



The European Synchrotron

Page 1 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

ANOMALOUS SCATTERING AND THE PHASE PROBLEM

GORDON LEONARD
ESRF STRUCTURAL BIOLOGY GROUP

Page 2 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron |

CONTENTS

1. The Phase Problem
2. Isomorphous Replacement
 1. SIR
 2. MIR
3. Anomalous Scattering
 1. What is it?
 2. How can we exploit it?
 1. SIRAS
 2. MAD
 3. SAD
4. Successful Phasing
 1. Data collection
 2. Data analysis

Page 3 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron |

STRUCTURE DETERMINATION

Integrated Intensities

Electron density
(= protein structure)!!

Page 4 | CCP4-DLS Workshop December 2016| Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

THE ELECTRON DENSITY EQUATION

$$\rho_{(x,y,z)} = \frac{1}{V_c} \sum_h \sum_k \sum_l F_{hkl} \exp(-2\pi i(hx + ky + lz))$$

$$F_{hkl} = \sum_j f_j \exp(-2\pi i(hx_j + ky_j + lz_j))$$

$$F_{hkl} = \sum_j |F_{hkl}| \exp(-2\pi i(\phi))$$

↑ Amplitude ↑ Phase

From an x-ray diffraction experiment we can 'measure' the amplitude ($|F| \propto \sqrt{I}$) but get **no information about the phase**. Deriving the phase is known as **'the phase problem'**

Page 5 | CCP4-DLS Workshop December 2016| Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

'GOOD' PHASES ARE MORE IMPORTANT THAN 'GOOD' AMPLITUDES

Amplitudes (colours) from duck transform

Phases (brightness) from cat transform

<http://www.yesbi.york.ac.uk/~cowtan/fourier/magic.html>

Page 6 | CCP4-DLS Workshop December 2016| Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

METHODS FOR RESOLVING THE PHASE PROBLEM IN MX

1. Direct Methods

- requires high (near atomic; $d_{\min} \sim 1.2 \text{ \AA}$) resolution data, still quite rare

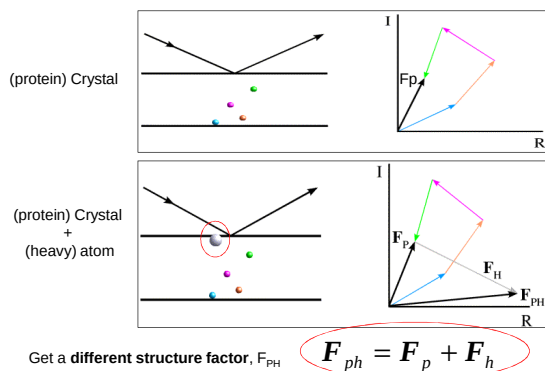
2. Molecular Replacement

- requires some prior knowledge of the crystal structure you want to solve (homologous protein, etc.)

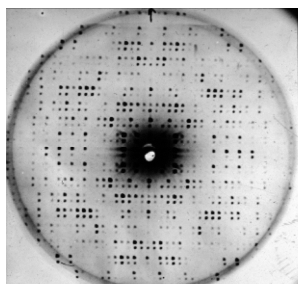
3. Experimental Phasing

- Isomorphous Replacement
- Anomalous Dispersion Methods

INTRODUCING A NEW (HEAVY) ATOM INTO A CRYSTAL



INTRODUCING A NEW (HEAVY) ATOM INTO A CRYSTAL #2



Taylor, G. L. (2010). Introduction to phasing. *Acta Cryst.* D66, 325-338.

Two protein diffraction patterns shifted vertically relative to one another. One is from native bovine β -lactoglobulin, the other from a crystal soaked in a mercury-salt solution. Note: 1) intensity changes for reflections in the latter; 2) identical unit cells suggesting isomorphism.

$$|\Delta F| = |F_{ph} - F_p|$$

Addition of 1 Hg atom to a protein of 1000 atoms will produce an average fractional change of intensity of ~25% so differences should be easy to measure and data doesn't have to be that accurate

Isomorphism + Heavy Atom = Isomorphous Replacement

$$\frac{\Delta I}{I} \approx \sqrt{\frac{N_H}{2N_P}} \frac{f_H}{f_P}$$

HOW IS ISOMORPHOUS REPLACEMENT DONE?

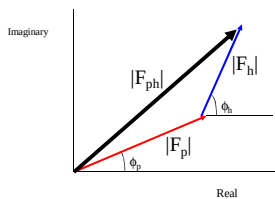
1. Collect data set from **native** crystal (structure factors, F_p)
2. Introduce a small number of heavy atoms (Pt, Hg,.) into another crystal of the same protein
3. Collect data set from **derivative** crystal (structure factors, F_{ph})
4. Use differences (**isomorphous differences**) to find heavy atom positions in unit cells
5. Use these positions to initiate phase calculations

Page 10 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

SINGLE ISOMORPHOUS REPLACEMENT

$$F_{ph} = F_p + F_h$$

Want to calculate ϕ_p :

$$\Phi_p = \Phi_h \pm \cos^{-1}(F_{ph}^2 - F_p^2 - F_h^2 / 2F_p F_h)$$

$$\Phi_p = \Phi_h \pm \Phi$$

Phase Ambiguity

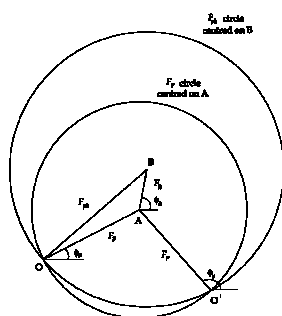
$$|F_h| \approx ||F_{ph}| - |F_p||$$

Measure $|F_p|$, $|F_{ph}|$; Calculate $|F_h|$, ϕ_h from heavy atom positions

Page 11 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

GRAPHICAL ILLUSTRATION OF PHASE AMBIGUITY



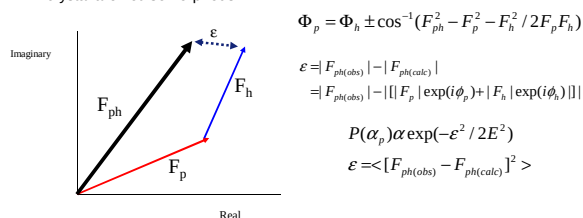
No errors shown here. What we actually get are phase probability distributions

Page 12 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

LACK OF CLOSURE (ϵ) AND PHASE PROBABILITIES

1. Measurement errors in $|F_p|$, $|F_{ph}|$
2. Model of heavy atom substructure is incorrect (minor sites not included)
3. Native and heavy atom derivative crystal are not isomorphous



$$\Phi_p = \Phi_h \pm \cos^{-1}(F_{ph}^2 - F_p^2 - F_h^2 / 2F_p F_h)$$

$$\epsilon = F_{ph(obs)} - |F_{ph(calc)}|$$

$$= |F_{ph(obs)}| - ||F_p| \exp(i\phi_p) + |F_h| \exp(i\phi_h)||$$

$$P(\alpha_p) \propto \exp(-\epsilon^2 / 2E^2)$$

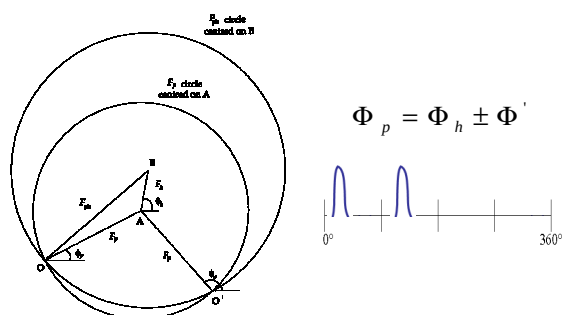
$$\epsilon = \langle [F_{ph(obs)} - F_{ph(calc)}]^2 \rangle$$

Blow, D. M. & Crick, F. H. C. (1959). Acta Cryst. **12**, 794-802.

Page 13 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

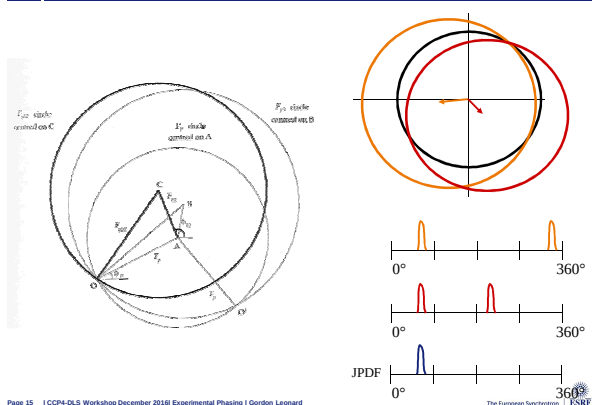
SIR PHASE PROBABILITY DISTRIBUTION



Page 14 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

A SECOND DERIVATIVE RESOLVES THE SIR PHASE AMBIGUITY

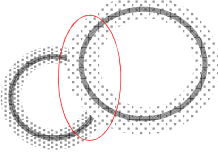
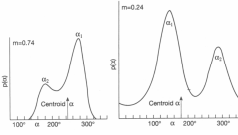


Page 15 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

PHASE PROBABILITY DISTRIBUTIONS

1. Measurement errors in $|F_P|$, $|F_{PH}|$
2. Model of heavy atom substructure is incorrect (minor sites not included)
3. Native and heavy atom derivative crystal are not isomorphous

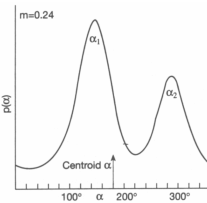
$$F_{ph} \neq F_p + F_h$$



The most probable phase is not the best phase to use in map calculations. The **centroid phase** (best phase) gives the least error in the Fourier map.

$$\rho_{(x,y,z)} = \frac{1}{V_c} \sum_h \sum_k \sum_l m_{hkl} |F_{hkl}| \exp[-2\pi i(hx + ky + lz) - \alpha_{hkl}]$$

Page 16 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard | The European Synchrotron | ESRF

THE FIGURE OF MERIT

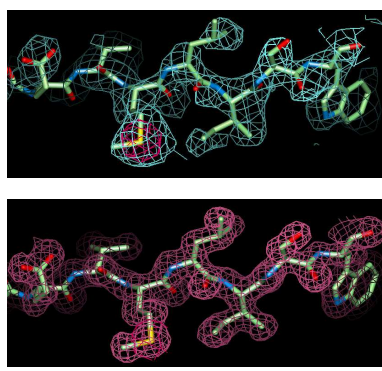
$$m = \frac{\sum P(\Delta\alpha) \cos(\Delta\alpha)}{\sum P(\Delta\alpha)}$$


m = probability that centroid phase is correct
 $\Delta\alpha$ = phase angle measured from the centroid phase.
 $P(\Delta\alpha)$ = probability that $\Delta\alpha$ is the true phase.

$m = 0.866$	$\Rightarrow$ phase error = 30°
$m = 1.0$	$\Rightarrow$ phase error = 0°
$m = 0.0$	$\Rightarrow$ phase error = 90°
$m = 0.5$	$\Rightarrow$ phase error = 60°

Page 17 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard | The European Synchrotron | ESRF

THE RESULT OF IMPROVING EXPERIMENTAL PHASES



Mean phase error = 53°

Phase Improvement

Mean phase error = 41°

Page 18 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard | The European Synchrotron | ESRF

WHAT CAN HAPPEN WHEN A HEAVY ATOM BINDS TO A PROTEIN IN A CRystal?

If heavy atoms don't bind to protein - or are not ordered in the crystal - then there will be no intensity differences

Binding of heavy atom to protein in crystal can cause protein to rotate and/or translate in unit cell and/or unit cell dimensions to change

Non-isomorphism: For a 100 Å cubic unit cell a 0.5% change in unit-cell dimensions or a 0.5° rotation of the molecule within the unit cell would produce an average 15% change in intensity.

Page 19 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard | The European Synchrotron | ESRF

NON ISOMORPHISM (AGAIN)

Adding heavy atoms to protein crystals can cause the protein molecules to rotate and/or translate inside the unit cell. This induces non-isomorphism between native and derivative crystals

$$F_{ph} \neq F_p + F_h$$

To help get around this take advantage of **anomalous scattering**

Page 20 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard | The European Synchrotron | ESRF

NORMAL SCATTERING

X-ray photon interacts with the electrons in the atoms of the sample. Atomic scattering factor f proportional to atomic number, Z , of atoms

$$f = f_o$$

$$F_{hkl} = \sum_j f_j \exp(-2\pi i(hx_j + ky_j + lz_j))$$

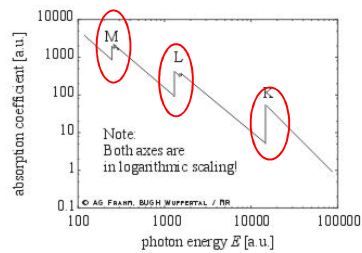
$$F_{hkl} = \sum_j |F_{hkl}| \exp(-2\pi i(\phi))$$

All atoms scatter X-rays in the same way. **Friedel's Law is obeyed.**

$$|F_{hkl}| = |F_{\bar{h}\bar{k}\bar{l}}|$$

Page 21 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard | The European Synchrotron | ESRF

BUT, ELECTRON SHELLS IN ATOMS HAVE X-RAY ABSORPTION EDGES



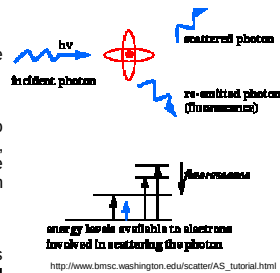
At incident X-ray energies close to absorption edges scattering from atoms is no longer normal

Page 22 | CCP4-OLS Workshop December 2016! Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

NORMAL SCATTERING IS MODIFIED CLOSE TO ABSORPTION EDGE

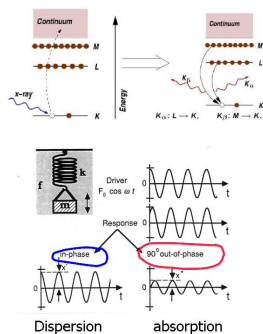
- Photons interact with the electrons in the atoms of the sample.
- Depending on the photon energy the probability of absorption changes.
- At photon energies close to absorption edges of a given atom, some photons are absorbed by the atom and electrons are ejected from its inner shell.
- Electrons from higher energy shells decay to lower energy shells **emitting fluorescence as they do so**



Page 23 | CCP4-OLS Workshop December 2016! Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

SCATTERING FACTORS AT X-RAY ENERGIES CLOSE TO AN ABSORPTION EDGE



Treat electron shells as *harmonic oscillators* with resonance frequencies close to energy of shell absorption edge. Scattering factor modified so that:

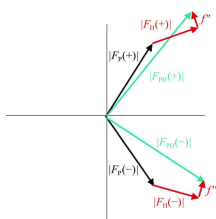
$$f = f_o + f'(\lambda) + if''(\lambda)$$

Scattering factor modified in such a way that it has additional, wavelength(energy)-dependent dispersive (f') (real) and absorptive (f'') (imaginary) terms

Page 24 | CCP4-OLS Workshop December 2016! Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

UNDER ANOMALOUS SCATTERING FRIEDEL'S LAW IS NO LONGER OBEYED



$$f = f_o + f'(\lambda) + if''(\lambda)$$

$$|F_{hkl}| \neq |F_{-h-k-l}| \text{ i.e. } |F(+)| \neq |F(-)|$$

$$\Delta F_{ano} = ||F(+)| - |F(-)||$$

$$F_{mean} = \frac{|F(+)| + |F(-)|}{2}$$

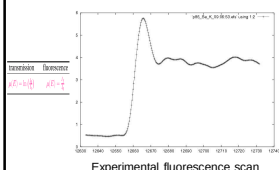
We can use the **anomalous difference** to (help) solve the phase problem. The larger f'' , the bigger anomalous difference and the easier (in principle) it will be to solve the phase problem

Page 25 | CCP4-OLS Workshop December 2016| Experimental Phasing | Gordon Leonard

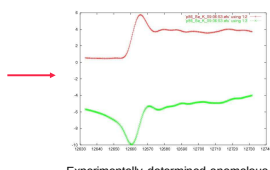
The European Synchrotron | ESRF

DERIVING f'' AND f' FROM FLUORESCENCE SCANS

$$f = f_o + f'(\lambda) + if''(\lambda)$$



Experimental fluorescence scan



Experimentally determined anomalous scattering factors

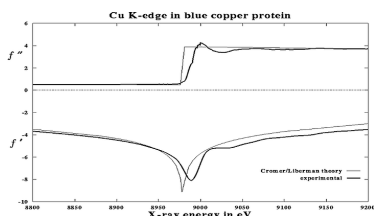
$$f''(\omega) = \left(m\omega \cdot c / 4\pi N e^2 \right) \mu(\omega)$$

$$f'(\omega) = (2/\pi) \int_0^\infty \left[\omega' f''(\omega') / (\omega^2 - \omega'^2) \right] d\omega'$$

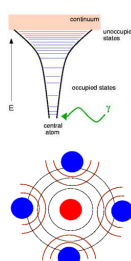
Page 26 | CCP4-OLS Workshop December 2016| Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

MAXIMUM SIGNAL: EXPERIMENT DIFFERS FROM THEORY. DON'T RELY ON TABLES



JM Guss, EA Merritt, RP Phizackerley, B Hedman, M Murata, KO Hodgson, & HC Freeman (1989). "Phase determination by multiple-wavelength X-ray diffraction: crystal structure of a basic blue copper protein from cucumbers". *Science* 241, 806-811.

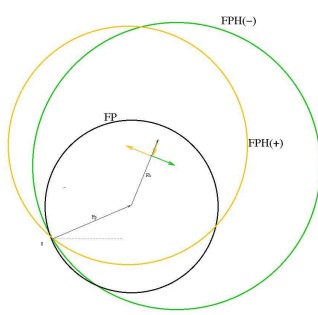


We're not dealing with free atoms. Interaction of anomalous scatterer with other atoms causes deviation from ideality, as do changes in oxidation state. So, we **MUST** measure absorption edge experimentally if we want to be sure to obtain maximum 'signal'

Page 27 | CCP4-OLS Workshop December 2016| Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

USING ANOMALOUS SCATTERING - SIRAS PHASING



$$f = f_o + f'(\lambda) + if''(\lambda)$$

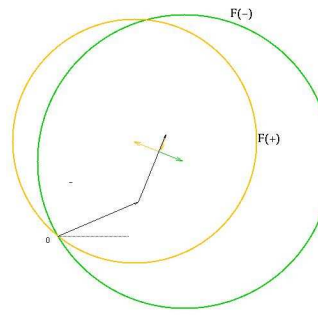
If in an SIR experiment we measure F_p (native crystal) and both F_{ph}^+ and F_{ph}^- for the derivative crystal we get three phasing circles. Hence no phase ambiguity

So, need only one derivative thus less chance of non-isomorphism

Page 28 | CCP4-OLS Workshop December 2016| Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

PHASING USING SAD



$$f = f_o + f'(\lambda) + if''(\lambda)$$

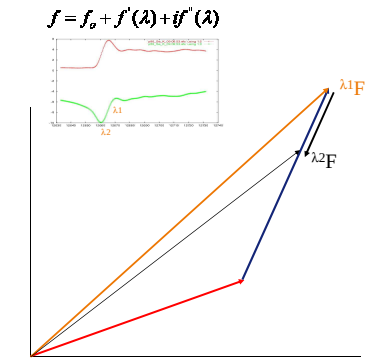
SAD results (like SIR) in phase ambiguity. But can use this technique to solve the phase problem.

Also allows the solution of the phase problem by collecting data from native crystals of macromolecules (i.e. no derivative). Exploits small f'' available from S or P at the wavelengths we can routinely access.

Page 29 | CCP4-OLS Workshop December 2016| Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

WHY DO WE CARE ABOUT f'' ?



$$f = f_o + f'(\lambda) + if''(\lambda)$$

If we collect data at different wavelengths about absorption edge then F_{hkl} at different wavelengths changes

$$\lambda^1 F \neq \lambda^2 F$$

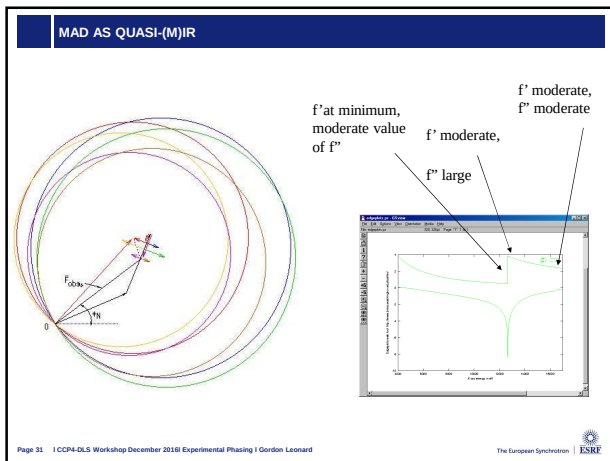
DISPERSIVE difference:

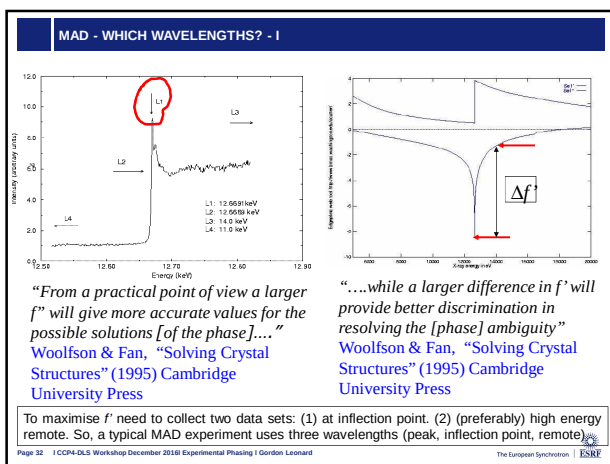
$$\Delta F_{disp} = |\lambda^1 F_{hkl}| - |\lambda^2 F_{hkl}|$$

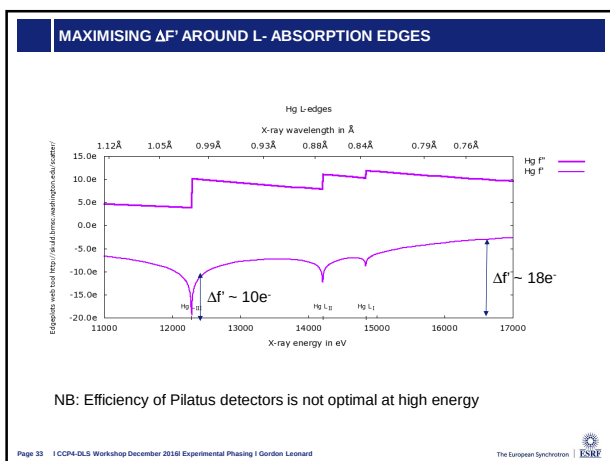
By combining measurement of anomalous and dispersive differences we can solve the phase problem using HA derivatives in MX without measuring data from a native crystal. How?

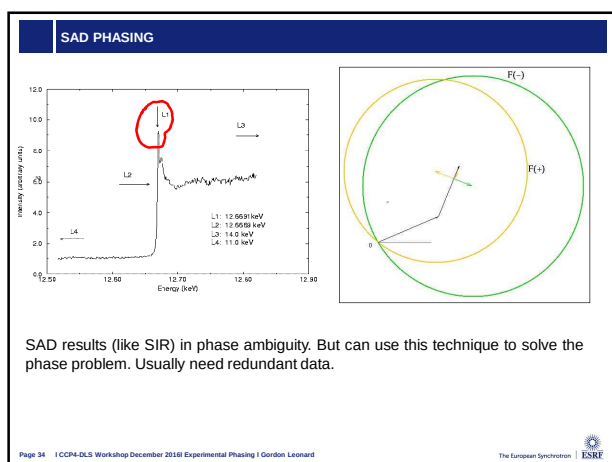
Page 30 | CCP4-OLS Workshop December 2016| Experimental Phasing | Gordon Leonard

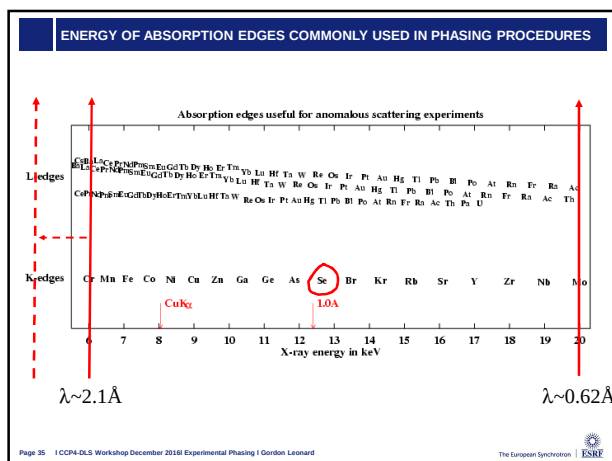
The European Synchrotron | ESRF











PHASING EXPLOITING ANOMALOUS SCATTERING

1. Introduce a small number of anomalously scattering atoms into crystals
2. Collect both F^+ and F^- at one (or more) wavelength(s). If more than one wavelength make sure these are chosen using absorption edge scan.
3. Derive anomalous and (if more than one wavelength) dispersive differences.
4. Use anomalous and/or dispersive differences to locate anomalously scattering atoms
5. Use these positions to initiate phase calculations

Page 36 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

COLLECTING BOTH F(+) AND F(-) (AFTER DAUTER, Z. METHODS IN ENZYMOLOGY, 276)

Crystal System	Laue Group	Total Oscillation required for	
		Standard data collection	Anomalous data collection
Triclinic	-1	180°	360°
Monoclinic	2/m	180° (b*), 90° (a*, c*)	180° (a*, b*, c*)
Orthorhombic	mmm	90° (a*, b*, c*)	90° (a*, b*, c*)
Tetragonal	4/m	90° (a*, b*, c*)	90° (c*), 180° (a*, b*)
	4/mmm	45° (c*), 90° (a*, b*)	45° (c*), 90° (a*, b*)
Trigonal	3/m	60° (c*), 90° (a*, b*)	120° (c*), 180° (a*, b*)
	-31m	30° (c*), 90° (a*, b*)	60° (c*), 180° (a*, b*)
	-3m1	30° (c*), 90° (a*, b*)	30° (c*), 180° (a*, b*)
Hexagonal	6/m	60° (c*), 90° (a*, b*)	60° (c*), 180° (a*, b*)
	6/mmm	30° (c*), 90° (a*, b*)	60° (c*), 90° (a*, b*)
Cubic	23	~60°	~70°
	432	~35°	~45°

Page 37 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

(S)MAD: GOOD & BAD

1. No 'native' crystal – no problems with isomorphism when using heavy atom derivatives
2. Selenomethionine can be introduced into many protein sequences and can be used as a "magic bullet"
3. The 'method of choice' for *de novo* structure solution in macromolecular crystallography
4. Signals are small in comparison to those from isomorphous replacement – data have to be good!
5. In MAD have to collect a lot of data from one crystal – radiation damage is a problem

Page 38 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

(S)MAD INTENSITY DIFFERENCES ARE MUCH SMALLER THAN FOR IR

$$'Dispersive' \quad f' \approx \frac{1}{\sqrt{2}} \frac{\sqrt{N_A} |f'_{A_i} - f'_{A_j}|_{\max}}{<|F_T|>}$$

$$'Anomalous' \quad f'' \approx \frac{1}{\sqrt{2}} \frac{\sqrt{N_A} 2 f''_{A_i, \max}}{<|F_T|>}$$

(J. Smith, Curr. Opin. Struct. Biol. 1, 1991)

K absorption edges: $f'' \sim 4e^-$; can rise to **6-7e⁻** with 'white line', $\Delta f'$ usually has a maximum around **7e⁻**; can rise to **~10e⁻** if far from edge.

L absorption edges: f'' usually has a maximum around **12e⁻**; can rise to **~30e⁻** with 'white line', $\Delta f'$ usually has a maximum around **13e⁻**; can rise to **> 20e⁻** if far from edge

For comparison, Hg has $Z = 78 e^-$.In comparison to signals available in IR, AD signals are small

This knowledge must inform all decisions when collecting data for AD experiments

Need to: MAXIMISE SIGNAL and/or REDUCE ERRORS

Page 39 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

EXPECTED ANOMALOUS SIGNAL

$$\langle \Delta I / I \rangle_{\text{anom}} \approx \frac{1}{\sqrt{2}} \frac{\sqrt{N_A} 2 f^-}{\langle |F_r| \rangle} \quad (\text{J. Smith, Curr. Opin. Struct. Biol. 1, 1991})$$

Screenshot of the Anomalous Scattering Ratios web application. The interface includes a navigation bar with 'Bookstore', 'My Book', and 'Workshops'. The main content area is titled 'Anomalous Scattering Ratios' and contains a form for inputting data. The form includes fields for 'Anomalous scatterer (element symbol)', 'How many anomalous scatterers in molecule', 'Number of residues in molecule', and 'Edge used for MAD experiment'. The 'Edge used for MAD experiment' field is set to '1.2' with a dropdown menu. The 'Submit' button is visible at the bottom of the form. The URL at the bottom of the page is http://www.ruppweb.org/new_comp/anomalous_scattering.htm.

METHODS FOR SMAD DATA COLLECTION - I

The Random Orientation 'Technique' – Here the crystal is mounted in a random orientation with respect to the X-ray beam and goniometer axes. Data are then collected in one contiguous oscillation range such **that complete and redundant anomalous data** are obtained.

The rotation range required can be obtained using various data collection strategy programs (i.e. BEST).

Takes no real account of factors that can adversely affect intensities (absorption, different detector position for diffraction spots)

DATA COLLECTION STRATEGIES AND DATA QUALITY PREDICTIONS

Diagram illustrating the data collection strategy process: Indexing, Integration, and Strategy. The process involves a sequence of steps: Indexing (using X-ray and Laser), Integration (using X-ray and Laser), and Strategy (using X-ray and Laser). The final output is an 'Optimized plan for data collection' and 'Predicted data quality statistics'.

Optimal Plan of data collection
Resolution is selected according to the user's choice
Attenuation = 1.0000

Wavelength (Å)	# of Images	Resolution (Å)	Exposure (seconds)	Overhead
1.0000	247	1.35	0.19	232.4

Resolution limit: 2.38 Angstroms
Resolution goal: 2.38 Angstroms
Total rotation range: 360.0 degrees
Total # of images: 247
Overall Completion: 100.00
factor (center offset): 3.15 (1.00)
factor (center offset): 3.15 (1.00)
Total Data Collection time: 737.8 min (5.19 hours)

Data collection statistics according to the plan

Resolution (Å)	Images	Average	3-Sigma	0-Sigma	Beam	Overhead
10.00	0.00	100.0	2511.0	10.0	1.0	0.00
9.00	0.00	100.0	1812.0	22.0	1.0	0.00
8.00	0.00	100.0	1312.0	40.0	1.0	0.00
7.00	0.00	100.0	1012.0	60.0	1.0	0.00
6.00	0.00	100.0	812.0	80.0	1.0	0.00
5.00	0.00	100.0	612.0	100.0	1.0	0.00
4.00	0.00	100.0	412.0	120.0	1.0	0.00
3.00	0.00	100.0	212.0	140.0	1.0	0.00
2.00	0.00	100.0	112.0	160.0	1.0	0.00
1.00	0.00	100.0	12.0	180.0	1.0	0.00

Resolution goal: 2.38 Angstroms
Resolution limit: 2.38 Angstroms
Total rotation range: 360.0 degrees
Total # of images: 247
Overall Completion: 100.00
factor (center offset): 3.15 (1.00)
factor (center offset): 3.15 (1.00)
Total Data Collection time: 737.8 min (5.19 hours)

Optimized plan for data collection
Predicted data quality statistics

Bourvenkov, G.P. & Popov, A.N. (2010). Optimization of data collection taking radiation damage into account. *Acta Cryst. D66*, 409-419.

METHODS FOR S/MAD DATA COLLECTION - II

The Inverse-beam Method – Crystal mounted in a random orientation. Data are collected in one contiguous oscillation range (from $\phi = X^\circ$ to $\phi = Y^\circ$) such that data is as complete as possible for a 'standard' data collection. The Friedel mates of these reflections are then collected by inverting the orientation of the crystal and collecting the equivalent data (i.e. by collecting from $\phi=180+X^\circ$ to $\phi=180+Y^\circ$). Can also collect in smaller wedges.

Friedel pairs are measured with the same path-length through crystal (absorption).

Friedel pairs are collected close together in time (if we use small wedges).

Friedel pairs measured with the same multiplicity.

METHODS FOR S/MAD DATA COLLECTION - III

The 'Set-Crystal' Method – Here the crystal is mounted such that the principle rotation axis of the crystal point group is parallel to the spindle axis of the goniometer. If this rotation axis is even-fold (i.e. 2, 4, 6) then:

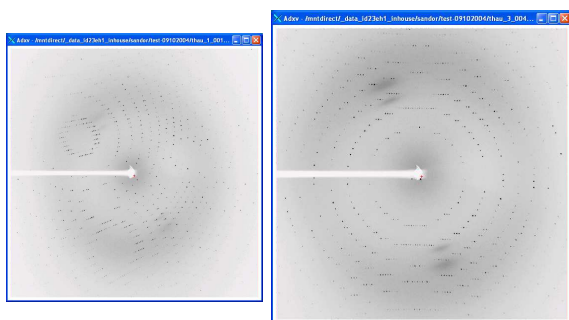
Bijvoet mates are collected on the same oscillation image.

Bijvoet mates are collected after same total exposure time.

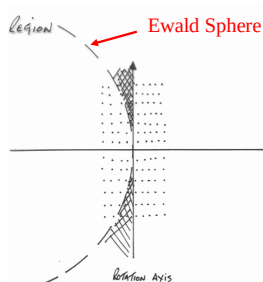
Bijvoet mates suffer from 'same' absorption and radiation damage.

(Mini-)Kappas make this straightforward

USING KAPPA GONIOMETERS



A 'PROBLEM' WITH SET CRYSTAL METHOD



The 'Blind' Region:

Shaded region will never pass through Ewald Sphere. Need to collect data in a second orientation to ensure data is properly complete.

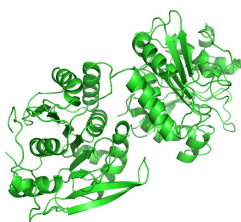
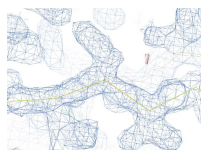
$$\text{Size of blind region: } \frac{\lambda^2}{2d_{\min}^2}$$

Page 46 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

HOW DO WE KNOW THAT OUR PHASING EXPERIMENT HAS WORKED?

Electron density map is the best metric



Page 47 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

BUT, 'SUCCESS' INDICATORS FOR EXPERIMENTAL PHASING PROTOCOLS ARE AVAILABLE AT ALL STAGES OF THE EXPERIMENT

- Before arriving at beamline

If possible, check that heavy atom is bound

What is the expected anomalous signal?

- Before data collection

Experimentally measure absorption edge

What will the data quality be?

- After data integration and scaling

What is the real data quality?

Is there any anomalous signal in the data?

How far (to what resolution) does the anomalous signal extend?

Are more data needed to try to improve things?

- Substructure determination

- Correct/reasonable number of heavy atom sites
- Correlation coefficients

- Phasing

- One hand significantly better than the other?
- What does 'best' electron density map look like?
- Does auto-building work?

Page 48 | CCP4-DLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

The European Synchrotron | ESRF

AFTER DATA PROCESSING – ARE THE DATA GOOD?

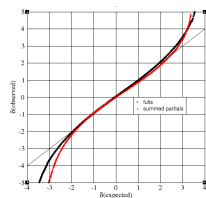
```

N 1/d^2 Delta(h) FWHM Reflnt Recn Panoce Nmon Av_1 SIGMA 1/sigma ad Rn(I/ed) Rhoan Rref Rcent FRCR1AS Rbias
11
1 0.0050 4.32 0.023 0.019 0.023 0.045 1368 3514 168.0 21.0 159 51.9 10844 1496 314 -0.003 4164
2 0.0050 4.47 0.024 0.023 0.025 0.058 2491 3951 185.1 21.4 191 47.7 18062 1044 400 -0.004 6979
3 0.0700 1.45 0.028 0.023 0.027 0.031 3545 5180 189.1 18.1 143 46.0 24152 3980 421 -0.005 9194
4 0.1000 1.16 0.028 0.023 0.027 0.033 4920 3845 191.5 19.9 202 43.9 23460 4742 443 -0.001 11352
5 0.1250 1.45 0.029 0.024 0.028 0.033 4920 3128 111.4 20.9 127 41.1 33306 5145 499 0.002 12929
6 0.1500 2.58 0.034 0.020 0.029 0.072 5515 1654 89.3 18.5 95 38.5 37161 5971 484 0.005 14544
7 0.1750 2.39 0.037 0.011 0.028 0.072 6015 1384 77.1 17.9 84 35.4 40621 6472 495 0.002 15844
8 0.2000 2.14 0.039 0.033 0.030 0.072 6481 1241 72.3 17.2 79 33.5 43244 4934 495 0.006 17079
9 0.2250 2.11 0.044 0.030 0.031 0.075 6852 1098 71.5 15.3 73 31.9 44190 7104 487 0.003 18090
10 0.1500 2.00 0.054 0.046 0.033 0.080 7265 855 66.1 12.9 61 29.2 49015 7709 494 0.006 19151

```

$$R_{\text{merge}} = \sum_{\text{all}} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{\text{all}} I_i(hkl)$$

Normal probability plot



$$\delta_{\text{all}} = (I_{\text{all}} - \langle I_{\text{all}} \rangle) / \sigma^2(I_{\text{all}}); \text{Slope} = 1; \text{Intercept} = 0$$

```

=====
Normal probability analysis, by run & partiality
=====
Run number: 1, fulls =====
All data: 10820 Number: 1.051 Slope Intercept:
Data within expected delta 0.90: 11709 4.841 0.004
=====
Run number: 1, summed partials =====
All data: 10809 Number: 1.051 Slope Intercept:
Data within expected delta 0.90: 98013 0.884 0.013
=====
Run number: 1, fulls against fulls only =====
All data: 128207 Number: 1.387 Slope Intercept:
Data within expected delta 0.90: 10907 4.832 0.021
=====

```

AFTER DATA PROCESSING – IS THERE ANY 'SIGNAL' IN THE DATA?

```

N 1/d^2 Delta(h) FWHM Reflnt Recn FWHM Nmon Av_1 SIGMA 1/sigma ad Rn(I/ed) Rhoan Rref Rcent FRCR1AS Rbias
11
1 0.0050 4.32 0.023 0.019 0.023 0.045 1368 3514 168.0 21.0 159 51.9 10844 1496 314 -0.003 4164
2 0.0050 4.47 0.024 0.023 0.025 0.058 2491 3951 185.1 21.4 191 47.7 18062 1044 400 -0.004 6979
3 0.0700 1.45 0.028 0.023 0.027 0.031 3545 5180 189.1 18.1 143 46.0 24152 3980 421 -0.005 9194
4 0.1000 1.16 0.028 0.023 0.027 0.033 4920 3845 191.5 19.9 202 43.9 23460 4742 443 -0.001 11352
5 0.1250 1.45 0.029 0.024 0.028 0.033 4920 3128 111.4 20.9 127 41.1 33306 5145 499 0.002 12929
6 0.1500 2.58 0.034 0.020 0.029 0.072 5515 1654 89.3 18.5 95 38.5 37161 5971 484 0.005 14544
7 0.1750 2.39 0.037 0.011 0.028 0.072 6015 1384 77.1 17.9 84 35.4 40621 6472 495 0.002 15844
8 0.2000 2.14 0.039 0.033 0.030 0.072 6481 1241 72.3 17.2 79 33.5 43244 4934 495 0.006 17079
9 0.2250 2.11 0.044 0.030 0.031 0.075 6852 1098 71.5 15.3 73 31.9 44190 7104 487 0.003 18090
10 0.1500 2.00 0.054 0.046 0.033 0.080 7265 855 66.1 12.9 61 29.2 49015 7709 494 0.006 19151

```

$$R_{\text{merge}} = \sum_{\text{all}} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{\text{all}} I_i(hkl)$$

$$R_{\text{anom}} = \sum_{\text{all}} |I^+(hkl) - I^-(hkl)| / \sum_{\text{all}} |I^+(hkl) + I^-(hkl)| / 2$$

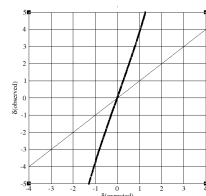
$$R_{\text{anom}} > R_{\text{merge}} ?$$

AFTER DATA PROCESSING – IS THERE ANY 'SIGNAL' IN THE DATA?

The presence of anomalous signal can be deduced from a normal probability plot of the normalized measured anomalous differences. The slope of the central region of this plot will be > 1 if the anomalous differences are larger than expected from their standard deviations (i.e. if there is significant measured anomalous signal).

$$\delta_{\text{obs}} = (I^+ - I^-) / [\sigma^2(I^+) + \sigma^2(I^-)]^{1/2}$$

Anomalous differences

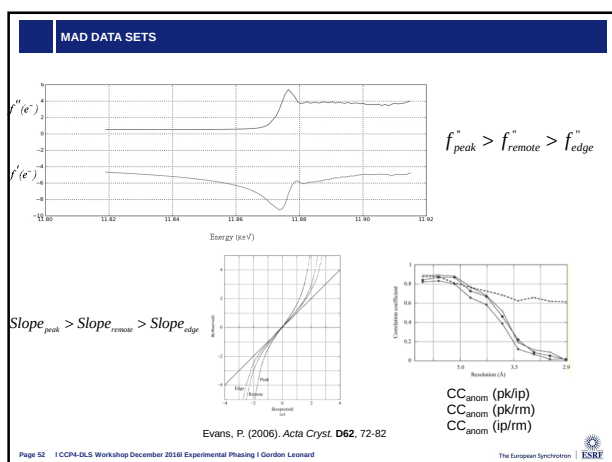


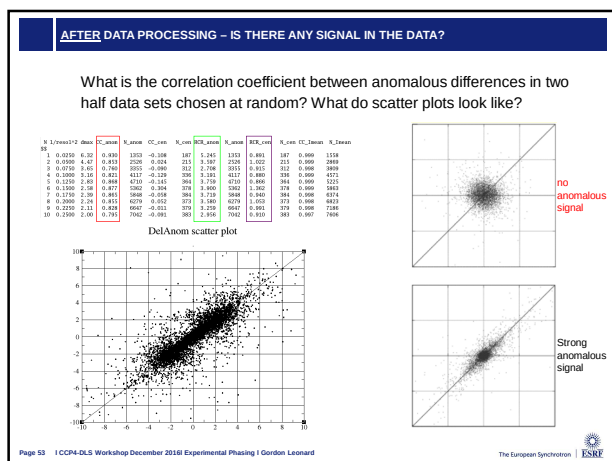
```

=====
Normal probability analysis of anomalous differences
=====
Anomalous differences =====
All data: 48958 Number: 4.851 Slope Intercept:
Data within expected delta 0.90: 30954 3.725 0.043
=====

```

(Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81-86; Evans, P. (2006). *Acta Cryst.* **D62**, 72-82.)





DATA PROCESSING – SUMMARY

Summary data for Project: Unspecified Crystal: Unspecified Dataset: Unspecified

	Overall	InnerShell	OuterShell
Low resolution limit	39.81	39.81	2.11
High resolution limit	2.00	6.32	2.00
Rmerge	0.033	0.023	0.054
Rmerge in top intensity bin	0.025	-	-
Rmerge (within I+/I-)	0.039	0.027	0.065
Rmerge (all I+ & I-)	0.075	0.073	0.103
Rpin (within I+/I-)	0.021	0.015	0.035
Rpin (all I+ & I-)	0.030	0.028	0.040
Fractional partial bias	0.000	-0.003	0.006
Total number of observations	333117	10954	49295
Total number unique	53020	1724	7768
Mean(I)/sd(I)	35.6	54.3	24.7
Completeness	98.1	93.4	97.6
Multiplicity	6.2	6.4	6.3
Anomalous completeness	98.0	97.5	97.5
Anomalous multiplicity	3.2	3.6	3.3
DelAnom correlation between half-sets	0.833	0.930	0.795
Mid-Slope of Anom Normal Probability	3.724	-	-

FAE: ΔF/F ~5.4%; λ = 0.97 Å (S.G. = P2₁2₁2₁; 160° data)

Page 54 | CCP4-OLS Workshop December 2016 | Experimental Phasing | Gordon Leonard

Figure 1 displays the iterative refinement process of a protein structure, showing the model at different stages of refinement (a, b, c, d) and the corresponding plots of Scaled Difference CC vs. Resolution, 2Fo-Fo Difference Map, and Peak Number vs. Peak Number.

Panel (a): Initial model (143KDa version 01, 5HELIC model). The plot shows Scaled Difference CC vs. Resolution (Å) from 0 to 40. The R-factor is 0.28 and R-free is 0.30. The 2Fo-Fo Difference Map shows a clear signal at 2.0 Å resolution. The Peak Number vs. Peak Number plot shows a sharp drop at 2.0 Å resolution.

Panel (b): Model after 10 cycles of refinement (143KDa version 02, 5HELIC model). The plot shows Scaled Difference CC vs. Resolution (Å) from 0 to 40. The R-factor is 0.20 and R-free is 0.25. The 2Fo-Fo Difference Map shows a clear signal at 2.0 Å resolution. The Peak Number vs. Peak Number plot shows a sharp drop at 2.0 Å resolution.

Panel (c): Model after 20 cycles of refinement (143KDa version 03, 5HELIC model). The plot shows Scaled Difference CC vs. Resolution (Å) from 0 to 40. The R-factor is 0.15 and R-free is 0.20. The 2Fo-Fo Difference Map shows a clear signal at 2.0 Å resolution. The Peak Number vs. Peak Number plot shows a sharp drop at 2.0 Å resolution.

Panel (d): Final model (143KDa version 04, 5HELIC model). The plot shows Scaled Difference CC vs. Resolution (Å) from 0 to 40. The R-factor is 0.12 and R-free is 0.18. The 2Fo-Fo Difference Map shows a clear signal at 2.0 Å resolution. The Peak Number vs. Peak Number plot shows a sharp drop at 2.0 Å resolution.

A slightly less clear-cut example....

AFTER DATA PROCESSING - IS THERE ANY SIGNAL IN THE DATA?

AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

R_{p.l.m.} (Precision-Indicating Merging R factor; Weiss, M. S. (2001). *J. Appl. Cryst.* **34, 130-135; a measure of the quality of the data after averaging the multiple measurements.**

A better indicator is the ratio $R_{\text{anom}}/R_{p.l.m.}$; R_{anom} = signal, $R_{p.l.m.}$ = noise. (Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9, 771-777; Panjikar, S. & Tucker, P. (2002). *J. Appl. Cryst.* **35**, 261-266.)**

$R_{\text{anom}} / R_{p.l.m.} > 1.2?$

Page 58 | CCP4-OLS Workshop December 2016 | Experimental Phasing I | Gordon Leonard

AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

Normal probability plot - anomalous differences

Anomalous differences

Normal probability analysis of anomalous differences

***** anomalous differences *****

Number: 11667

All data: 11667

Data within expected delta 0.00: 11667

Slope: 1.448

Intercept: -0.030

Delta: -0.047

(Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81-86; Evans, P. (2006). *Acta Cryst.* **D62**, 72-82.)

Page 59 | CCP4-OLS Workshop December 2016 | Experimental Phasing I | Gordon Leonard

AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA?

What is the correlation coefficient between anomalous differences in two half data sets chosen at random? Where is the point that CC_{anom} drops below 30%?

no anomalous signal

Strong anomalous signal

Page 60 | CCP4-OLS Workshop December 2016 | Experimental Phasing I | Gordon Leonard

AFTER DATA PROCESSING – IS THERE ANY SIGNAL IN THE DATA (XDS)?

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= 0.0 AS FUNCTION OF RESOLUTION													
RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	COMPLETENESS UNIQUE POSSIBLE	R-FACTOR OBSERVED	R-FACTOR COMPARED EXPECTED	1/SIGMA	R-merge	Rmerge-I	Rmerge-F	Rmerge-Corr	Sigma/I	Sigma/F	Nano	
9.77	2351	654	718	91.1%	3.8%	4.0%	2315	31.07	4.4%	2.4%	88%	3.023	302
6.91	4854	1319	1352	97.6%	3.8%	4.1%	4849	30.13	4.4%	2.6%	89%	3.162	646
5.64	6028	1673	1750	95.6%	4.3%	4.3%	6017	27.53	5.0%	3.1%	87%	3.610	808
4.88	6924	1949	2038	95.6%	4.3%	4.3%	6900	27.19	5.1%	3.2%	84%	2.917	933
4.37	7948	2216	2311	95.8%	4.0%	4.2%	7921	27.92	4.7%	2.6%	78%	2.307	1062
3.99	8783	2457	2604	94.4%	4.2%	4.3%	8740	26.59	5.0%	3.1%	76%	2.200	1183
3.69	9501	2642	2771	95.2%	4.4%	4.6%	9480	24.86	5.2%	3.6%	72%	2.046	1274
3.45	10579	2899	3034	95.6%	4.7%	4.8%	10555	23.39	5.6%	3.9%	67%	1.895	1408
3.26	11467	3092	3250	95.1%	5.4%	5.4%	11447	21.10	6.3%	4.6%	60%	1.668	1512
3.09	12284	3276	3426	95.6%	6.6%	6.7%	12271	17.85	7.7%	6.3%	63%	1.771	1612
2.84	12761	3350	3538	94.7%	6.0%	6.3%	12750	15.51	9.3%	7.4%	59%	1.597	1661
2.62	13405	3504	3696	97.5%	11.4%	10.1%	13399	13.58	17.9%	9.2%	54%	1.445	1745
2.43	13954	3626	3886	93.3%	11.7%	12.8%	13944	11.32	13.6%	12.4%	49%	1.333	1809
2.41	7491	3628	4084	88.4%	11.4%	11.4%	7389	7.84	15.7%	21.0%	34%	1.134	1609
2.52	7084	3602	4128	87.3%	11.6%	12.0%	6964	6.86	16.3%	23.7%	28%	1.023	1643
2.44	7454	3600	4362	86.7%	12.4%	13.4%	7308	6.09	18.1%	25.0%	29%	1.003	1700
2.37	7440	3609	4462	85.2%	15.9%	16.7%	7274	5.14	22.4%	33.3%	24%	0.933	1710
2.30	7359	3779	4526	83.3%	17.4%	18.0%	7160	4.51	24.6%	36.9%	19%	0.921	1677
2.24	7779	4032	4876	82.7%	21.2%	22.5%	7494	3.93	29.9%	44.3%	12%	0.851	1725
2.18	1113	2622	4738	59.4%	23.3%	24.9%	4662	2.47	32.9%	46.3%	12%	0.840	1057
total	170599	58120	65586	88.6%	5.5%	5.7%	168859	14.16	6.6%	10.3%	60%	1.569	27156

Bovine Trypsin; S-SAD; S.G. = P₂₁₂₁₂₁; 360° data; λ = 1.77Å (7.0 keV)
(ΔF/F ~1.4%)

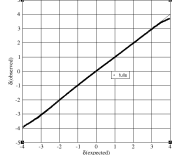
AFTER DATA PROCESSING – ARE THE DATA GOOD?

N	1/d ² Bin(A)	Ravg	RFull	Rms	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Pd(I/sd)	Nmeas	Nref	Ncent	PROB15	Nbias	
1	0.0250	6.32	0.028	0.028	0.028	0.020	446	1582	76	20.7	64	74.2	8681	684	239	-	0
2	0.0501	4.47	0.028	0.028	0.028	0.015	923	1850	92	20.2	76	75.4	16192	1167	244	-	0
3	0.0751	3.45	0.029	0.029	0.029	0.014	1211	2288	116	19.7	96	72.5	19596	1446	234	-	0
4	0.1002	3.16	0.036	0.036	0.033	0.014	1440	1530	87	17.5	74	65.4	23564	1675	225	-	0
5	0.1252	2.83	0.040	0.040	0.033	0.010	1675	852	60	14.3	55	47.4	26943	1996	229	-	0
6	0.1503	2.58	0.060	0.060	0.037	0.023	1894	515	47	10.9	49	34.1	27966	2053	224	-	0
7	0.1753	2.39	0.090	0.090	0.041	0.029	1989	424	50	8.6	53	26.9	31613	2211	220	-	0
8	0.2003	2.23	0.127	0.127	0.046	0.029	2131	329	54	6.1	58	19.9	35941	2362	220	-	0
9	0.2254	2.11	0.200	0.200	0.053	0.027	2248	245	63	5.9	66	13.6	35181	2457	200	-	0
10	0.2504	2.00	0.376	0.376	0.062	0.109	2337	156	75	2.1	79	7.5	36106	2541	211	-	0

$$R_{\text{merge}} = \frac{\sum_i \sum_h |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_i \sum_h I_i(hkl)}$$

Normal probability plot

with statistics for XDS processed data



Normal probability analysis, by row & partiality

----- Row number: 1, full -----

All data: Nmeas: 8681, Nref: 684, Ncent: 239, Prob15: 0.000, Nbias: 0.000

Data within expected delta 0.50: 164591, 0.501: 0.003

AFTER DATA PROCESSING - ANY SIGNAL?

N	1/d ²	Delta(A)	Ravg	Rhalf	Rsum	Ranom	Nanom	Av_I	SDNA	I/sigma	sd	Pn(I/sd)	Nneas	Nref	Ncont	FRIBIAS	Nbias
\$																	
1	0.0050	6.32	0.028	0.028	0.028	0.020	446	1582	76	20.7	64	74.2	8681	684	238	-	0
2	0.0501	4.47	0.028	0.028	0.028	0.015	923	1850	22	20.2	75	75.4	16192	1157	244	-	0
3	0.0751	3.45	0.029	0.029	0.029	0.014	1211	2288	116	19.7	96	72.5	19596	1446	234	-	0
4	0.1002	3.16	0.036	0.036	0.033	0.014	1442	1532	87	17.5	74	65.4	23264	1675	229	-	0
5	0.1252	2.83	0.048	0.048	0.033	0.010	1675	852	60	14.3	55	47.4	26343	1906	229	-	0
6	0.1503	2.58	0.069	0.069	0.037	0.021	1854	515	47	10.9	49	34.1	29966	2053	224	-	0
7	0.1753	2.39	0.090	0.090	0.041	0.029	1989	404	50	8.6	53	26.9	31619	2211	220	-	0
8	0.2003	2.23	0.127	0.127	0.046	0.029	2131	329	54	6.1	58	19.9	35641	2362	220	-	0
9	0.2254	2.11	0.200	0.200	0.053	0.057	2248	245	63	3.9	66	13.6	35181	2457	208	-	0
10	0.2504	2.00	0.378	0.378	0.062	0.109	2337	156	75	2.1	79	7.5	36106	2541	211	-	0

$$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} I_i(hkl)$$

$$R_{anom} = \sum_{hkl} |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} |I_i(hkl) + I_j(hkl)| / 2$$

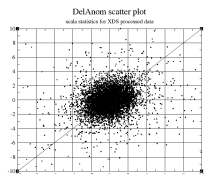
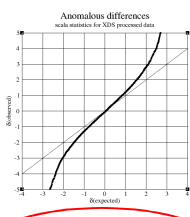
$$R_{p.i.a.} = \sum_{hkl} [1/(N-1)] \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} I_i(hkl)$$

$$R_{anom} < R_{merge}$$

Ranom	Rpin	N	1/resol ²	dnax	CR_anom	N_anom	CR_cen	N_cen	RCR_anom	N_anom	RCR_cen	N_cen	CR_Ineas	N_Ineas
\$														
0.020	0.011	1	0.0050	6.32	0.653	442	-0.139	230	2.107	442	0.863	230	1.000	683
0.015	0.011	2	0.0501	4.47	0.566	912	0.014	232	1.898	912	1.009	232	1.000	1167
0.014	0.012	3	0.0751	3.45	0.323	1199	0.009	224	1.413	1199	1.010	224	0.999	1442
0.014	0.014	4	0.1002	3.16	0.168	1430	-0.028	211	1.185	1430	0.772	211	0.999	1671
0.018	0.019	5	0.1252	2.83	0.270	1653	0.065	217	1.331	1653	1.073	217	0.999	1903
0.023	0.027	6	0.1503	2.58	0.211	1889	-0.065	216	1.242	1889	0.795	216	0.999	2047
0.029	0.035	7	0.1753	2.39	0.172	1972	0.049	212	1.190	1972	1.044	212	0.998	2208
0.039	0.050	8	0.2003	2.23	0.107	2131	-0.046	214	1.115	2131	0.761	214	0.996	2351
0.057	0.078	9	0.2254	2.11	0.049	2232	-0.056	195	1.051	2232	0.952	195	0.992	2454
0.100	0.149	10	0.2504	2.00	0.064	2311	-0.043	209	1.066	2311	0.760	209	0.975	2534

$$R_{anom} / R_{p.i.a.} < 1.2$$

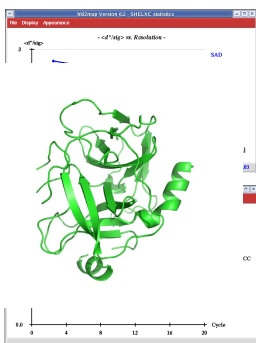
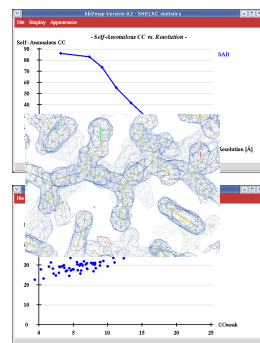
AFTER DATA PROCESSING - IS THERE ANY SIGNAL IN THE DATA?



Normal probability analysis of anomalous differences

Anomalous differences	Number	Significance
All data	10230	-0.000
Data within expected delta 0.30	10230	-0.000

AFTER DATA PROCESSING - SHELXC/SHELXD/SHELXE OUTPUT



- Taylor, G.L. (2010). Introduction to phasing. *Acta Cryst.* **D66**, 325-338.
- Dauter, Z., Dauter, M. & Dodson, E.J. (2002). Jolly SAD. *Acta Cryst.* **D58**, 494-506.
- Hendrickson W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*. **254**, 51-8.
- Shekdrick, G. M. (2008). A short history of SHELX. *Acta Cryst.* **A64**, 112-122.
- Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Cryst.* **D67**, 282-292.
- Bourenkov, G.P. & Popov, A.N. (2010). Optimization of data collection taking radiation damage into account. *Acta Cryst.* **D66**, 409-419.
- Incardona, M.-F., Bourenkov, G.P., Levik, K., Pieritz, R.A., Popov, A.N. & Svensson, O. (2009). EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Rad.* **16**, 872-879.
- Leonard, G.A., Sole, V.A., Beteva, A., Gabadinho, J., Gujarró, M., McCarthy, J., Marrocchelli, D., Nurizzo, D., McSweeney, S. & Mueller-Dieckmann, C. (2009). Online collection and analysis of X-ray fluorescence spectra on the macromolecular crystallography beamlines of the ESRF. *J. Appl. Cryst.* **42**, 333-335.