
Revisiting Some Basic Concepts: Resolution, Diffraction et al

Clemens Vornrhein
Global Phasing Ltd.

CCP4/Diamond workshop
12/2019

Too much focus on single number (resolution) to describe model quality from X-Ray diffraction?

intitle:resolution ccp4bb site:www.mail-archive.com

All News Videos Shopping Maps More Settings

About 564 results (0.34 seconds) **564**

intitle:diffraction ccp4bb site:www.mail-archive.com

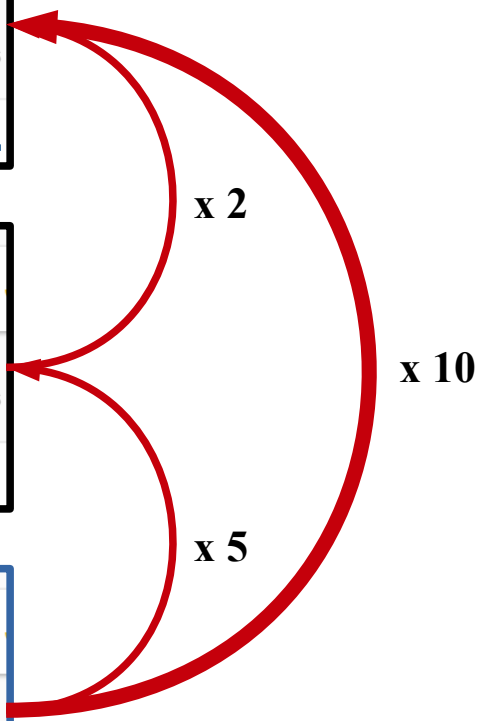
All Videos Shopping News Images More Settings

About 283 results (0.39 seconds) **283**

intitle:quality ccp4bb site:www.mail-archive.com

All Maps Shopping News Videos More Settings

About 58 results (0.31 seconds) **58**



x 2

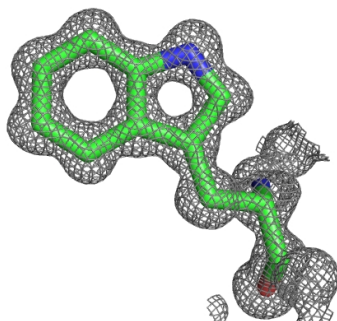
x 10

x 5

Resolution = ability to resolve detail (optical concept)

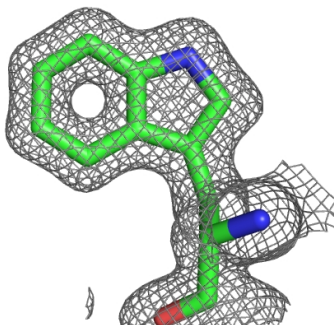
1.0 Å2CNQ (1.00 Å)
=13.5

B(A244)=7.3

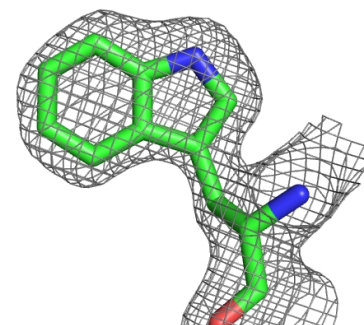
**1.5 Å**6N19 (1.50 Å)
=27.0

B(A522)=17.6

1.4 Å

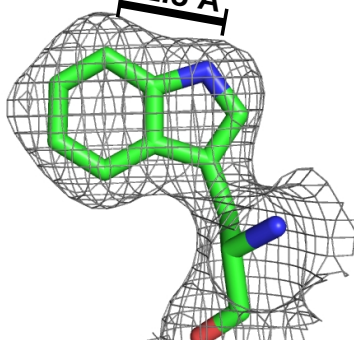
**2.0 Å**6J66 (1.95 Å)
=28.8

B(A181)=27.5

**2.5 Å**6NT2 (2.48 Å)
=28.4

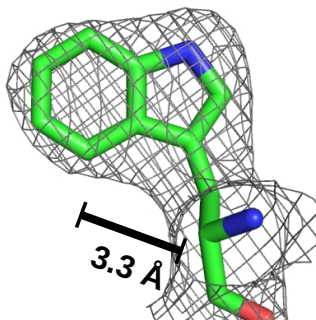
B(B216)=17.9

2.5 Å

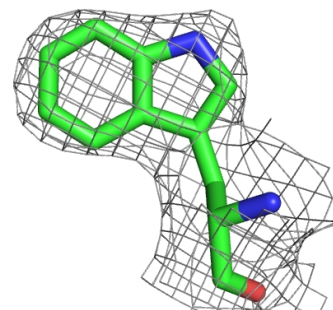
**3.0 Å**6JJ9 (3.00 Å)
=56.0

B(A201)=31.5

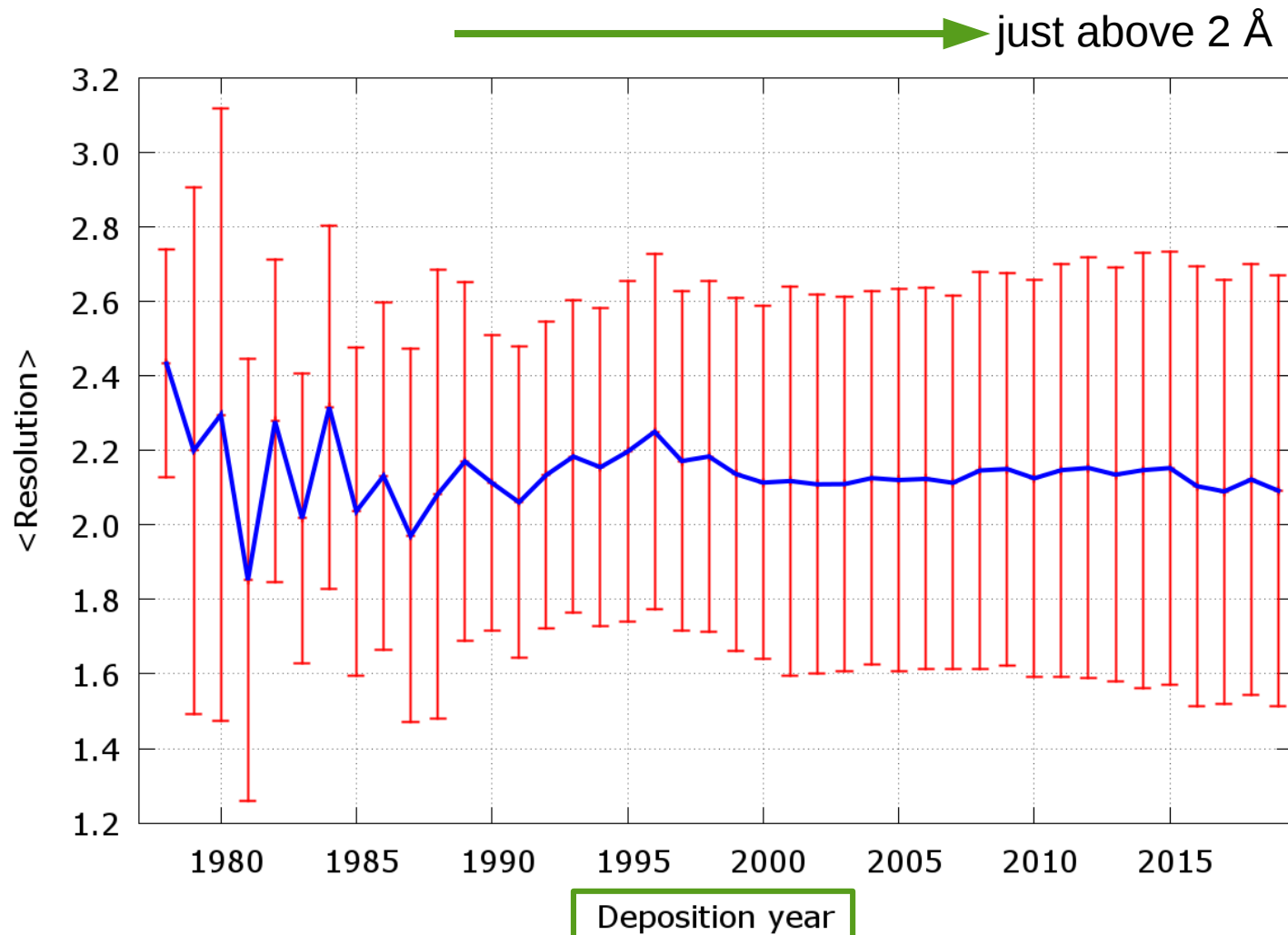
3.3 Å

**3.5 Å**6DC5 (3.50 Å)
=72.2

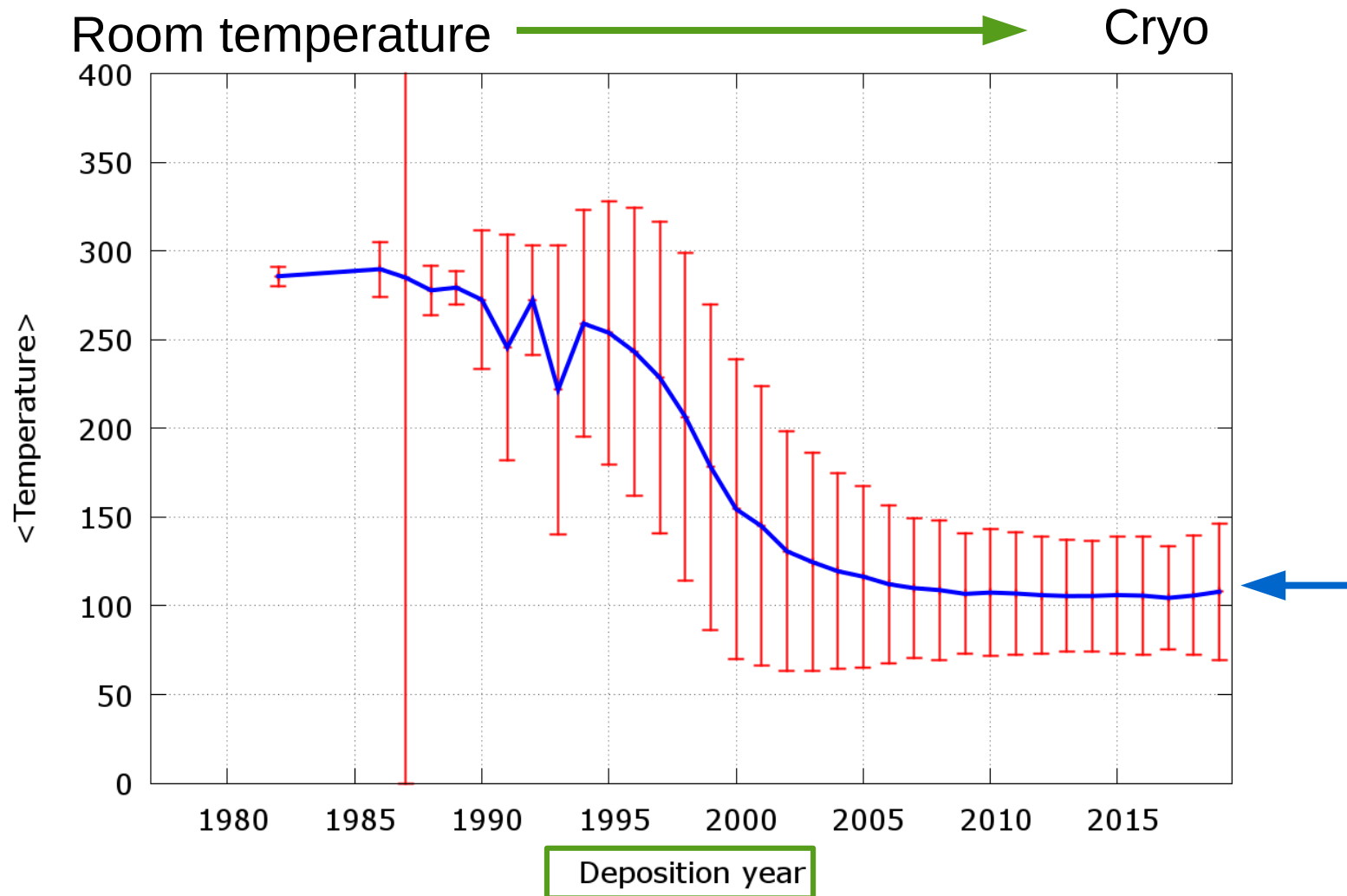
B(I35)=31.0

**2mFo-DFc maps after BUSTER refinement at 1.0 rms**

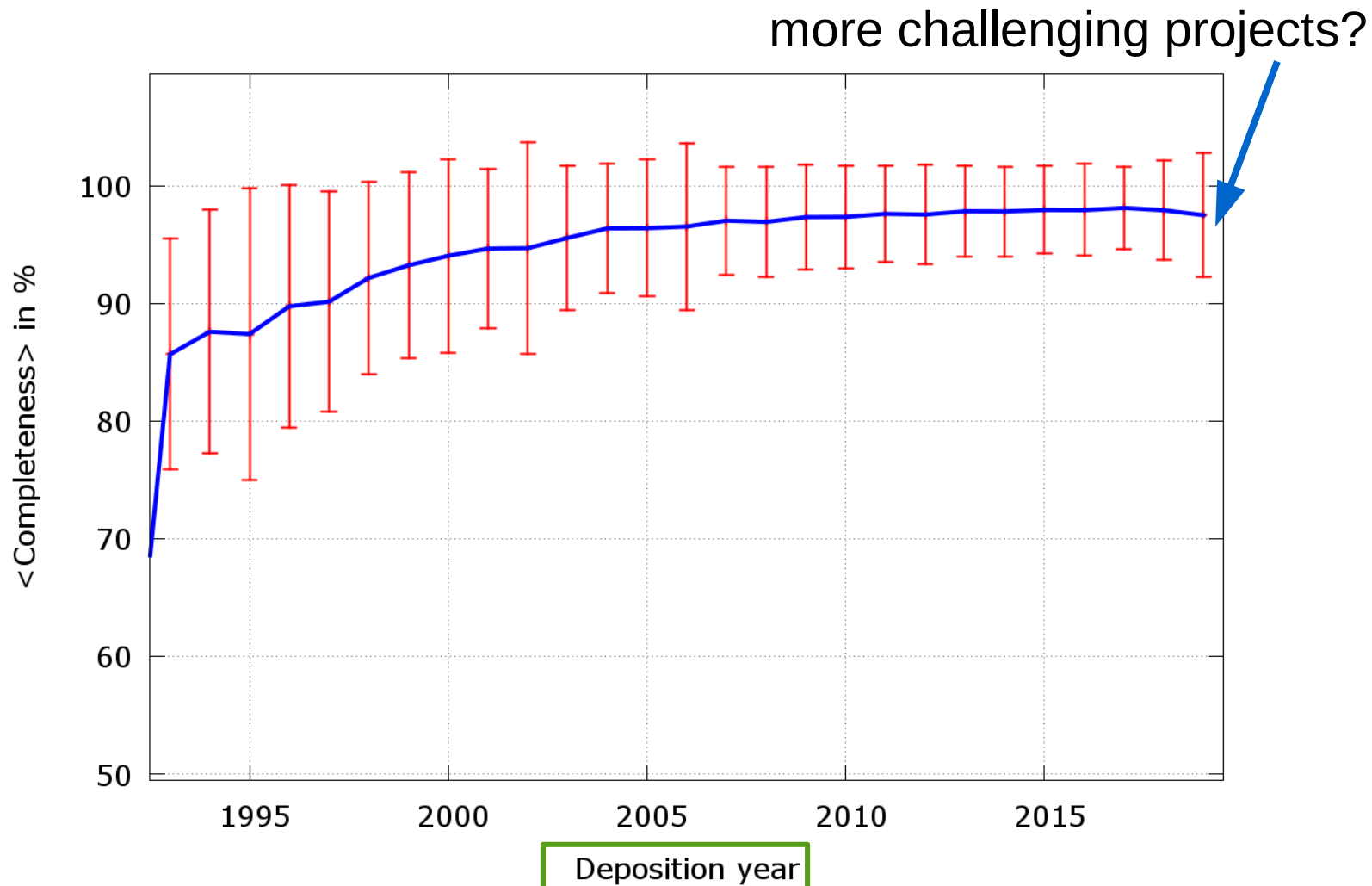
PDB (X-Ray): resolution



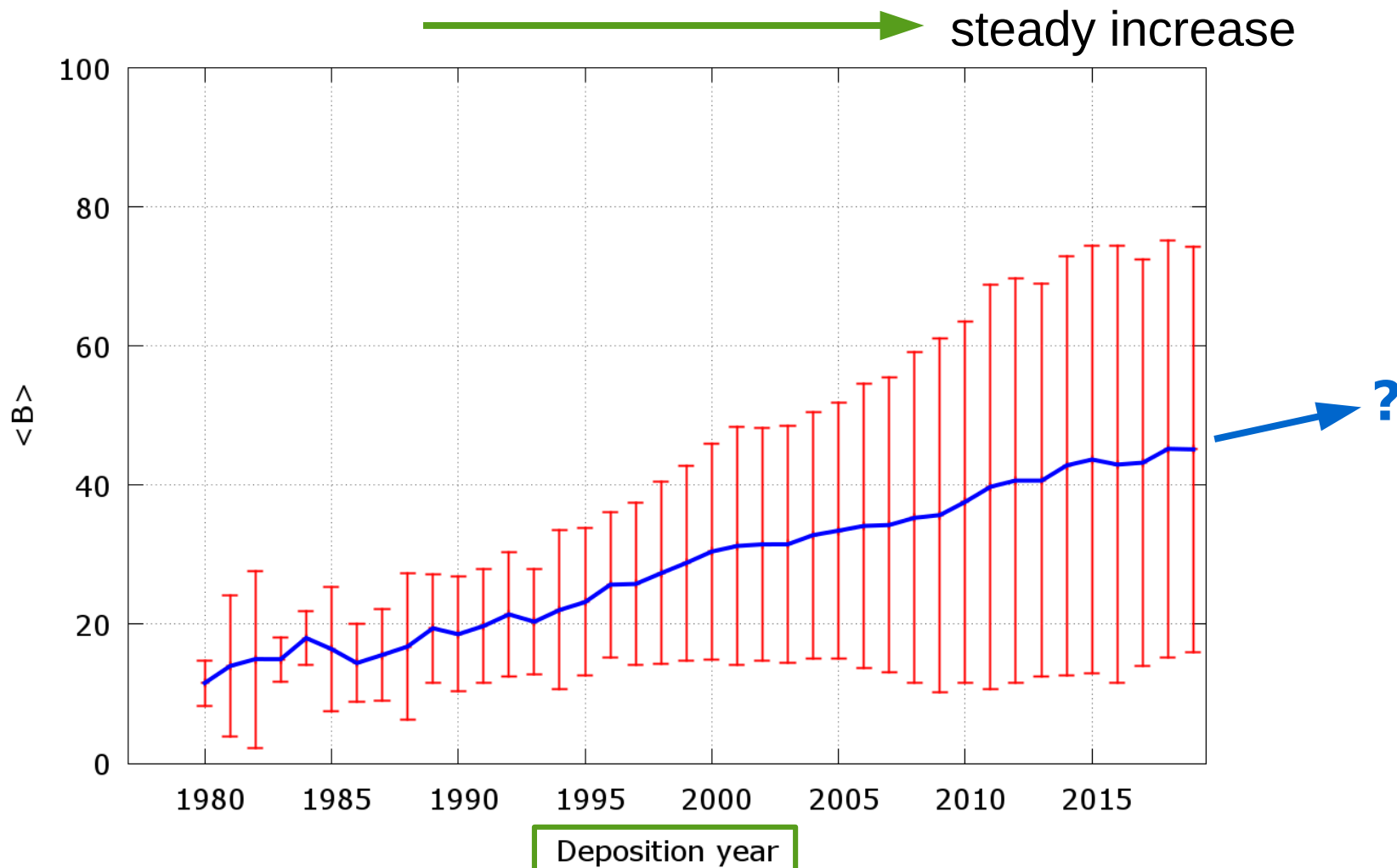
PDB (X-Ray): temperature



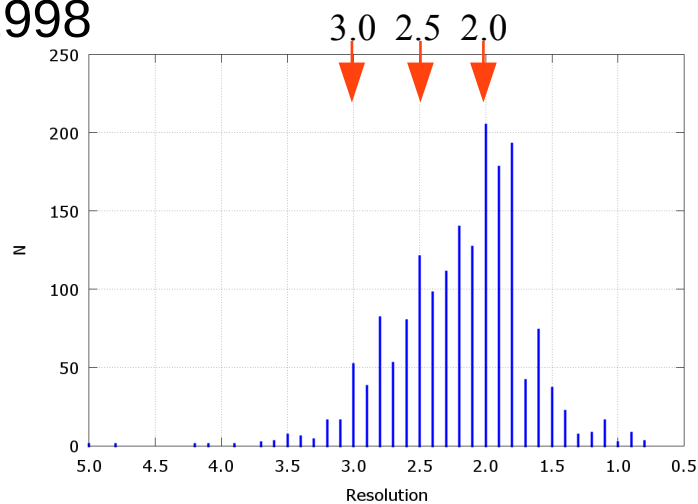
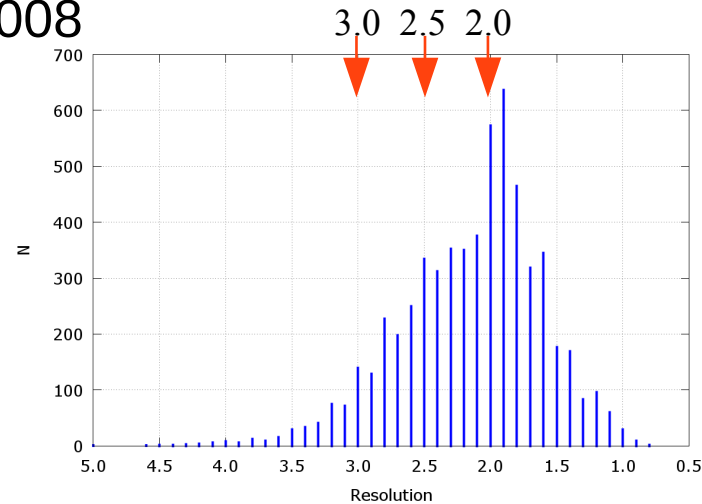
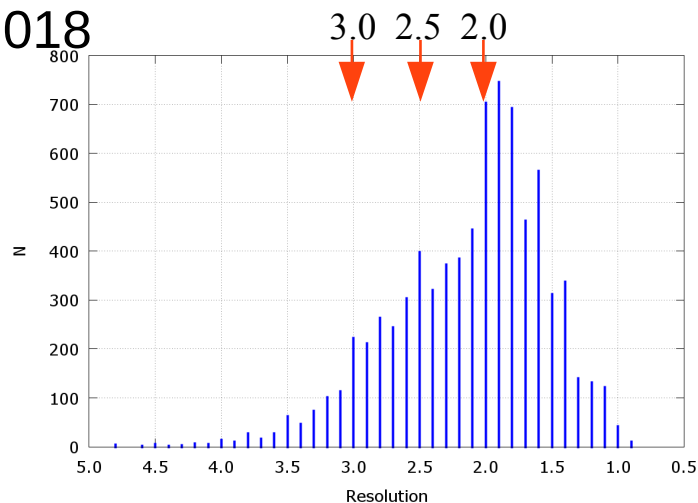
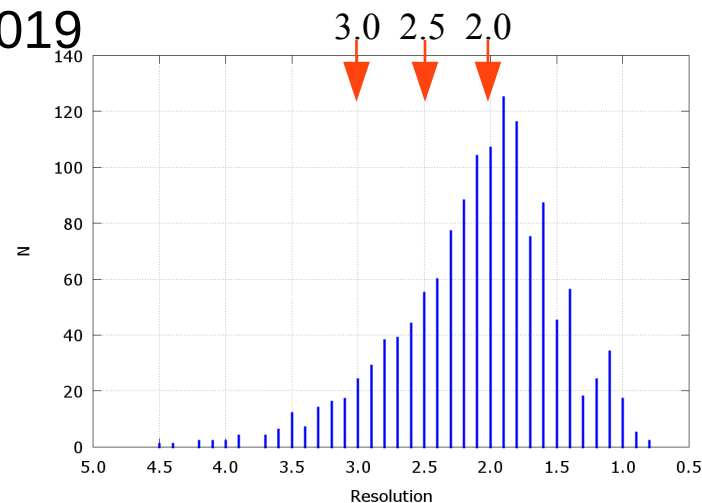
PDB (X-Ray): completeness



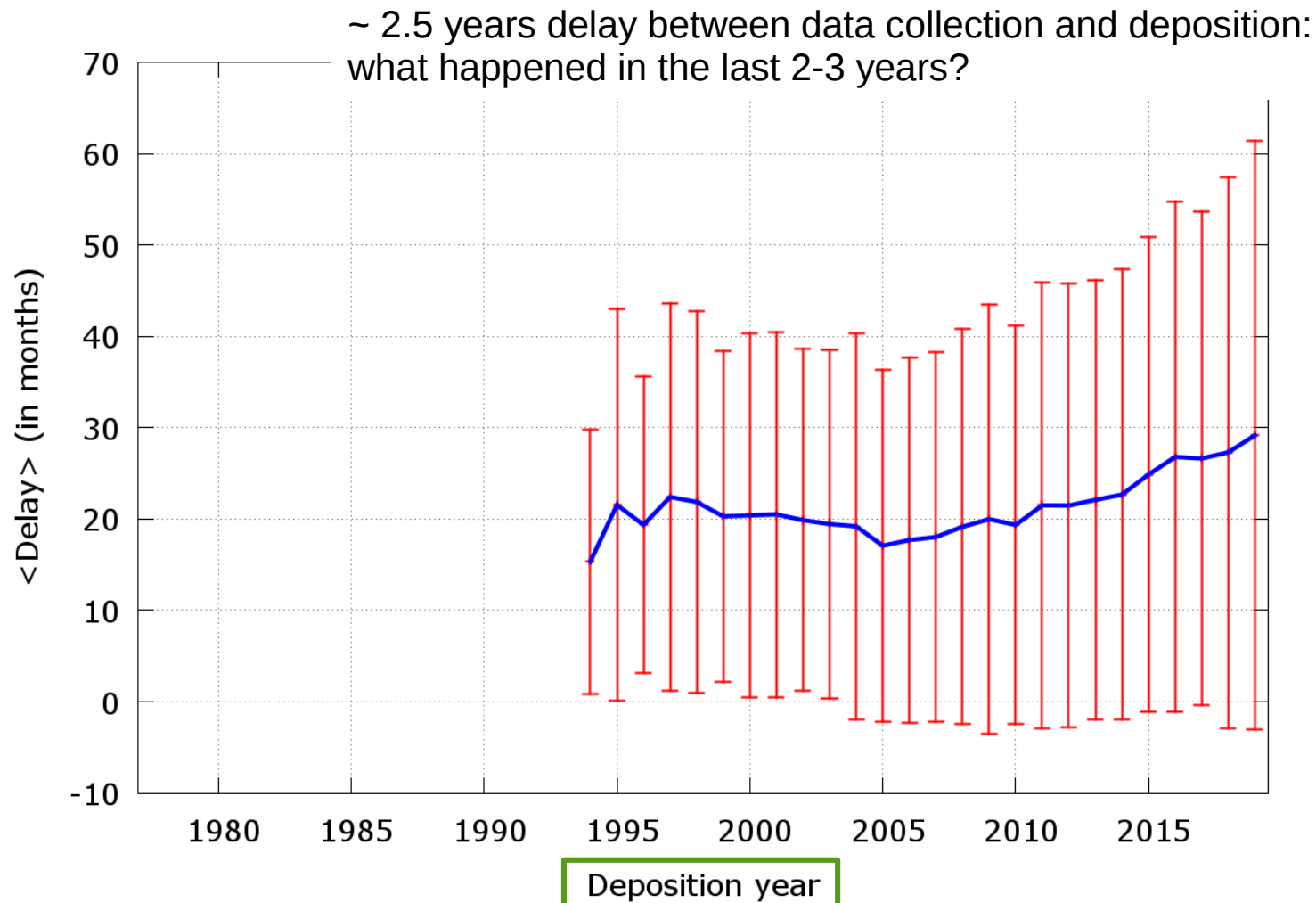
PDB (X-Ray): B-factor



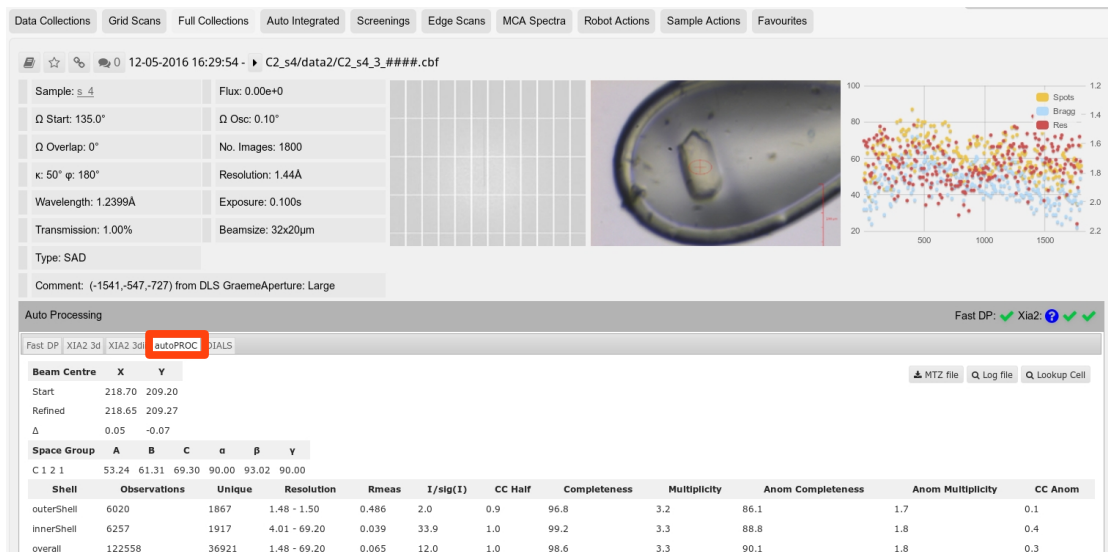
Popular high-resolution limits

1998**2008****2018****2019**

PDB: data collection → data deposition



Automatic processing at synchrotrons



Diamond
(DLS),
UK

OSC 17-05-2018 16:30:00
/data/visitor/mx415/td30a2/20180517/RAW_DATA/t4

Summary Bear

Pipeline	SpaceGroup	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	Shell	Resolution (Å)	Multiplicity	Completeness %	<I/Sigma>	Rmeas	Rmerge	Rpim	cc(1/2)	ccAno	sigAno	ISA	Download
BEST autoPROC	1 2 3	79.0	79.0	79.0	90.0	90.0	90.0	Overall Inner Outer	55.9-2.3 55.9-6.3 2.4-2.3	5.2 5.3 5.4	99.1 94.7 100.0	16.0 36.3 2.8	8.3 2.6 47.7		100 100 90					
autoPROC_staranisotropy	1 2 3	79.0	79.0	79.0	90.0	90.0	90.0	Overall Inner Outer	55.9-1.8 55.9-5.1 1.9-1.8	5.1 5.0 5.3	92.0 97.0 49.0	9.6 34.7 -0.2	10.5 3.1 -607.4		100 100 40					
ANOM autoPROC	1 2 3	79.0	79.0	79.0	90.0	90.0	90.0	Overall Inner Outer	55.9-2.3 55.9-6.3 2.4-2.3	5.2 5.3 5.5	98.8 94.7 99.5	16.1 36.7 2.7	9.1 2.6 51.8		100 100 90					
ANOM autoPROC_staranisotropy	1 2 3	79.0	79.0	79.0	90.0	90.0	90.0	Overall Inner Outer	55.9-1.8 55.9-5.2 1.9-1.8	5.1 5.0 5.4	91.1 97.0 48.9	10.0 35.0 -0.6	11.6 3.2 -168.9		100 100 60					

ESRF, F

automatic processing with autoPROC also at: ALBA (ES), APS (US), MaxIV (S), PETRA (D), SLS (CH), SSRF, ...

So where are we now?

- ❑ We (obviously) want **high quality structures** in the PDB: all users of this database benefit from this!
 - ❑ If we assume that more and more X-ray diffraction projects will have data **processed automatically** through automated, high-throughput processing pipeline systems (at synchrotrons):
 - those systems have to be able to provide a more complete picture about data quality than a single number (*the* high resolution value) can provide
 - ❑ We (software developers, synchrotrons, databases) need to **provide the user with the information to make decisions** about data quality, comparison of datasets, comparison of processing results etc.
 - synchrotrons become more powerful, crystal handling more automatic, detectors faster: can **collect on much more samples** (and not all have been pre-screened for quality)
 - computing becomes more powerful, allowing **automatic processing of every dataset collected** with a whole range of different programs and options
-

The STARANISO Server

*Anisotropy of the Diffraction Limit
and
Bayesian Estimation of Structure Amplitudes*



Illicium verum

main STARANISO server: staraniso.globalphasing.org

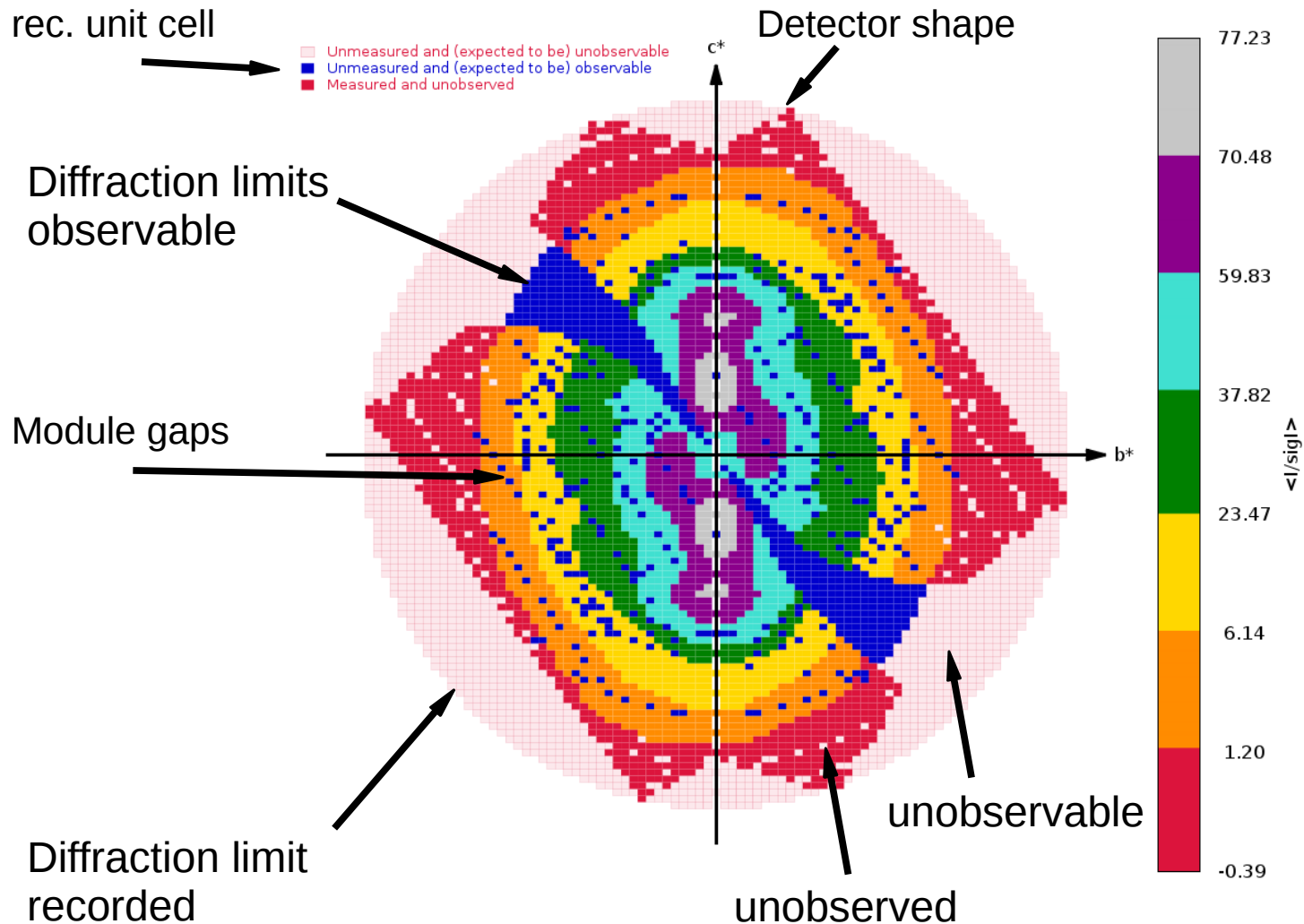
analyse deposited PDB datasets: staraniso.globalphasing.org/cgi-bin/PDBpeep.cgi

Remember: anisotropy means “not isotropic” (ellipsoid is approximation to simplify description of anisotropy).

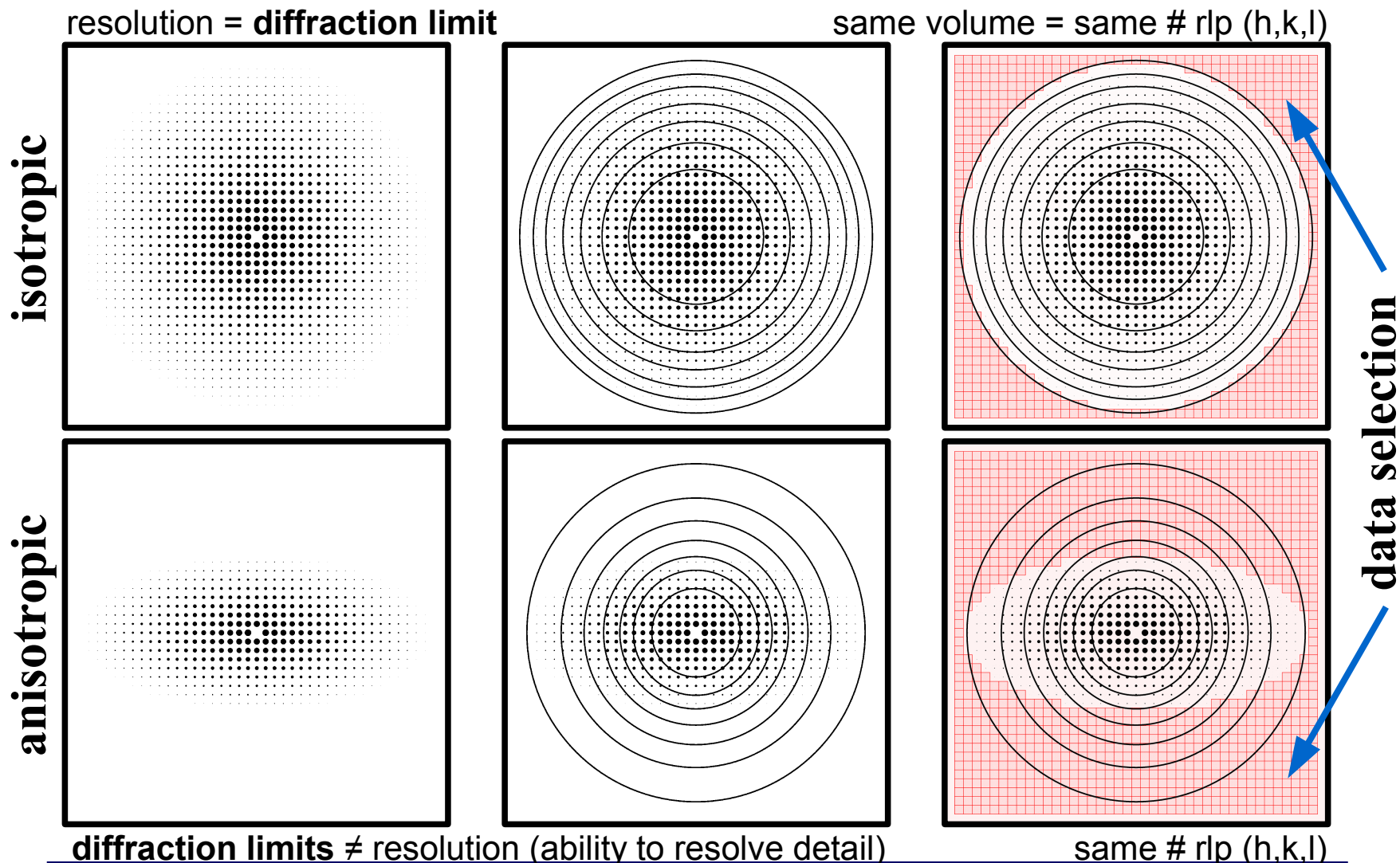
Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonnrhein, C., Bricogne, G. (2016). STARANISO. Cambridge, United Kingdom: Global Phasing Ltd.

Rupp, Bernhard. "Against Method: Table 1 - Cui Bono?." Structure (2018).

autoPROC: STARANISO 2D plots



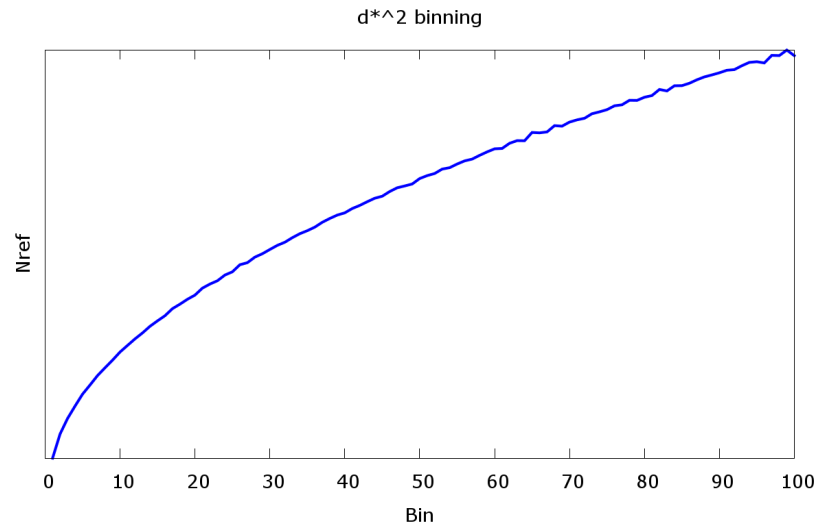
Resolution, diffraction limit and equal-observation-number binning



Revisiting binning (resolution shells)

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE ≥ -3.0

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENE OF DATA
3.13	75353	14441	14448	100.0%
2.22	136754	26193	26370	99.3%
1.81	178487	34010	34139	99.6%
1.57	207171	40372	40378	100.0%
1.40	214727	45702	45725	99.9%
1.28	183266	50533	50658	99.8%
1.18	142566	54470	55100	98.9%
1.11	57895	43932	59148	74.3%
1.04	15808	15171	63013	24.1%
total	1212027	324824	388979	83.5%



□ Equally spaced in $(d^*)^2$

- Number of measurements used for computing bin averages increases with resolution
- Low-resolution statistics can be unreliable (too few measurements)
- Low resolution issues “masked” by too coarse binning
- High-resolution statistics not finely enough sampled?
- Used by XDS, AIMLESS

Binning: decision making, data description

- ❑ **Equal volume** (to have same number of reciprocal lattice points in each bin)
 - Allows different sampling (coarse or fine) depending on requirements
 - Adequate for isotropic data that is complete with homogeneous multiplicity
- ❑ **Equal number** of (actual) observations
 - Can be seen as generalisation of idea behind “equal volume” binning
 - Automatically self-adjusting for anisotropic (STARANISO) and incomplete data (serial crystallography, LCP, microED, ...)

autoPROC

STARANISO

Does it matter? We use binning everywhere in crystallographic software for:

- ❑ Decision making:
 - Which images/datasets to include
 - Which reflections to include: isotropic/anisotropic diffraction limit
- ❑ Comparisons:
 - Between programs and pipelines
 - Between different processing options
- ❑ Basis of decision making
 - Raw data (binned statistics)
 - Smoothed (spline, Bezier, ...)
 - Ice-ring resolution ranges included/excluded/smoothed?

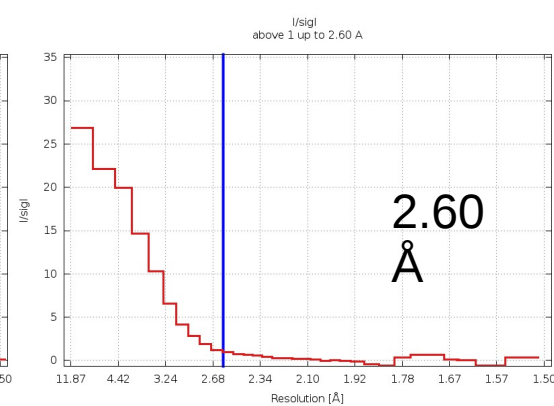
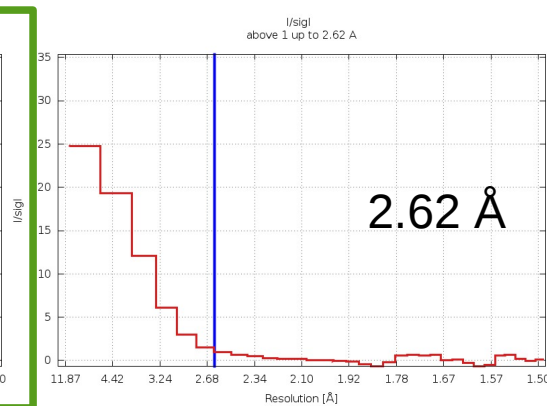
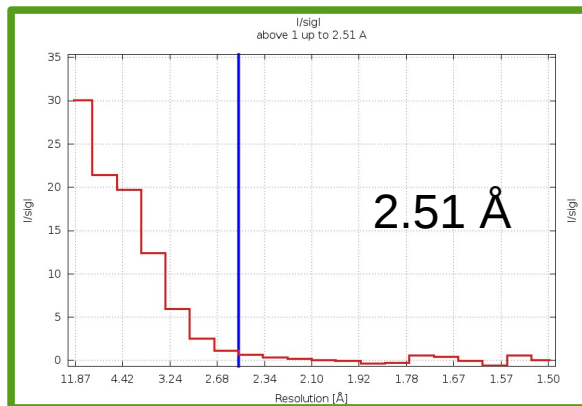
What is my high-resolution limit? ... depending on binning method

d*2 (20 bins)

equal-volume

equal-Nobs

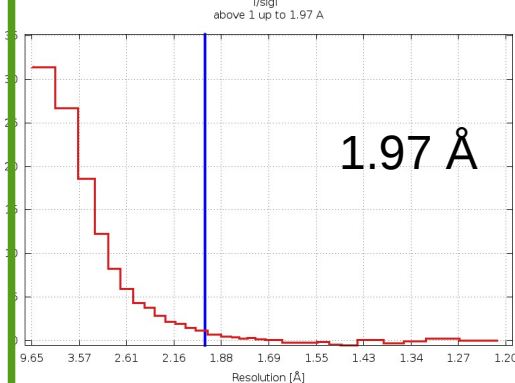
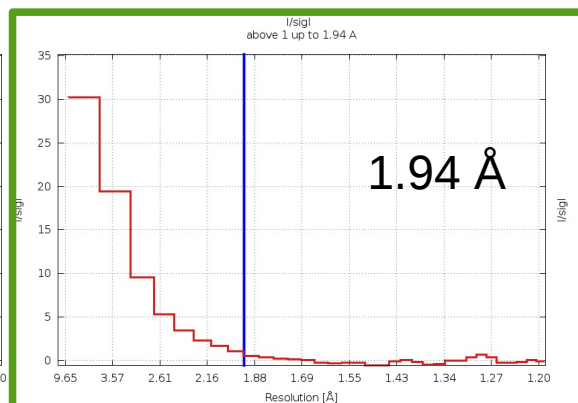
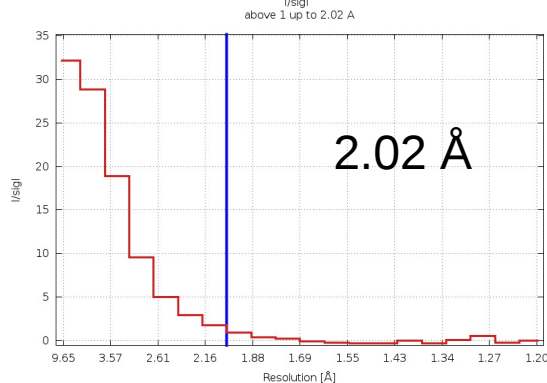
data A



I/sigI > 1

looking from the low-resolution end

data B



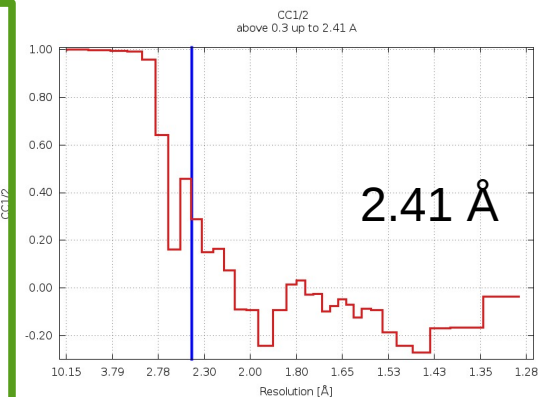
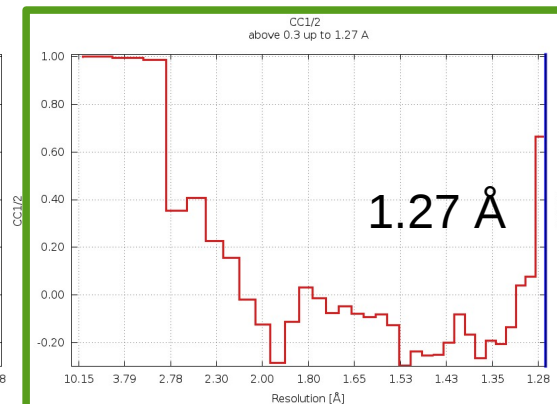
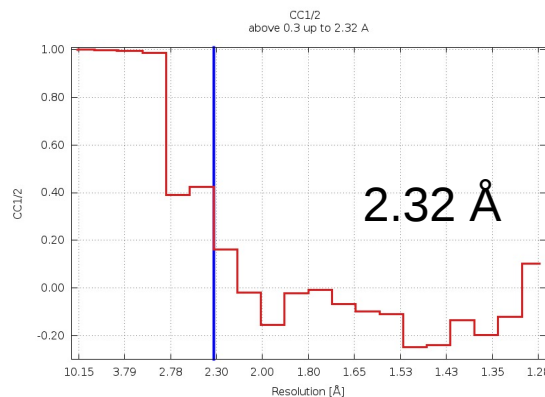
How do we look at data/statistics: from low- or high-resolution end?

d^*2 (20 bins)

equal-volume

equal-Nobs

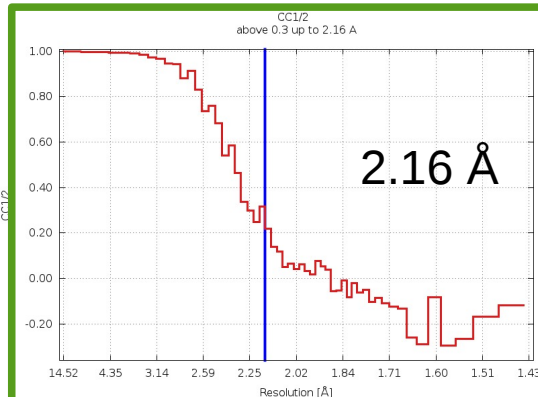
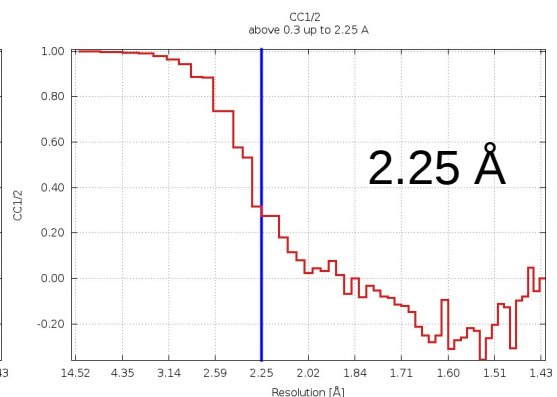
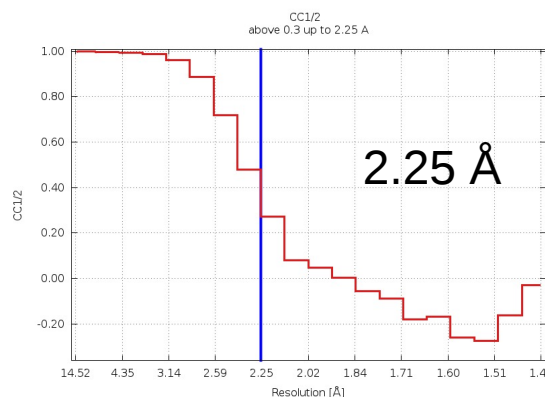
data C



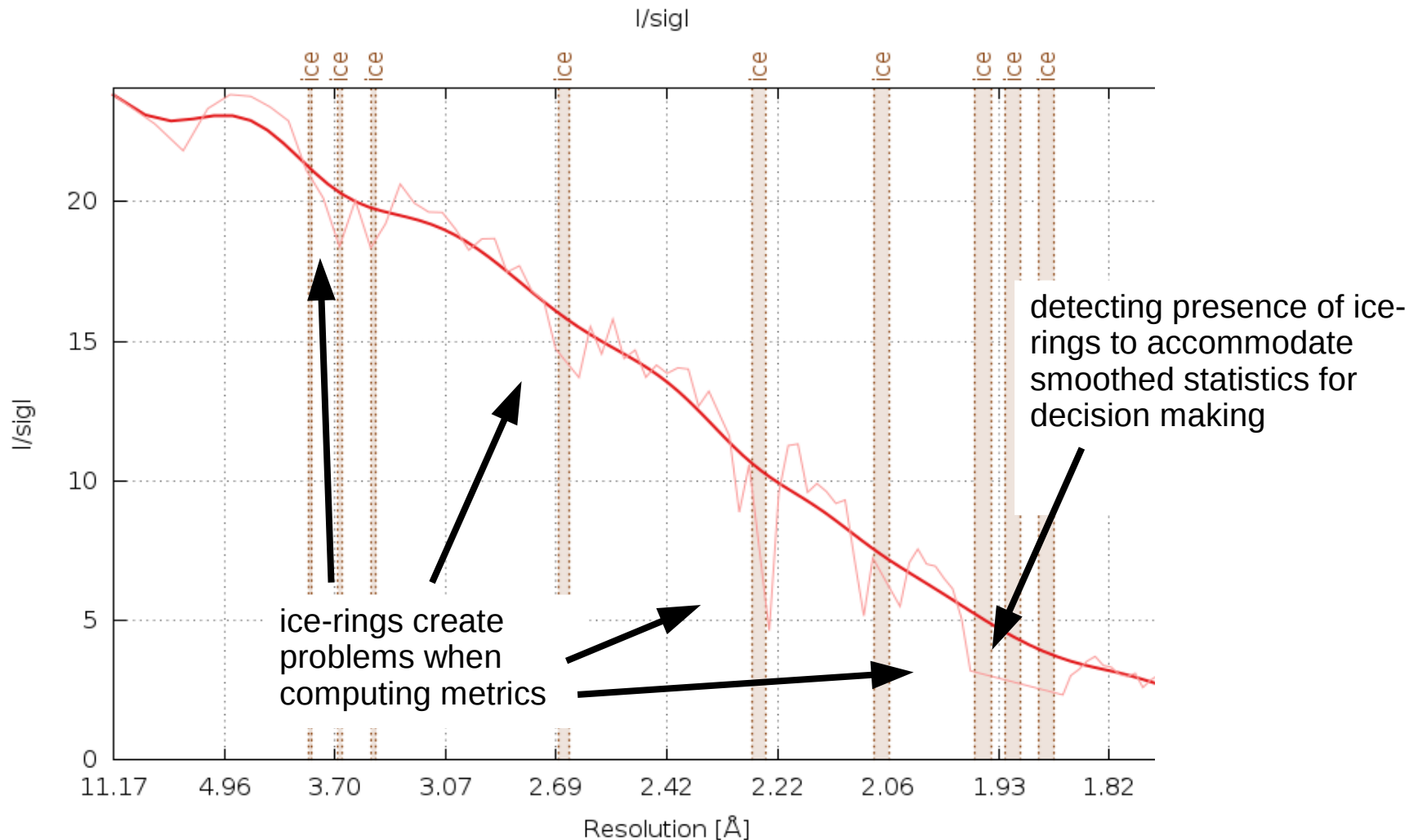
CC(1/2)>0.3

looking from the high-resolution end

data D



Raw vs. smoothed data as basis



Same dataset: which result is better?

RESOLUTION LIMIT	COMPLETENESS OF DATA	R-FACTOR observed	I/SIGMA	R-meas	CC (1/2)
2.83	99.3%	4.2%	29.69	4.7%	99.4*
2.00	99.7%	3.9%	27.49	4.3%	99.8*
1.63	99.9%	5.1%	22.68	5.7%	99.8*
1.42	100.0%	8.4%	15.39	9.4%	99.3*
1.27	100.0%	16.2%	8.92	18.3%	97.8*
1.16	99.9%	26.5%	5.65	29.8%	94.3*
1.07	99.7%	48.1%	3.06	54.3%	83.7*
1.00	87.7%	93.1%	0.97	115.1%	39.5*
0.94	38.3%	175.8%	0.31	236.5%	11.5*
total	88.1%	5.2%	10.18	5.8%	99.6*
2.83	98.7%	3.7%	22.81	4.5%	99.4*
2.00	98.5%	3.3%	20.65	4.2%	98.7*
1.63	99.3%	4.4%	16.46	5.5%	99.2*
1.42	99.3%	7.5%	10.77	9.4%	98.8*
1.27	99.1%	14.4%	6.04	18.3%	77.0*
1.16	99.3%	23.6%	3.76	30.0%	88.8*
1.07	98.9%	43.0%	2.02	54.8%	72.9*
1.00	72.7%	83.3%	0.72	113.3%	32.7*
0.94	24.9%	161.3%	0.26	228.1%	12.9*
total	83.1%	4.5%	7.61	5.6%	99.6*

Same dataset: which result is better?

RESOLUTION LIMIT	COMPLETENESS OF DATA	R-FACTOR observed	I/SIGMA	R-meas	CC (1/2)
2.83	99.3%	4.2%	29.69	4.7%	99.4*
2.00	99.7%	3.9%	27.49	4.3%	99.8*
1.63	99.9%	5.1%	22.68	5.7%	99.8*
1.42	100.0%	8.4%	15.39	9.4%	99.3*
1.27	100.0%	16.2%	8.92	18.3%	97.8*
1.16	99.9%	26.5%	5.65	29.8%	94.3*
1.07	99.7%	48.1%	3.06	54.3%	83.7*
1.00	87.7%	93.1%	0.97	115.1%	39.5*
0.94	38.3%	175.8%	0.31	236.5%	11.5*
total	88.1%	5.2%	10.18	5.8%	99.6*

2.83	98.7%	3.7%	22.81	4.5%	99.4*
2.00	98.5%	3.3%	20.65	4.2%	98.7*
1.63	99.3%	4.4%	16.46	5.5%	99.2*
1.42	99.3%	7.5%	10.77	9.4%	98.8*
1.27	99.1%	14.4%	6.04	18.3%	77.0*
1.16	99.3%	23.6%	3.76	30.0%	88.8*
1.07	98.9%	43.0%	2.02	54.8%	72.9*
1.00	72.7%	83.3%	0.72	113.3%	32.7*
0.94	24.9%	161.3%	0.26	228.1%	12.9*
total	83.1%	4.5%	7.61	5.6%	99.6*

XDS: FRIEDEL'S_LAW= TRUE | FALSE

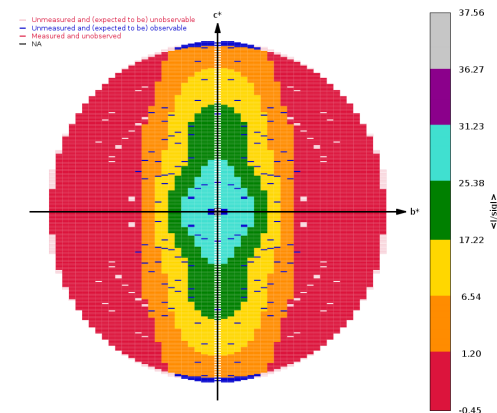
RESOLUTION LIMIT	COMPLETENESS OF DATA	R-FACTOR observed	I/SIGMA	R-meas	CC (1/2)	
2.83	99.3%	4.2%	29.69	4.7%	99.4*	“noano” TRUE
2.00	99.7%	3.9%	27.49	4.3%	99.8*	
1.63	99.9%	5.1%	22.68	5.7%	99.8*	
1.42	100.0%	8.4%	15.39	9.4%	99.3*	
1.27	100.0%	16.2%	8.92	18.3%	97.8*	
1.16	99.9%	26.5%	5.65	29.8%	94.3*	
1.07	99.7%	48.1%	3.06	54.3%	83.7*	
1.00	87.7%	93.1%	0.97	115.1%	39.5*	
0.94	38.3%	175.8%	0.31	236.5%	11.5*	
total	88.1%	5.2%	10.18	5.8%	99.6*	
2.83	98.7%	3.7%	22.81	4.5%	99.4*	“ano” FALSE
2.00	98.5%	3.3%	20.65	4.2%	98.7*	
1.63	99.3%	4.4%	16.46	5.5%	99.2*	
1.42	99.3%	7.5%	10.77	9.4%	98.8*	
1.27	99.1%	14.4%	6.04	18.3%	77.0*	
1.16	99.3%	23.6%	3.76	30.0%	88.8*	
1.07	98.9%	43.0%	2.02	54.8%	72.9*	
1.00	72.7%	83.3%	0.72	113.3%	32.7*	
0.94	24.9%	161.3%	0.26	228.1%	12.9*	
total	83.1%	4.5%	7.61	5.6%	99.6*	

Full "Table 1" from autoPROC (not CORRECT.LP/XSCALE.LP/aimless.log)

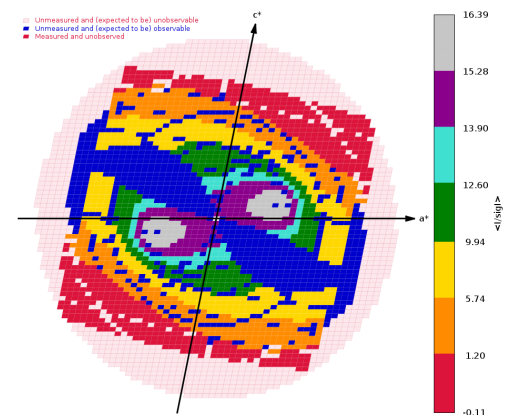
	Overall	InnerShell	OuterShell
Low resolution limit	97.979	97.979	0.970
High resolution limit	0.944	3.843	0.944
Rmerge (all I+ & I-)	0.052	0.049	2.832
Rmerge (within I+/I-)	0.045	0.041	2.574
Rmeas (all I+ & I-)	0.058	0.055	3.924
Rmeas (within I+/I-)	0.056	0.050	3.641
Rpim (all I+ & I-)	0.026	0.023	2.707
Rpim (within I+/I-)	0.033	0.029	2.574
Total number of observations	728801	15183	3500
Total number unique	169786	2979	2979
Mean(I)/sd(I)	10.2	29.7	0.2
Completeness	88.1	98.7	19.8
Multiplicity	4.3	5.1	1.2
CC (1/2)	0.997	0.991	-0.005
Anomalous completeness	77.8	95.8	3.0
Anomalous multiplicity	2.4	2.7	1.1
CC (ano)	-0.053	-0.063	NA
DANO /sd(DANO)	0.822	0.765	0.829

DFc completion for density maps

- ❑ How to handle missing data when computing maps?
- ❑ Difference map (**mFo-DFc**):
 - nothing we can do, i.e. have to treat as 0
- ❑ **2mFo-DFc** map:
 - using DFc instead of 0 should be better
 - if we had measured this observable reflection it would have $F > 0$
 - “DFc completion”
- ❑ Refinement programs (BUSTER, PHENIX, REFMAC) allow control
 - **BUT:** differences in default 2mFo-DFc map coefficients written by programs!



highly anisotropic



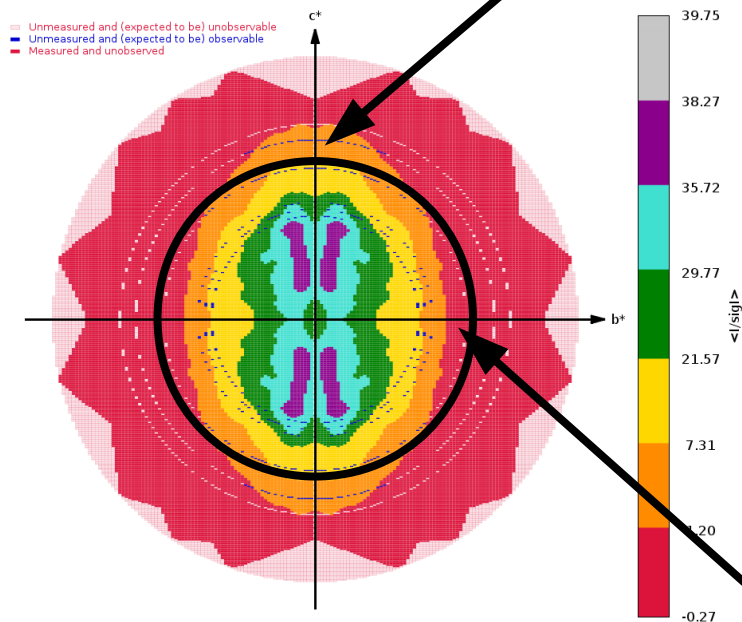
cusp + ice

DFc completion: traditional data analysis

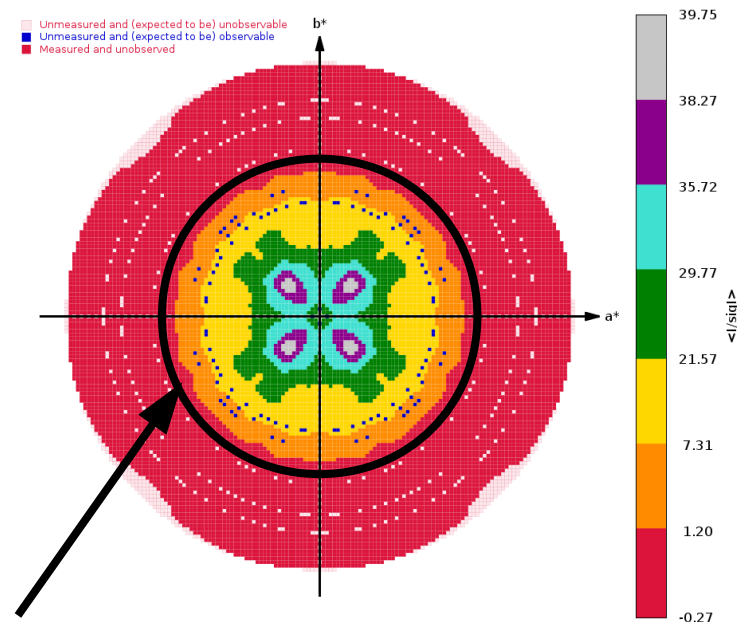
Assuming data is isotropic:

→ 2.8Å high resolution limit

0kl plane



hk0 plane



Diffraction limits & principal axes of ellipsoid fitted to diffraction cut-off surface:

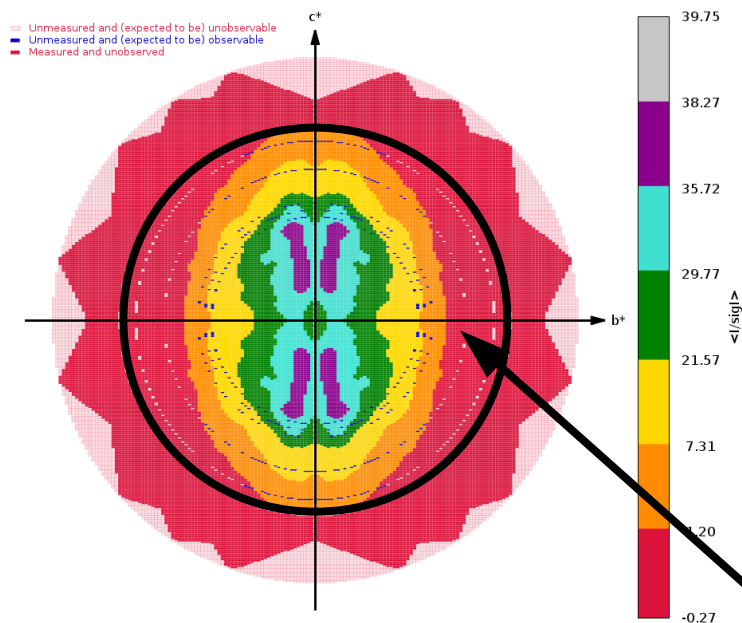
3.032	1.0000	0.0000	0.0000	<u>a</u> *
3.032	0.0000	1.0000	0.0000	<u>b</u> *
2.077	0.0000	0.0000	1.0000	<u>c</u> *

STARANISO analysis

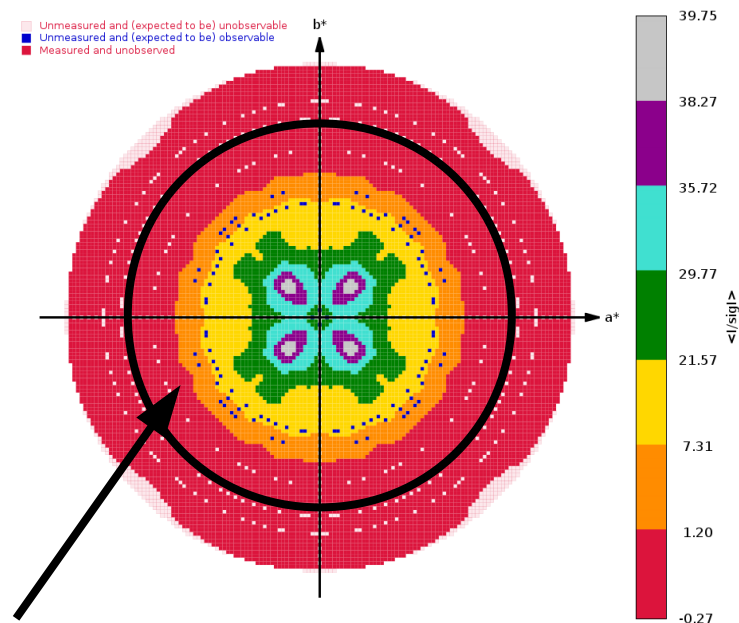
DFc completion: anisotropic data 1/3

REFMAC	FWT	PHWT
PHENIX	2FOFCWT_fill	PH2FOFCWT_fill
BUSTER	2FOFCWT_iso-fill	PH2FOFCWT_iso-fill

0kl plane



hk0 plane



include noise

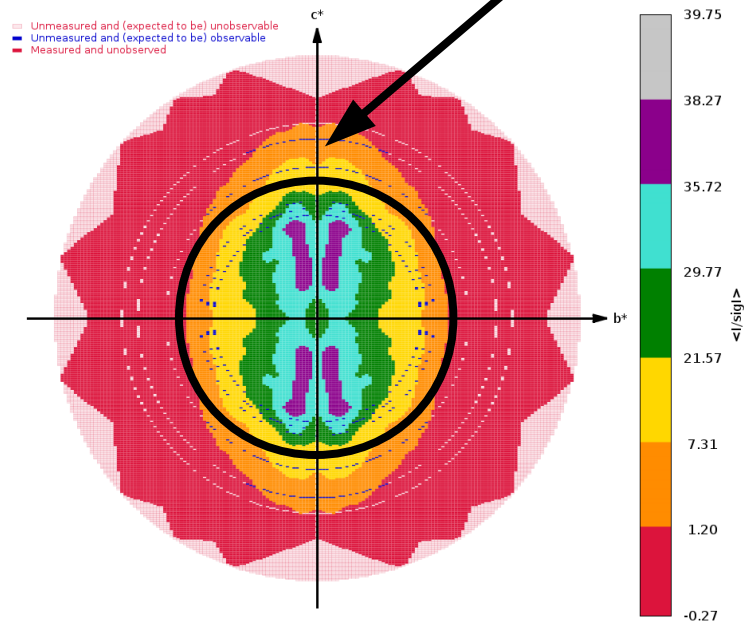
**anisotropic analysis of data (STARANISO)
using high resolution limit**

DFc completion: anisotropic data 2/3

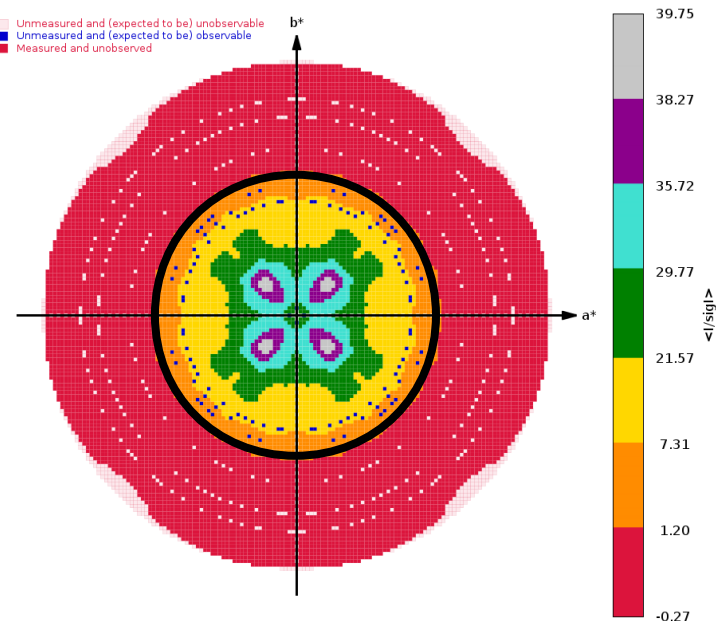
Diffraction limits & principal axes of ellipsoid fitted to diffraction cut-off surface:

3.032	1.0000	0.0000	0.0000	<u>a</u> *
3.032	0.0000	1.0000	0.0000	<u>b</u> *
2.077	0.0000	0.0000	1.0000	<u>c</u> *

0kl plane **missed signal**



hk0 plane



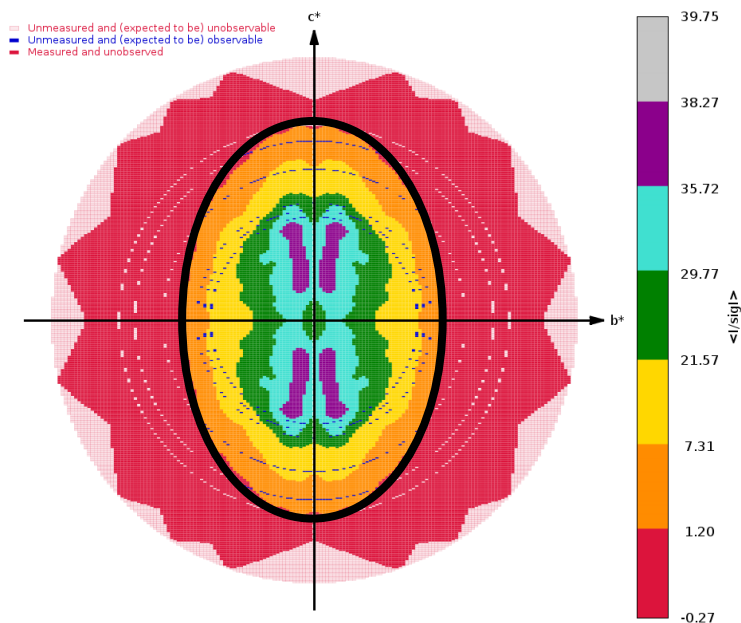
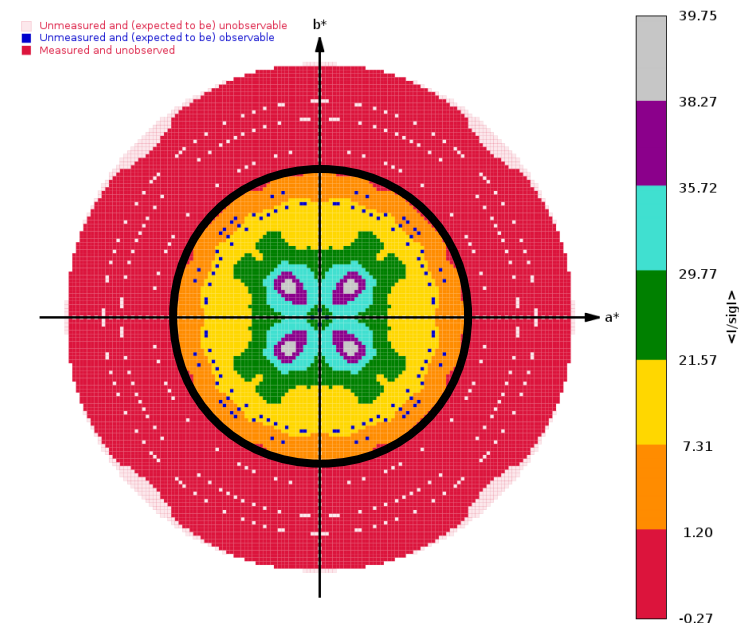
**anisotropic analysis of data (STARANISO)
using lowest diffraction limit**

DFc completion: anisotropic data 3/3

BUSTER

2FOFCWT_aniso-fill

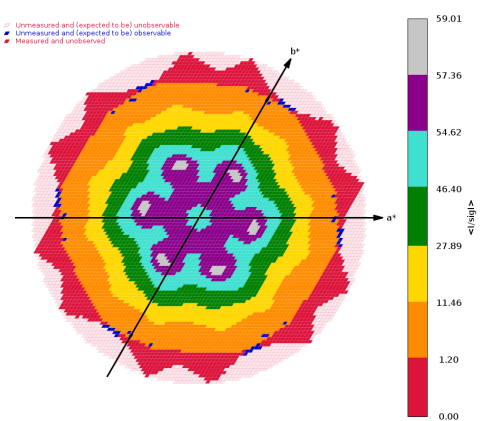
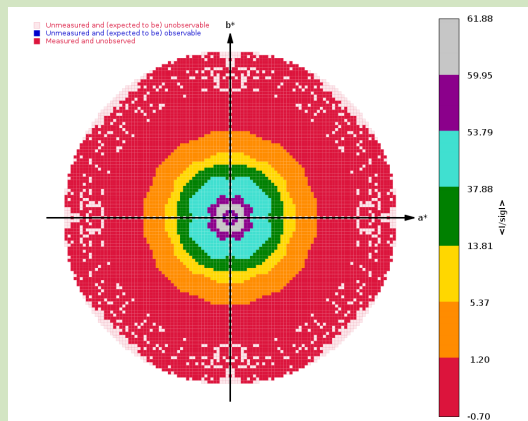
PH2FOFCWT_aniso-fill

0kl plane**hk0 plane**

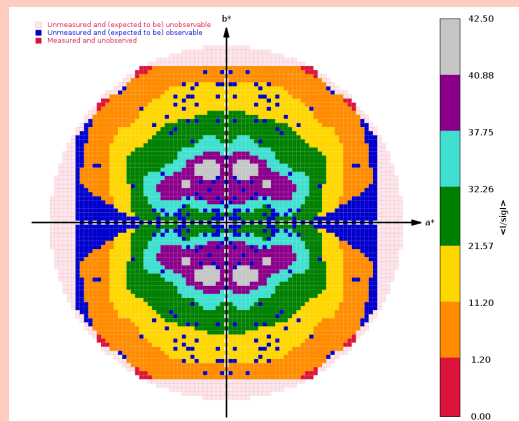
**anisotropic analysis of data (STARANISO, SA_flag)
using anisotropic diffraction limits**

Diffraction limit determination required to accurately describe collected data

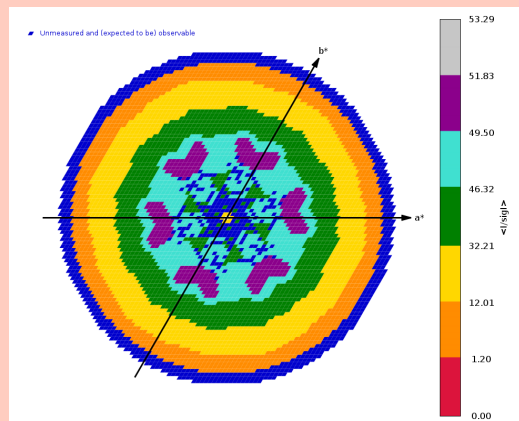
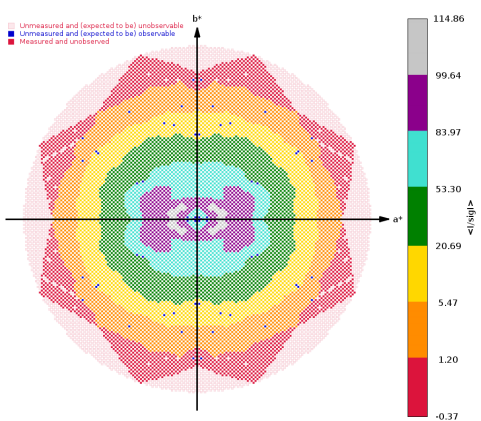
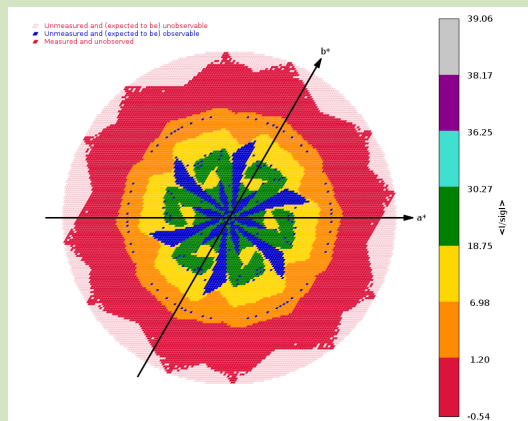
“Too close”



Too far



Crystal diffracts better than resolution of collected data



- Since we deal with X-Ray diffraction, let's call it “**diffraction limit(s)**”
 - Will the $\langle B \rangle$ of deposited structures keep rising?
 - **Data anisotropy** requires new ways of looking at data (and describing it)
 - **Binning methods** are not exciting, but important.
 - **Automated processing and decision** making depends on all of the above
 - **DFc completion** needs to be done correctly - taking observability into account
 - **Data quality metrics** need revisiting, consolidation and clarification: watch this space!
-

Data quality metrics workshop - 04/2019

Beamline scientists, synchrotron, software and database developers as well as power users and crystallography experts:

- ❑ to address the need for adequate and consistent calculation and presentation of **data quality metrics** for crystallographic X-Ray experiments.
- ❑ to make results **comparable** for **experts** as well as **non-expert** users.



☐ **Global Phasing Ltd (UK):**

- Gérard Bricogne
- Leigh Carter
- Claus Flensburg
- Rasmus Fogh
- Peter Keller
- Wlodek Paciorek
- Andrew Sharff
- Ian Tickle
- Marcin Wojdyr

☐ **Global Phasing Industrial Consortium members**

- ☐ Wolfgang Kabsch, Kay Diederichs
- ☐ Phil Evans
- ☐ PDBx/mmCIF working group
- ☐ Jose Marquez, Irina Cornaciu (EMBL/Grenoble)
- ☐ JCSG, SBGrid, proteindiffraction.org (raw data archives)
- ☐ ... many, many users!

DECTRIS®



www.globalphasing.com
staraniso.globalphasing.org
grade.globalphasing.org

autoPROC/STARANISO
 BUSTER / Grade / Rhofit / Pipedream
 SHARP/autoSHARP

Reflection data is different from model data



“daisy-chaining” reflection data files seems like a good idea

refinement PDB-1 → model building PDB-2 → PDB-3 → PDB-4 → ...
... but maybe not for reflection data!



Initial reflection data (intensities, amplitudes, test-set flag)
is the one to (normally) use at all stages