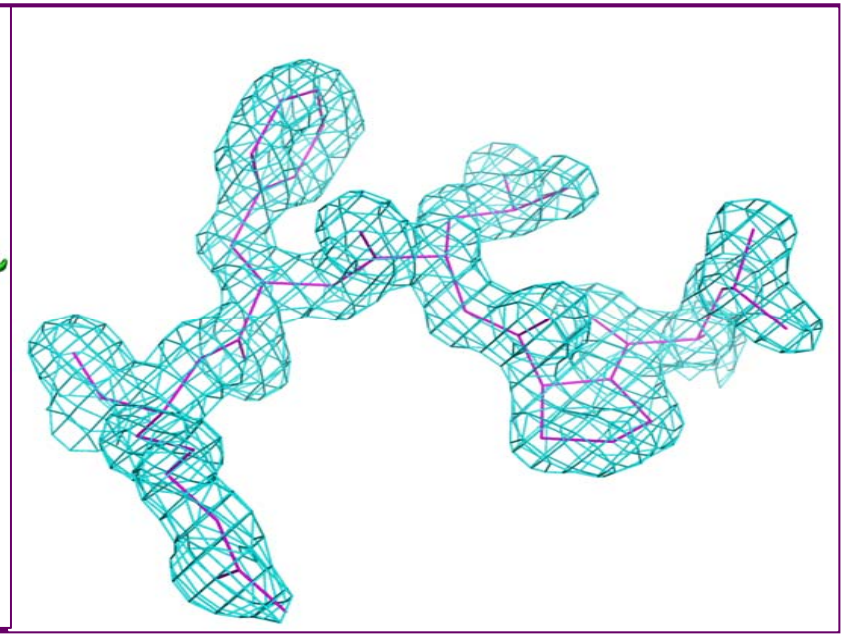

Ligand evaluation
Substructure determination
Validation of MR
Phase improvement and model completion
Heavy atom refinement
SAD
SIRAS
2W-MAD
3W-MAD
4W-MAD
MR
Model refinement
Density modification
Calculation of Anomalous difference Fourier or LLG map
Refinement of anomalous peaks (above 5σ) at max resolution
Dual fragment phasing using the partially refined/built model and anomalous peaks
Phase combination of resultant phases, density modification and/or NCS-averaging
Model building

Phasing at short wavelength

Heavy atoms/ASU	2 Zn + 12S
Wavelength (Å)	0.8
f”	1.76 (Zn) 0.15(S)
Bijvoet ratio	1.14

Space group	$P2_1$
Redundancy	3.7
Resolution (Å)	20-1.66
Completeness (%)	97.4 (94.3)
I/σ(I)	14.2(2.8)
R_{merge} (%)[b]	3.7 (37.0)

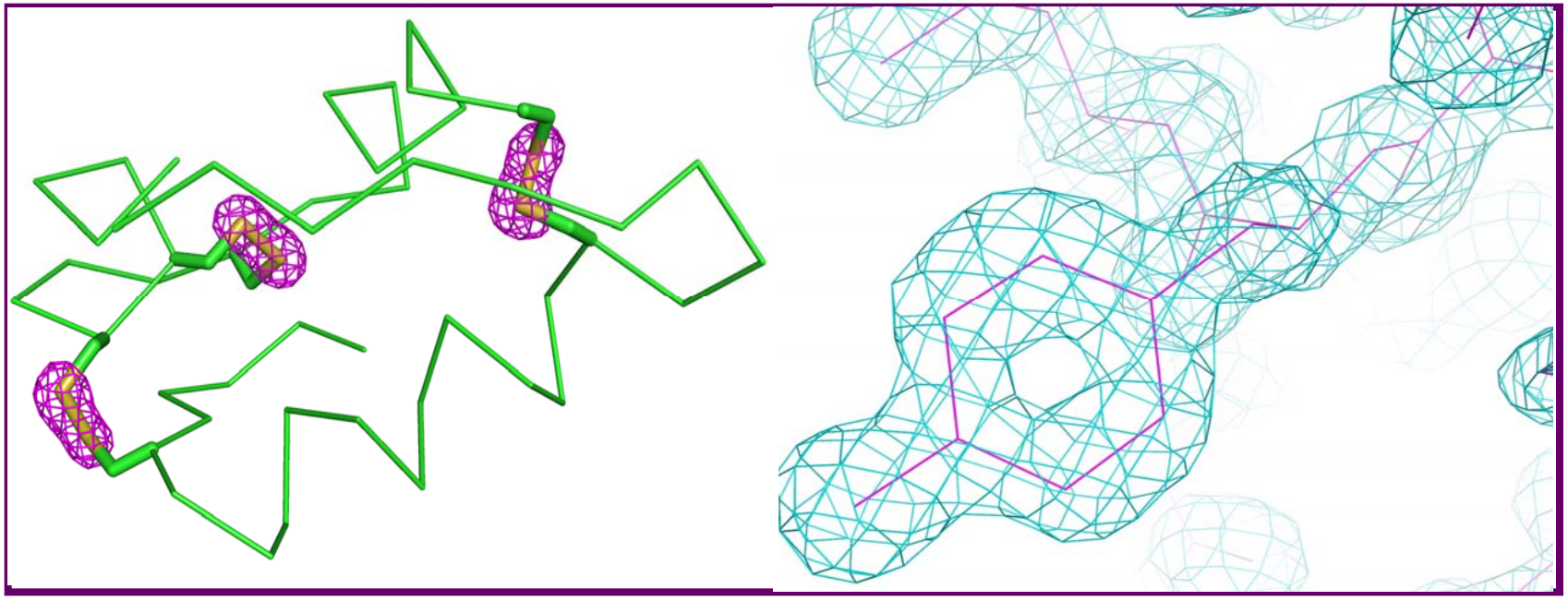




Phasing at short wavelength

Heavy atoms/ASU	6S
Wavelength (Å)	1.00
f”	0.24(S)
Bijvoet ratio	0.66

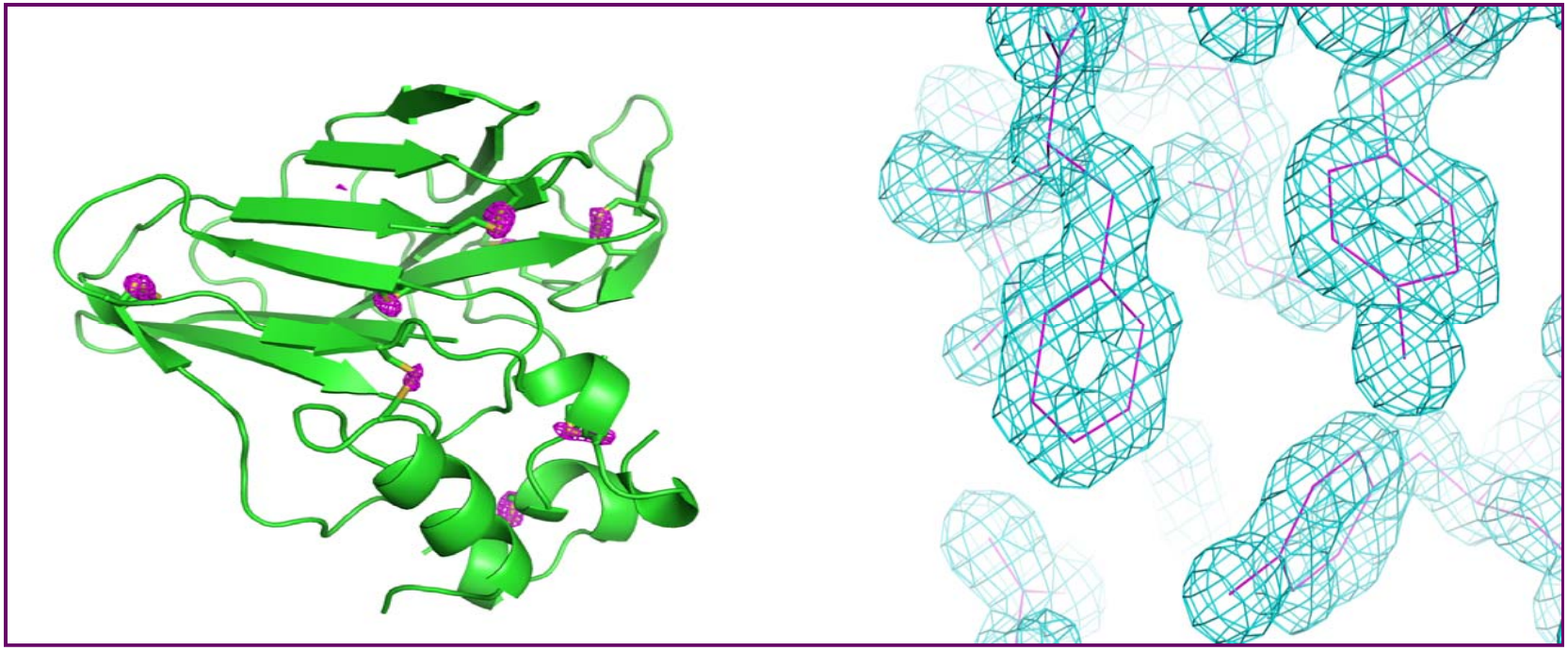
Space group	$I2_13$
Redundancy	22.6
Resolution (Å)	20-1.6
Completeness (%)	98.2(100.0)
I/σ(I)	42.0 (13.0)
R_{merge} (%)[b]	3.2 (12.2)



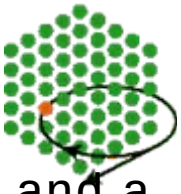
Phasing at short wavelength

Heavy atoms/ASU	17S
Wavelength (Å)	0.9
f”	0.20(S)
Bijvoet ratio (%)	0.45

Space group	$P4_12_12$
Redundancy	7.6
Resolution (Å)	20-1.4
Completeness (%)	94 (68)
I/σ(I)	33.3(9.8)
R_{merge} (%)[b]	3.9(13.7)



Exploiting Phase Information



So far, we have assumed that we have only native diffraction data and a molecular replacement model.

Of course, we can well find ourselves in a situation where we have experimental phases (from a heavy atom derivative) as well as a model.

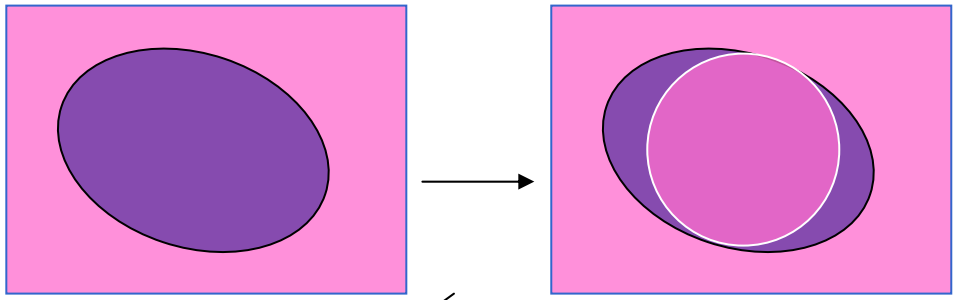
Phase information can, in fact, be exploited to help to solve the MR problem.

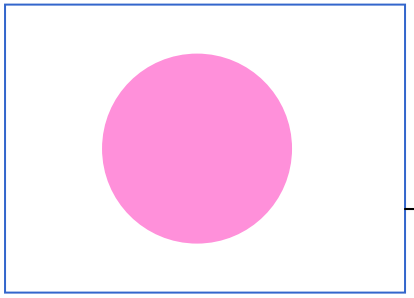
This is particularly useful when we have a molecular replacement model that is too poor to find a solution and phases that are too weak to provide an interpretable map.

By combining molecular replacement with experimental phases, the two together may be enough to succeed.

This technique of combining experimental phases with molecular replacement methods is called **phased molecular replacement**.

Phased Molecular Replacement

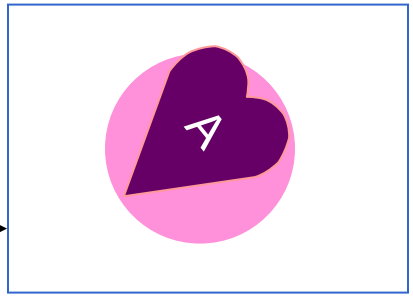




FT

F_A

Domain rotation function



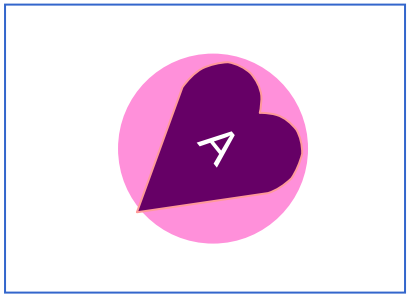
Orientation (α_A, β_A, γ_A)

FT

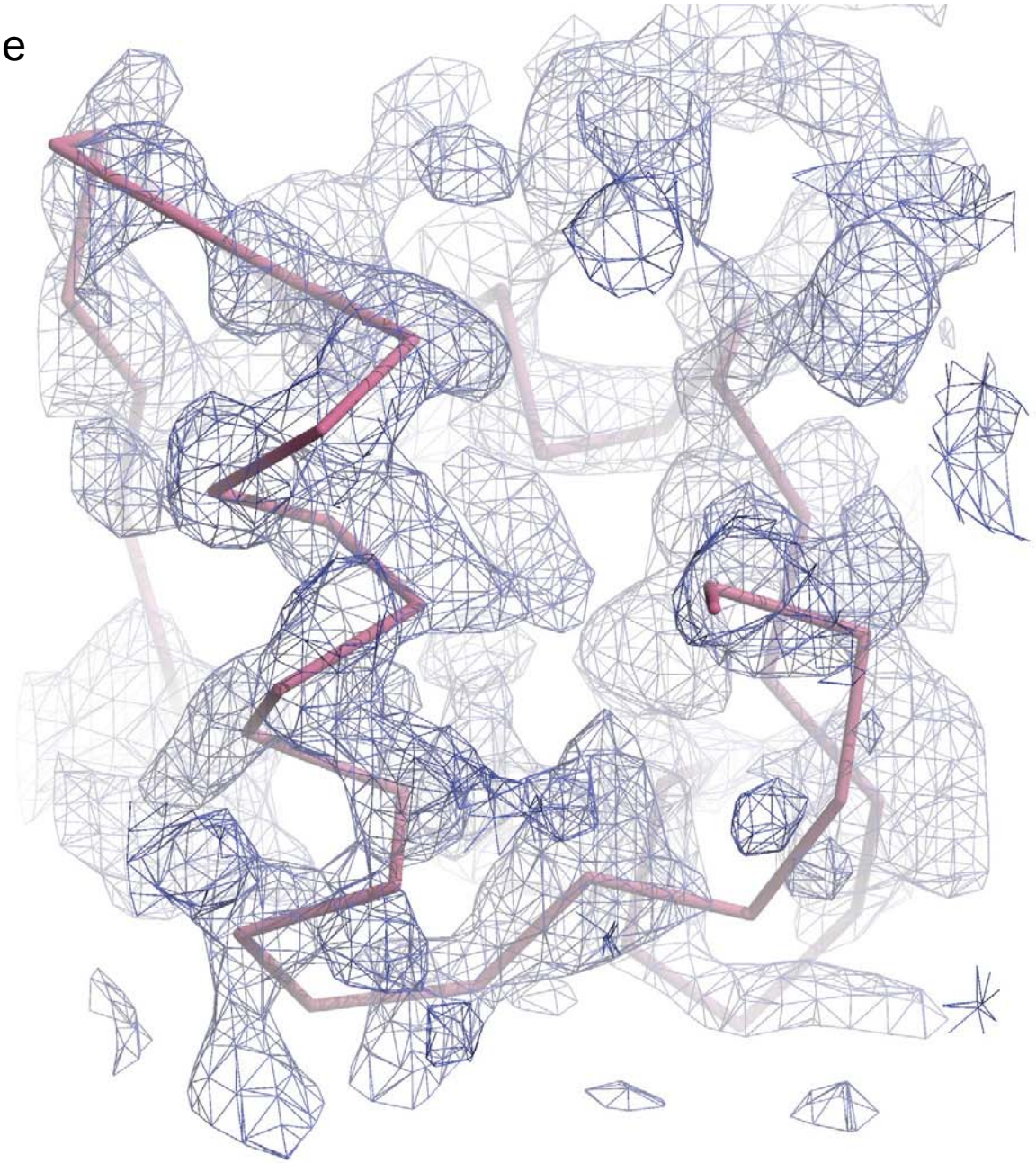
$F_{(\alpha_A, \beta_A, \gamma_A)}$

Phased translation function

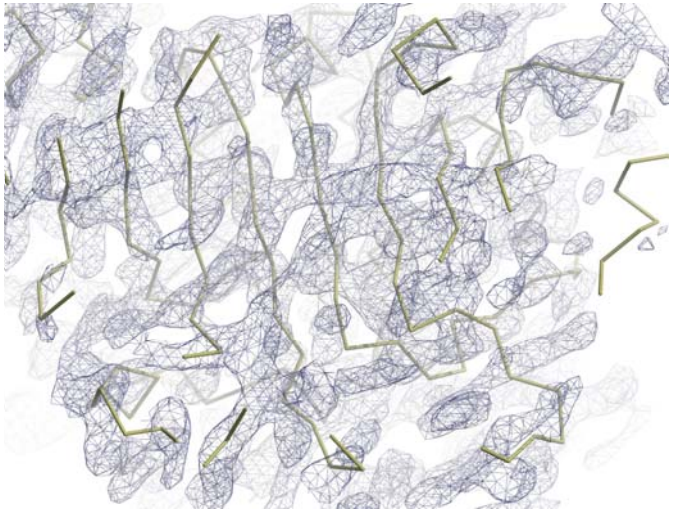
(T_x^A, T_y^A, T_z^A)

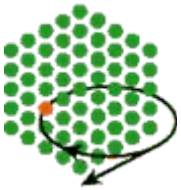


An Example

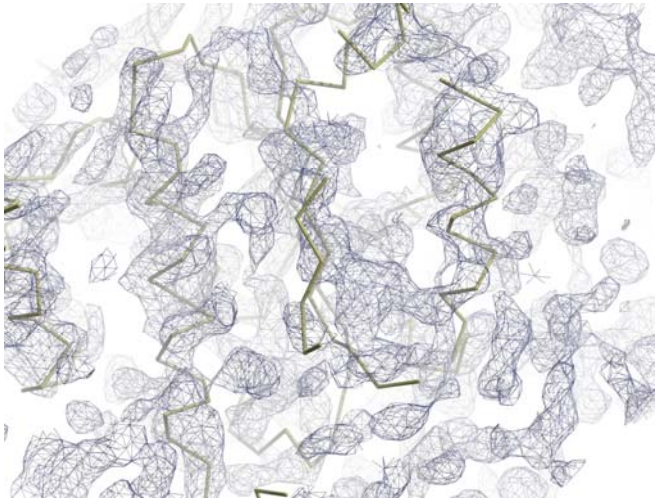


Another example





Phased MR model is useful for interpretability of the poor experimental map and this can be also useful in phase enhancement



90°

User interface and server

Auto-Rickshaw server

http://www.embl-hamburg.de/Auto-Rickshaw

Since 14th April 2008 the AR server is available to world wide crystallographic community

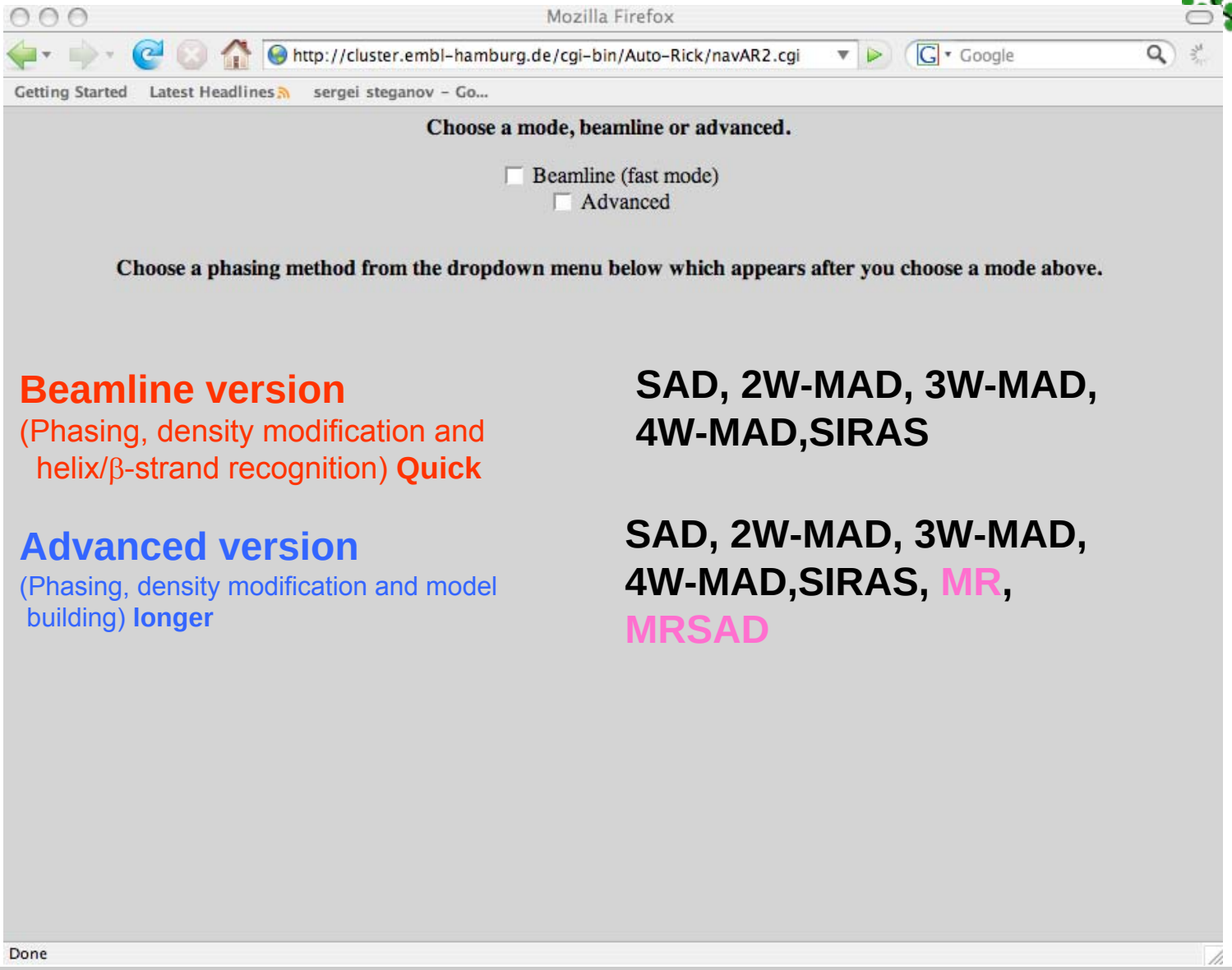
Web-browser

DPS2AR

GUI

Command line

Remote web service: AutoRickshaw



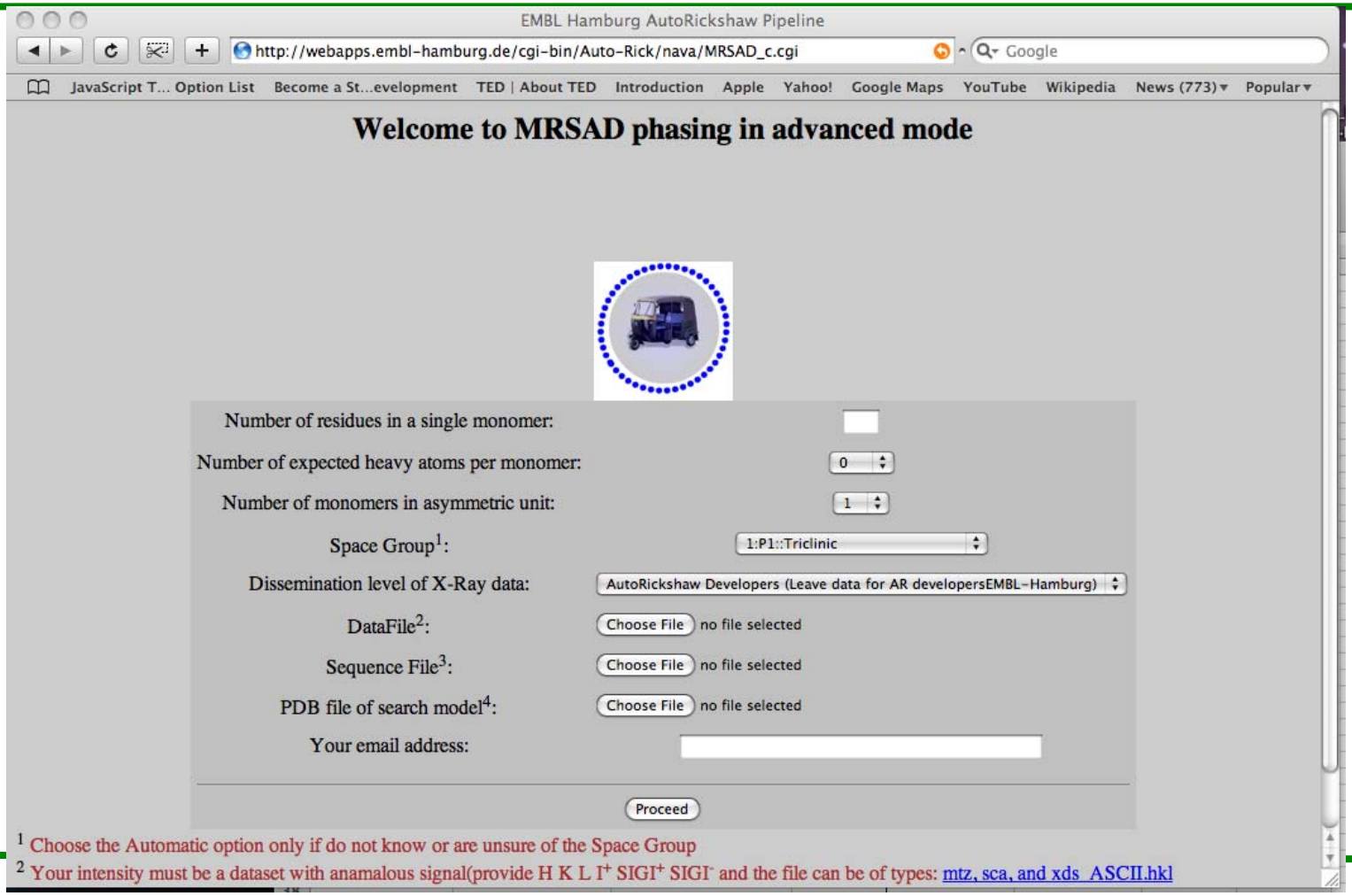
Beamline version
(Phasing, density modification and helix/β-strand recognition) **Quick**

SAD, 2W-MAD, 3W-MAD, 4W-MAD,SIRAS

Advanced version
(Phasing, density modification and model building) **longer**

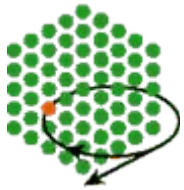
SAD, 2W-MAD, 3W-MAD, 4W-MAD,SIRAS, MR, MRSAD

Remote web service: AutoRickshaw



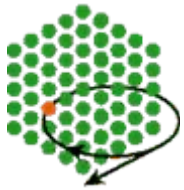
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Auto-Rickshaw* - An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Cryst.* D61, 449-457.

Necessary input



- No. of residues per monomer
- No. of expected heavy atom sites per monomer
- No. of monomers per a.u.
- Space group
- X-ray data
- Sequence file (optional)
- Model file PDB-format (optional)
- Email address

X-ray data input



- Integrated and scaled X-ray data files
  - formats — Extension
    - **SCALEPACK :** H K L I^+ SIGI$^+$ I^- SIGI$^-$ (.sca)
    - **MTZ :** H K L I^+ SIGI$^+$ I^- SIGI$^-$ (.mtz)
    - **XDS :** XDS-ASCII.HKL (.HKL) or (.hkl)
    - **XSCALE** : (.HKL) or (.hkl)
    - **D*TREK** (.REF) or (.ref)

Acknowledgements

Santosh Panjikar
Venkataraman Parthasarathy
Manfred S. Weiss
Victor S. Lamzin
Paul A. Tucker

All the developers of the various computer programs for their kind permission to use their software in this pipeline (Martyn Winn, Kevin Cowtan, Qang Hao, George Sheldrick, Gerard Bricogne, Clemens Vonrhein, Navraj Pannu, Tom Terwilliger, Gerard Kleywegt, Alwyn Jones, G. D. Smith, Victor Lamzin, Anastassis Perrakis, Rendy Read, Garib Murshudov, Hi-Fu Fan).

Annette Faust and Brice Kaufmann for short wavelength data and Christoph Mueller-Dieckmann, Manfred Weiss and Paul Tucker for Long wavelength data.

Hamburg beamline/Auto-Rickshaw users and JCSG for the provision of data

Thank you for your attention