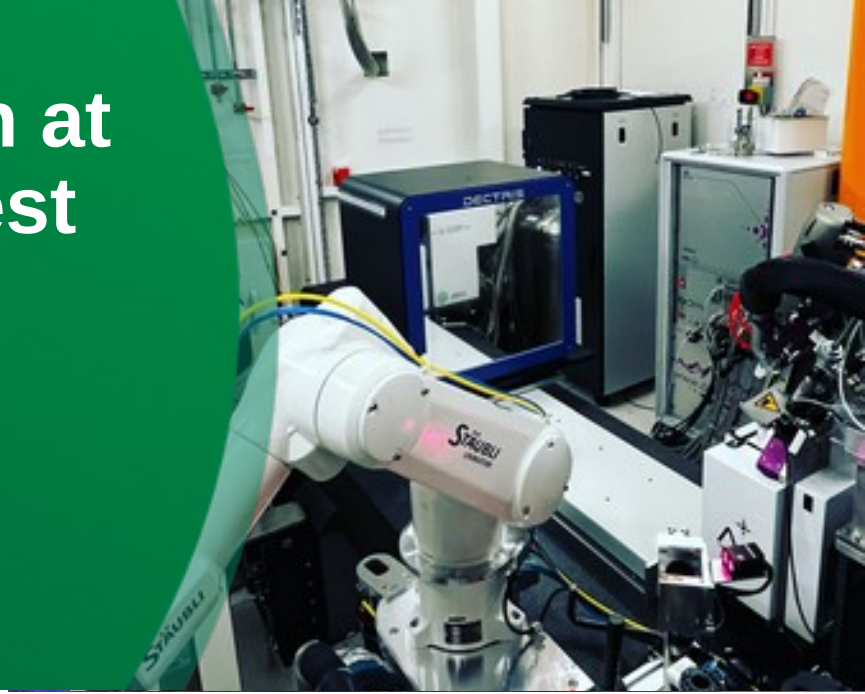
Automated data collection at MASSIF-1 – getting the best data from *all* samples



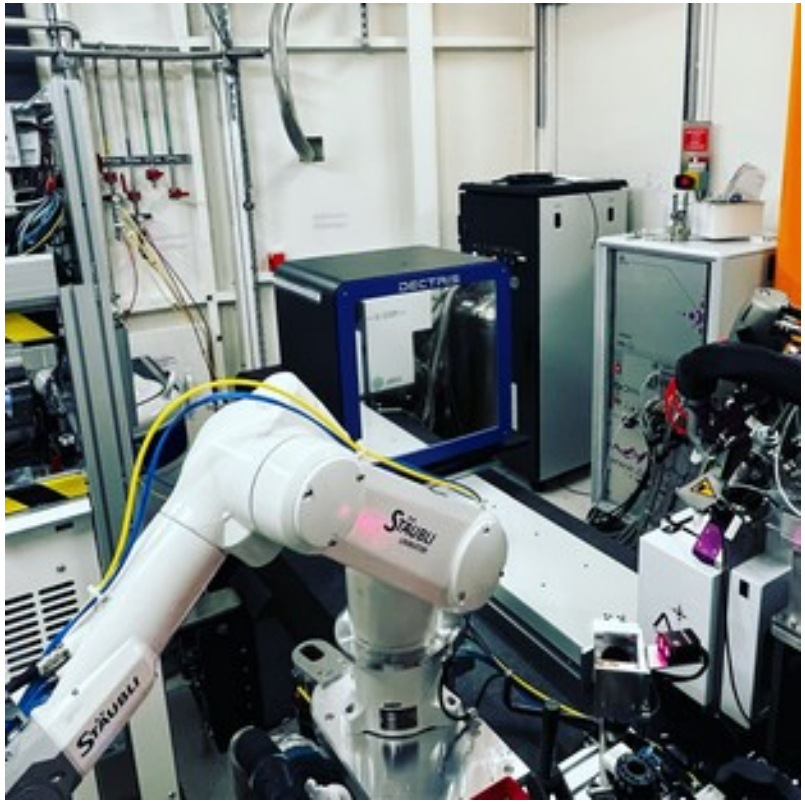
Matthew W. Bowler

Project leader, EMBL Grenoble

MXCuBE meeting, ELETTRA, Trieste, 2024

ESRF EMBL

MASSIF-1

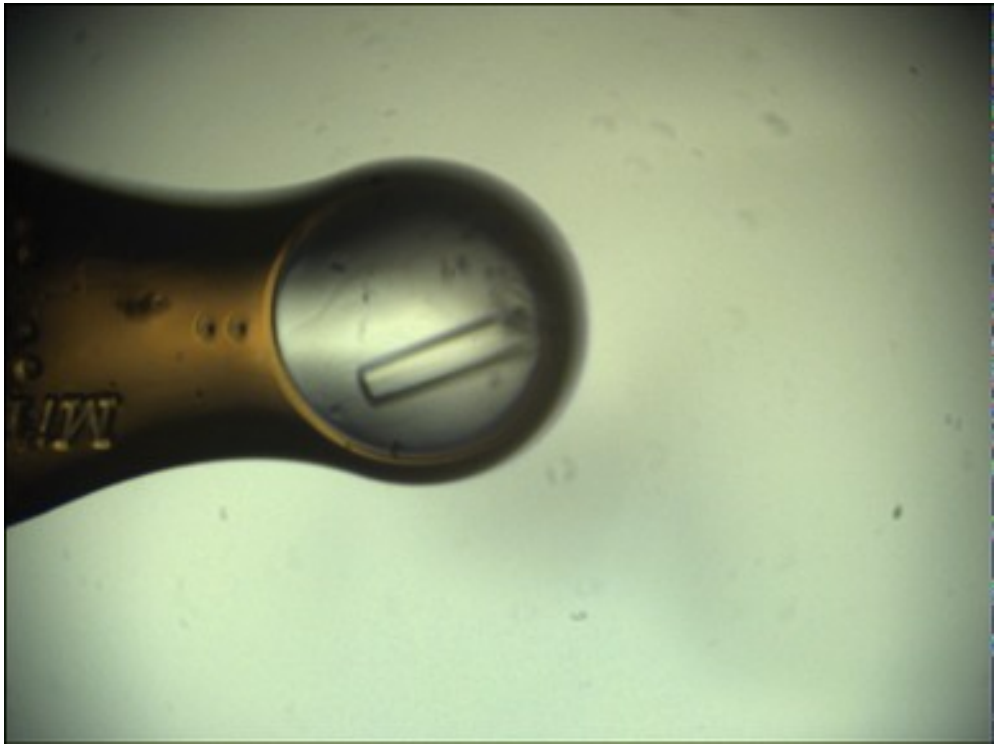


- Fully autonomous beamline
  - no user control
  - data collection optimised for every sample
- Flexible booking, queuing system
- Flex HCD – 368 samples capacity
- **Pilatus3 6M (Pilatus4 4M coming 2025)**
- **CrystalDirect Harvester**
- **Fully automated data collection from any sample either room or cryogenic temperatures with complex strategies and optimized parameters for each sample**

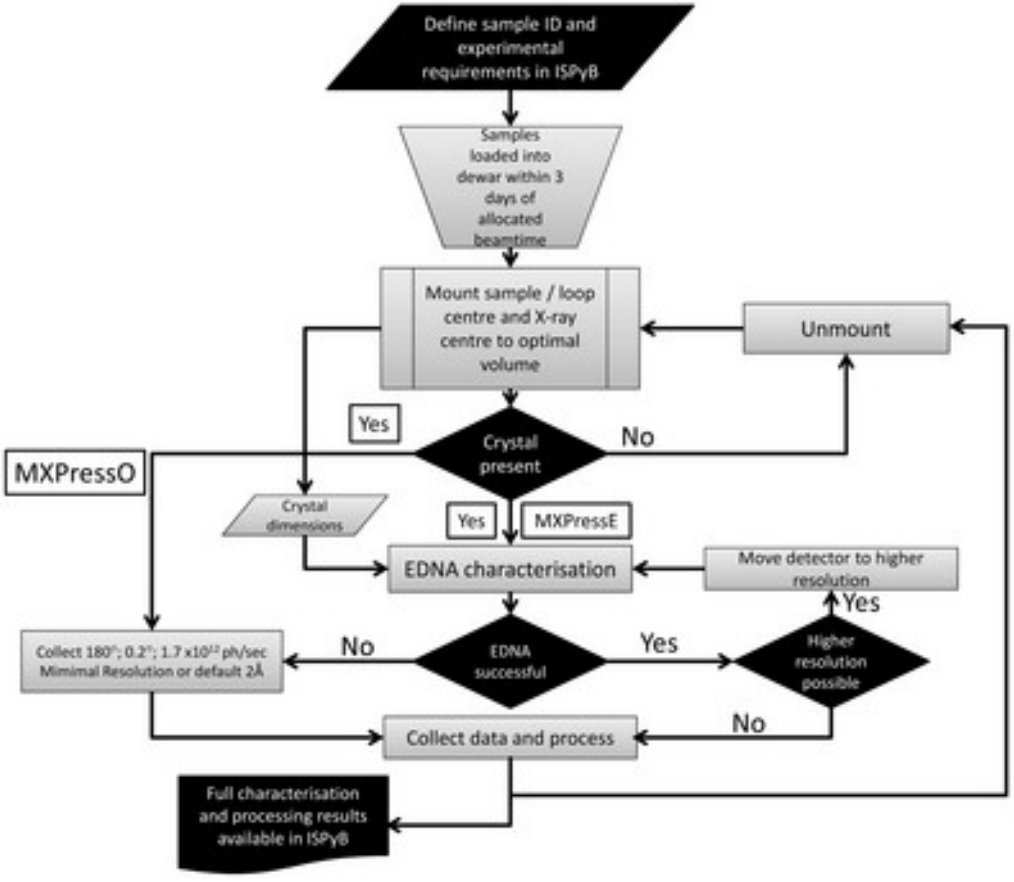
JOURNAL OF SYNCHROTRON RADIATION

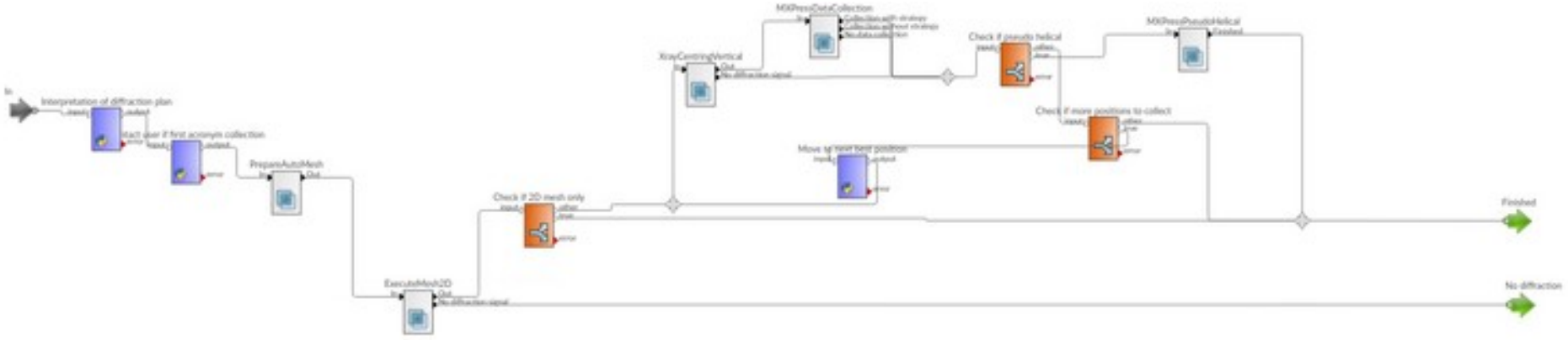
Bowler *et al.*

Volume 22 | **Part 6** | November 2015 | Pages 1540–1547 |



EMBL


Define sample ID and experimental requirements in ISPyB
Samples loaded into dewar within 3 days of allocated beamtime
Mount sample / loop centre and X-ray centre to optimal volume
Unmount
Yes
Crystal present
No
MXPressO
Crystal dimensions
Yes
MXPressE
EDNA characterisation
Move detector to higher resolution
Yes
Collect 180°; 0.2°; 1.7 x10^12 ph/sec Mimimal Resolution or default 2Å
No
EDNA successful
Yes
Higher resolution possible
No
Collect data and process
Full characterisation and processing results available in ISPyB

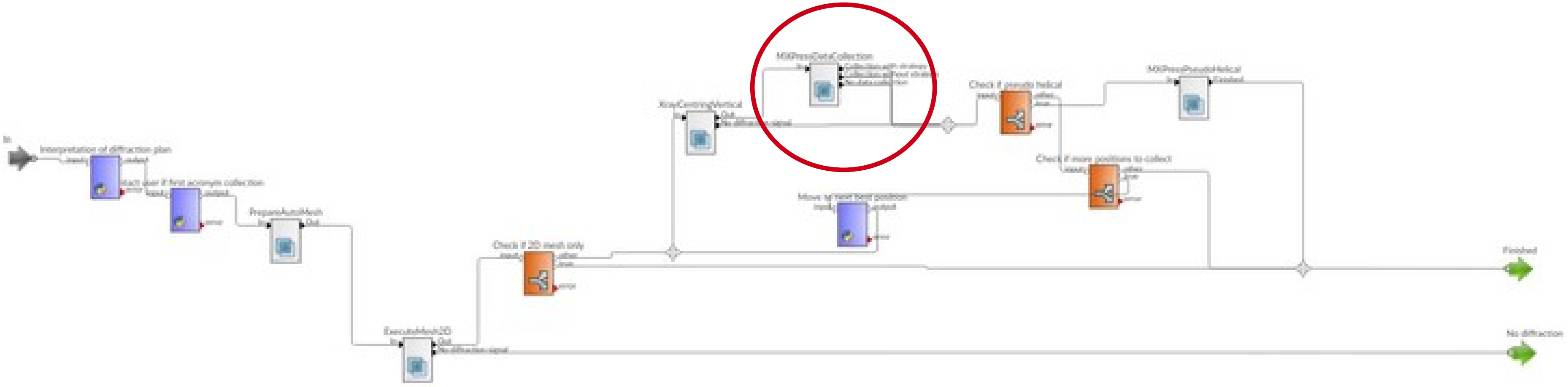


STRUCTURAL BIOLOGY

Svensson *et al.*

Volume 71 | **Part 8** | August 2015 | Pages 1757–1767

EMBL

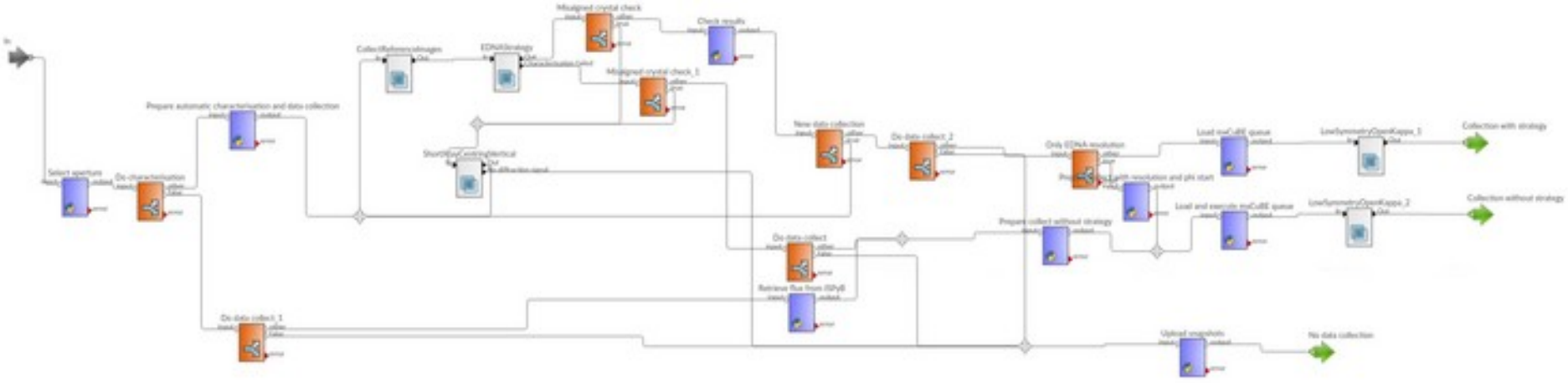


STRUCTURAL BIOLOGY

Svensson *et al.*

Volume 71 | **Part 8** | August 2015 | Pages 1757–1767

EMBL

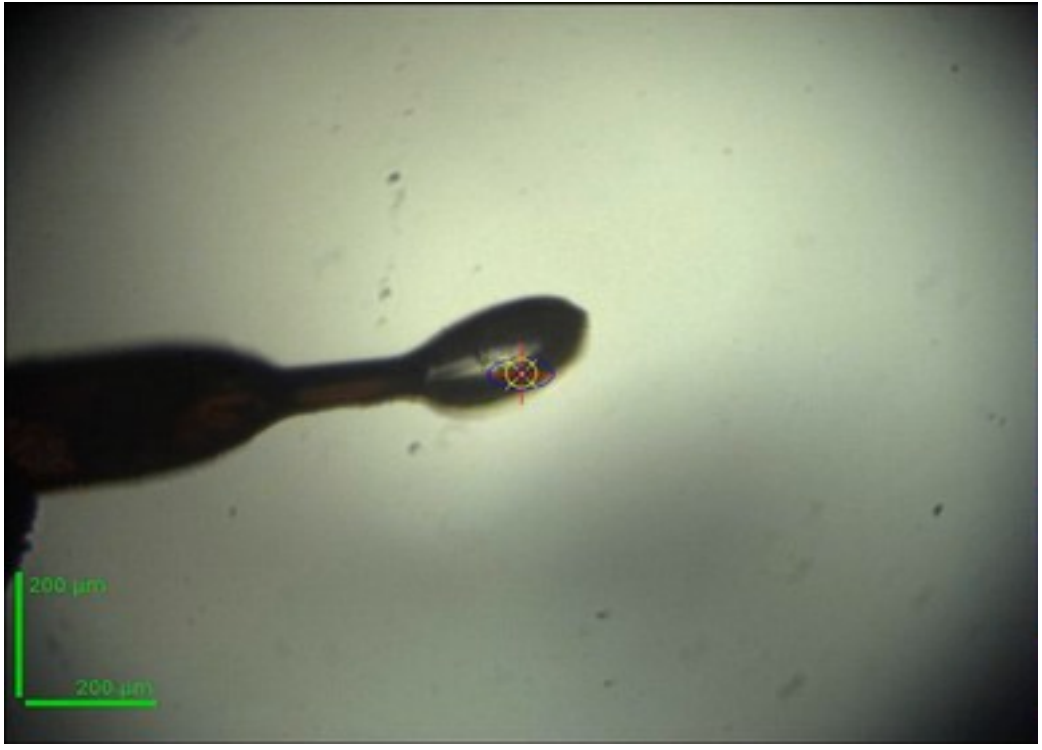


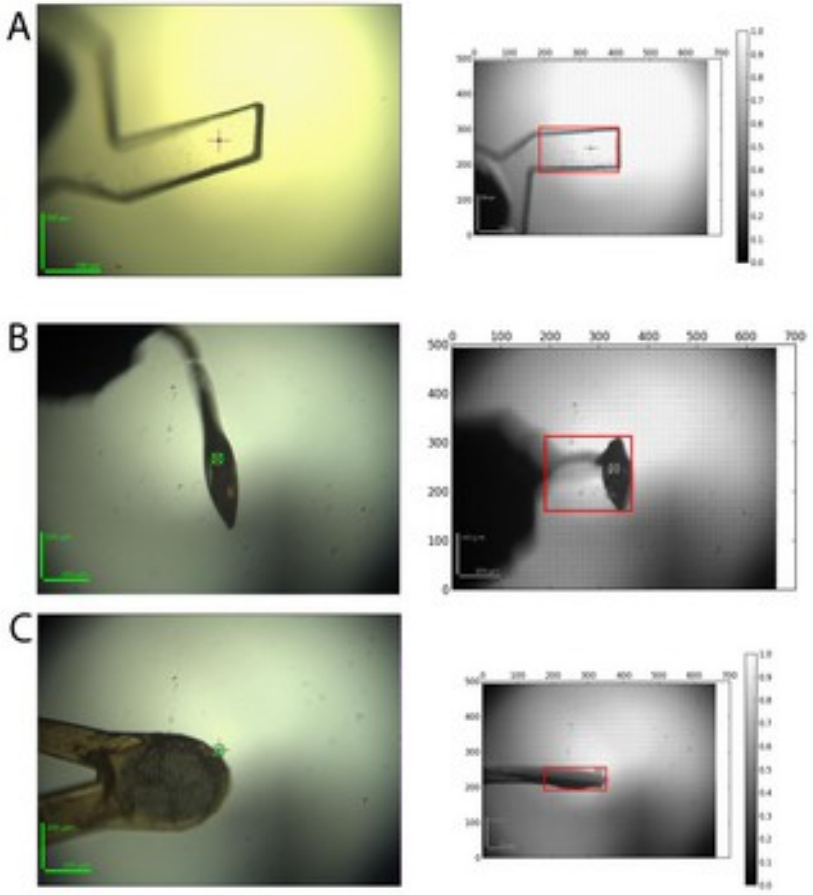
Acta Cryst D STRUCTURAL BIOLOGY

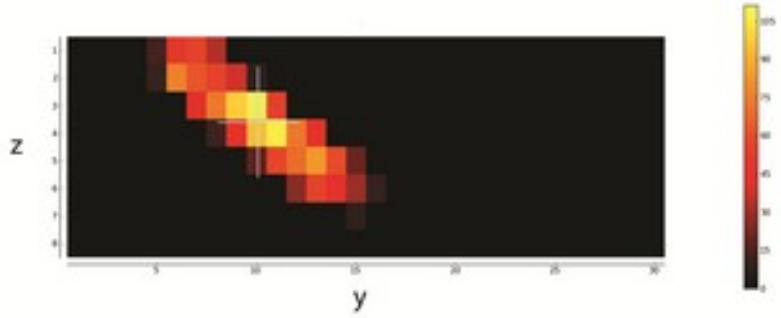
Svensson *et al.*

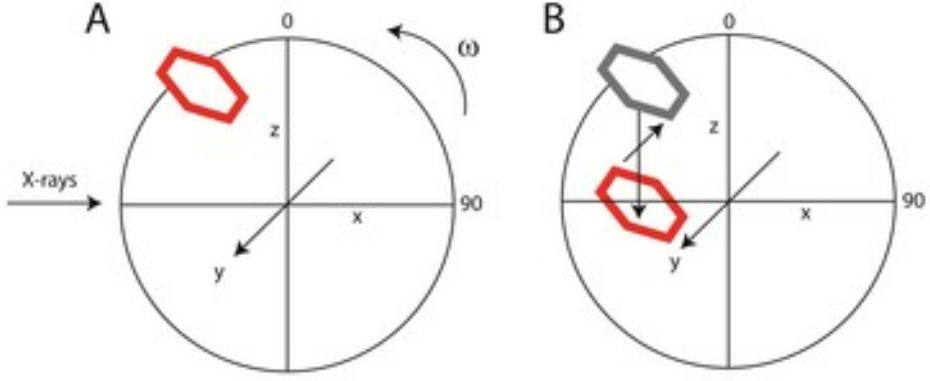
Volume 71 | **Part 8** | August 2015 | Pages 1757–1767

EMBL


200 µm
200 µm


A
B
C





Software routines locate crystals and centre to best volume.

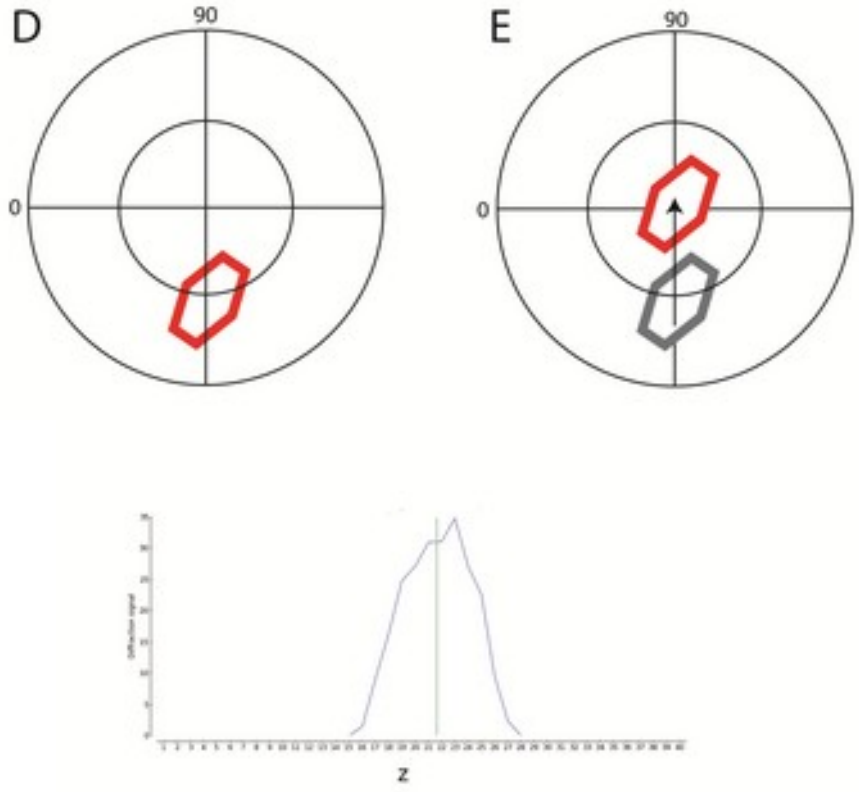
Characteristics such as beam size and flux as well as crystal volume lead to highly optimised data collection

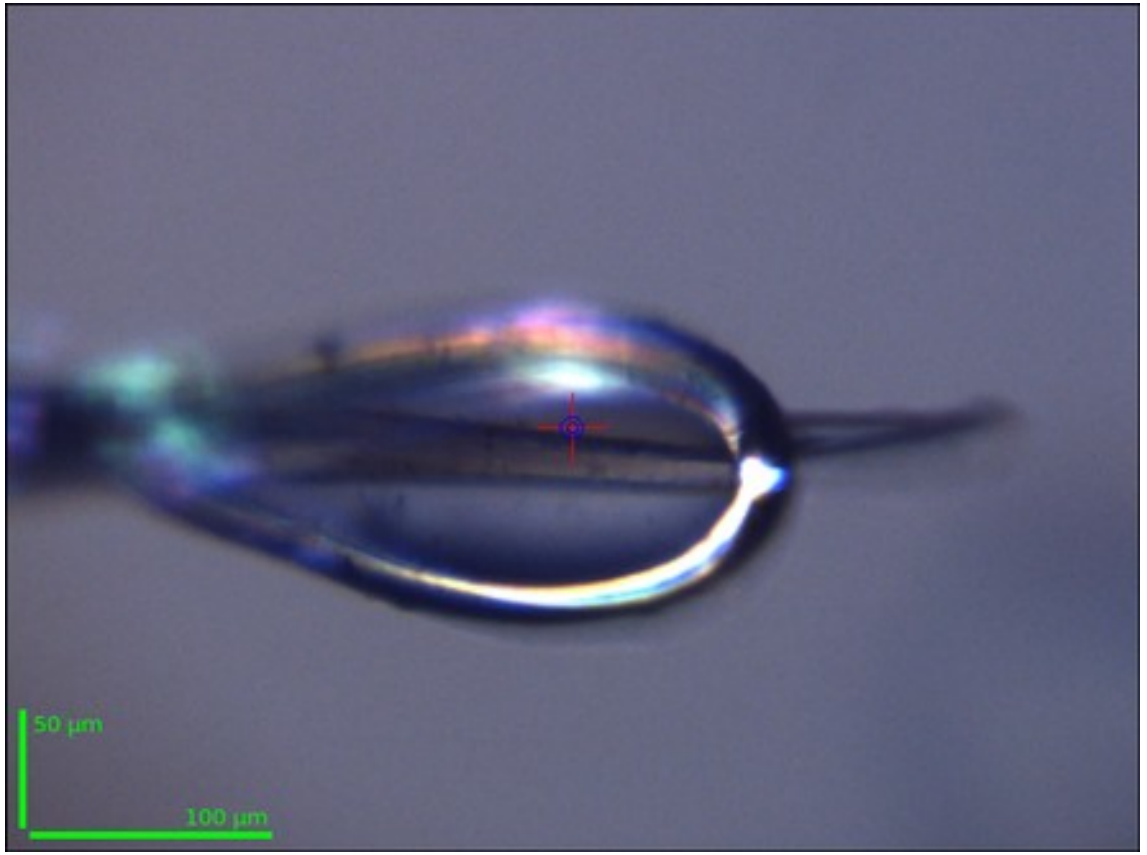
Acta Cryst D STRUCTURAL BIOLOGY

Svensson *et al.*

Volume 71 | **Part 8** | August 2015 | Pages 1757–1767

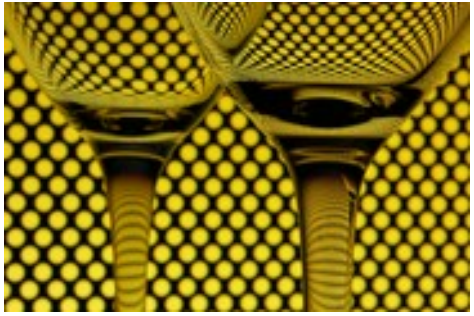
EMBL


D
90
0
E
90
0
z

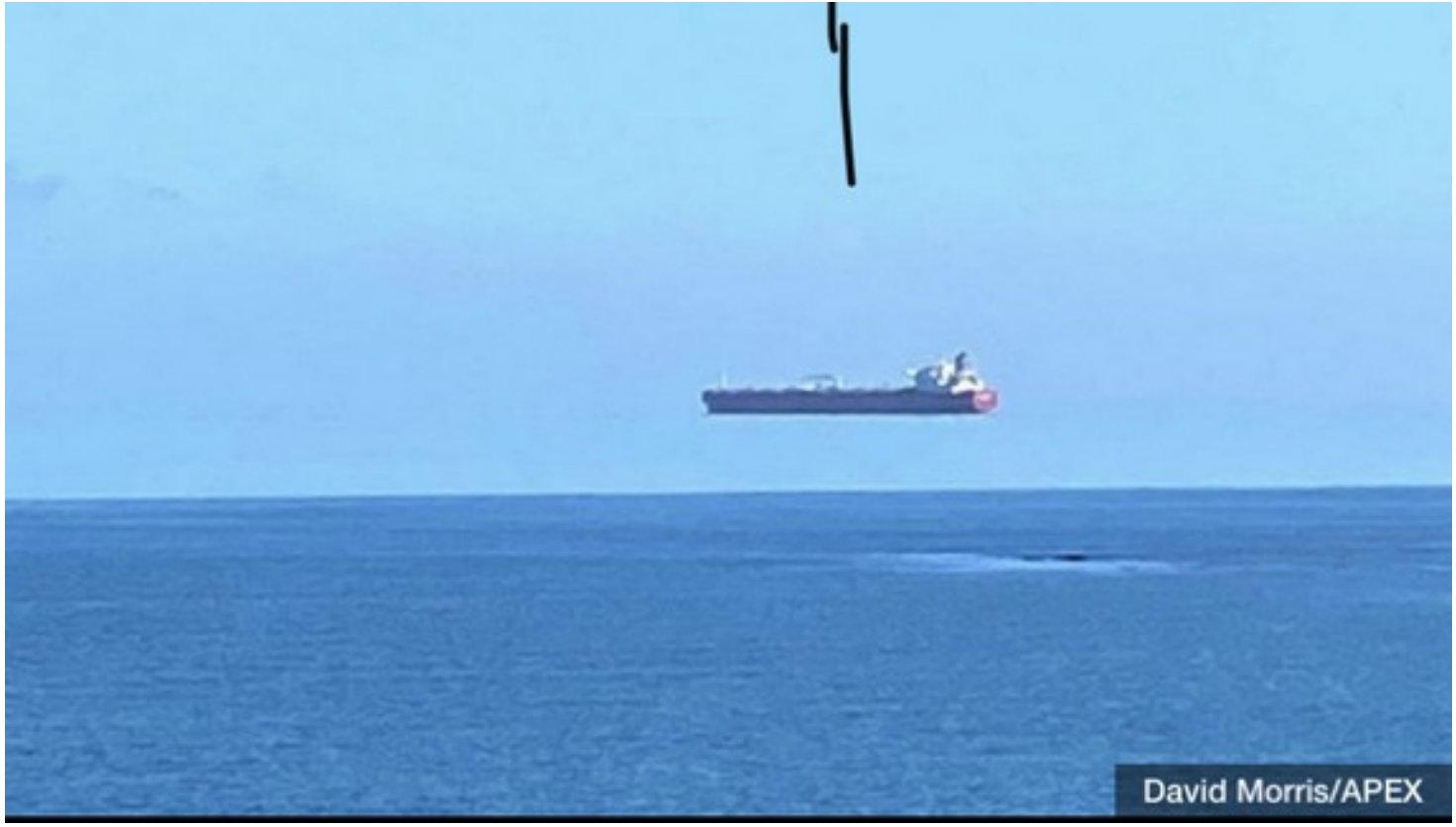


Bowler M.W**.**, Svensson, O and Nurizzo, D. (2016) *Cryst. Rev.*, **22**, 229-245

EMBL

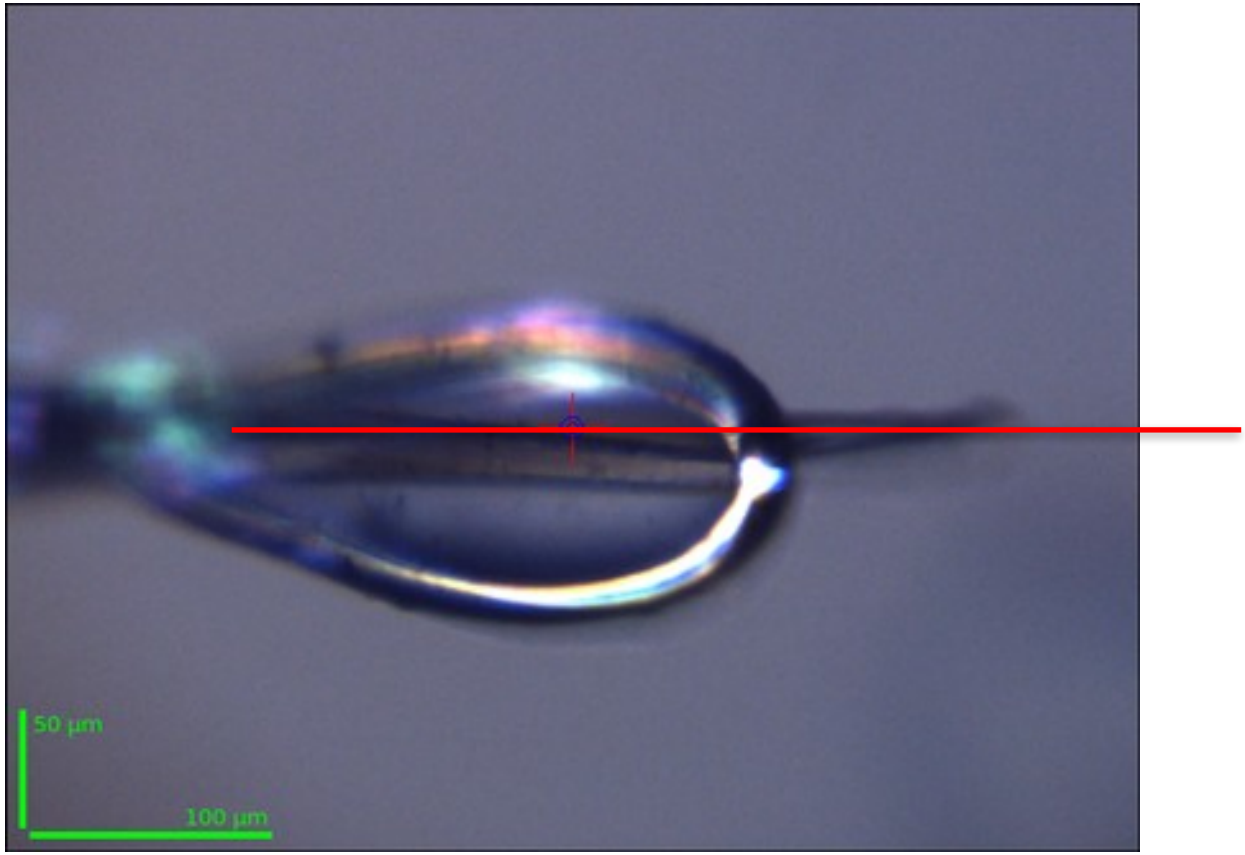


EMBL



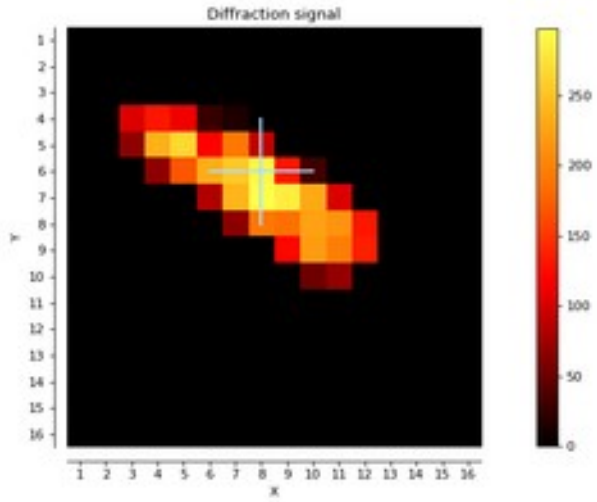
“Fata Morgana”

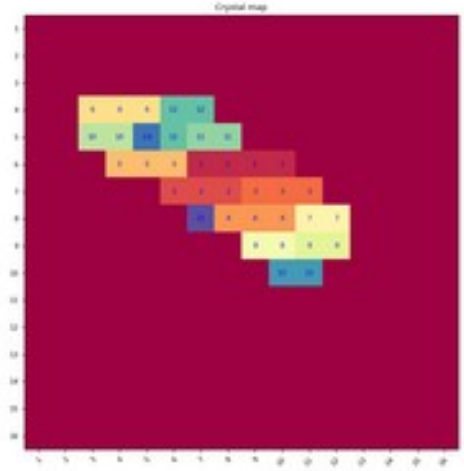
EMBL

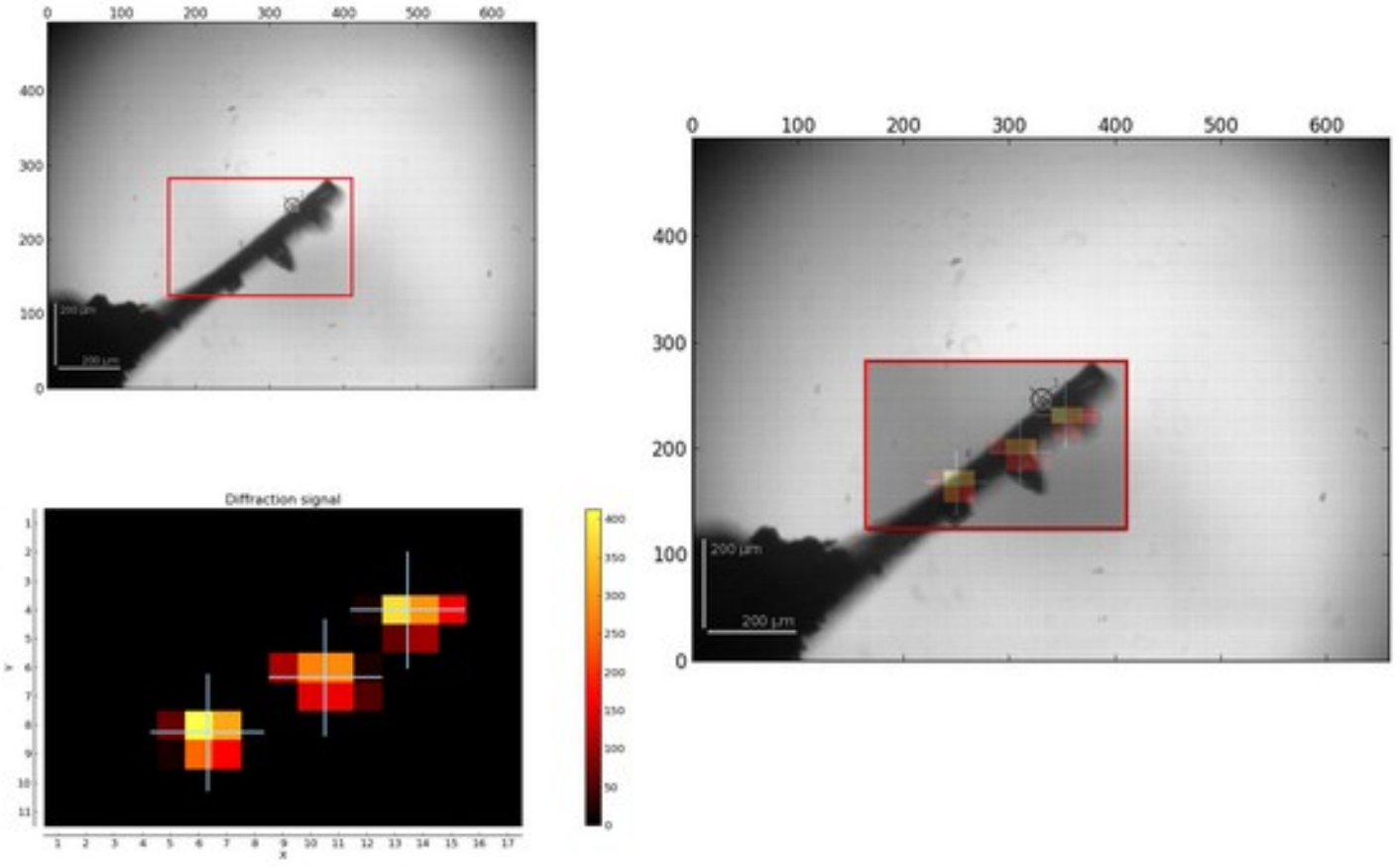


Bowler M.W., Svensson, O and Nurizzo, D. (2016) *Cryst. Rev.*, **22**, 229-245

EMBL


Diffraction signal
Y
X
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
250
200
150
100
50
0





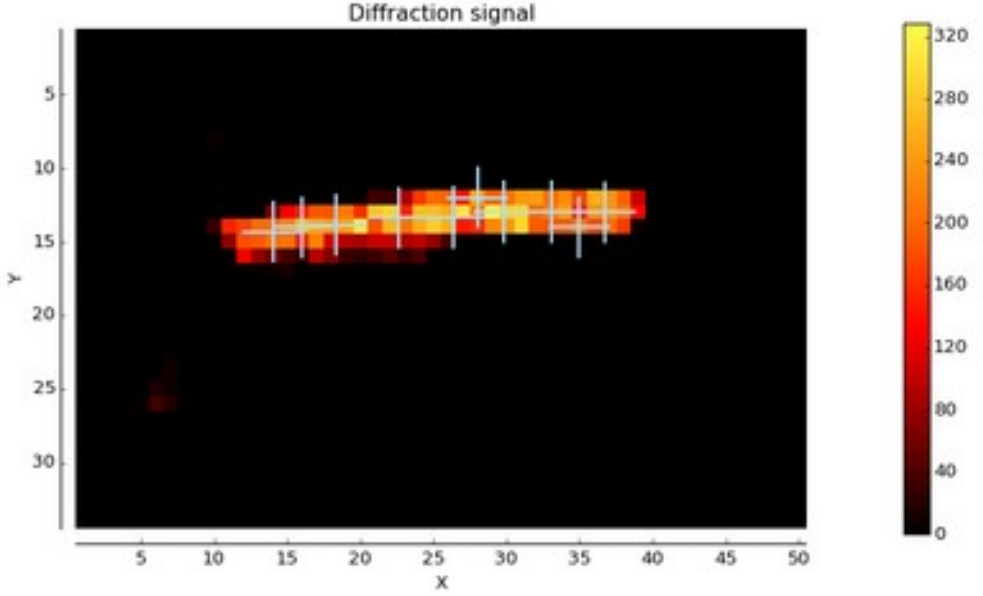
Acta Cryst D STRUCTURAL BIOLOGY

Svensson *et al.*

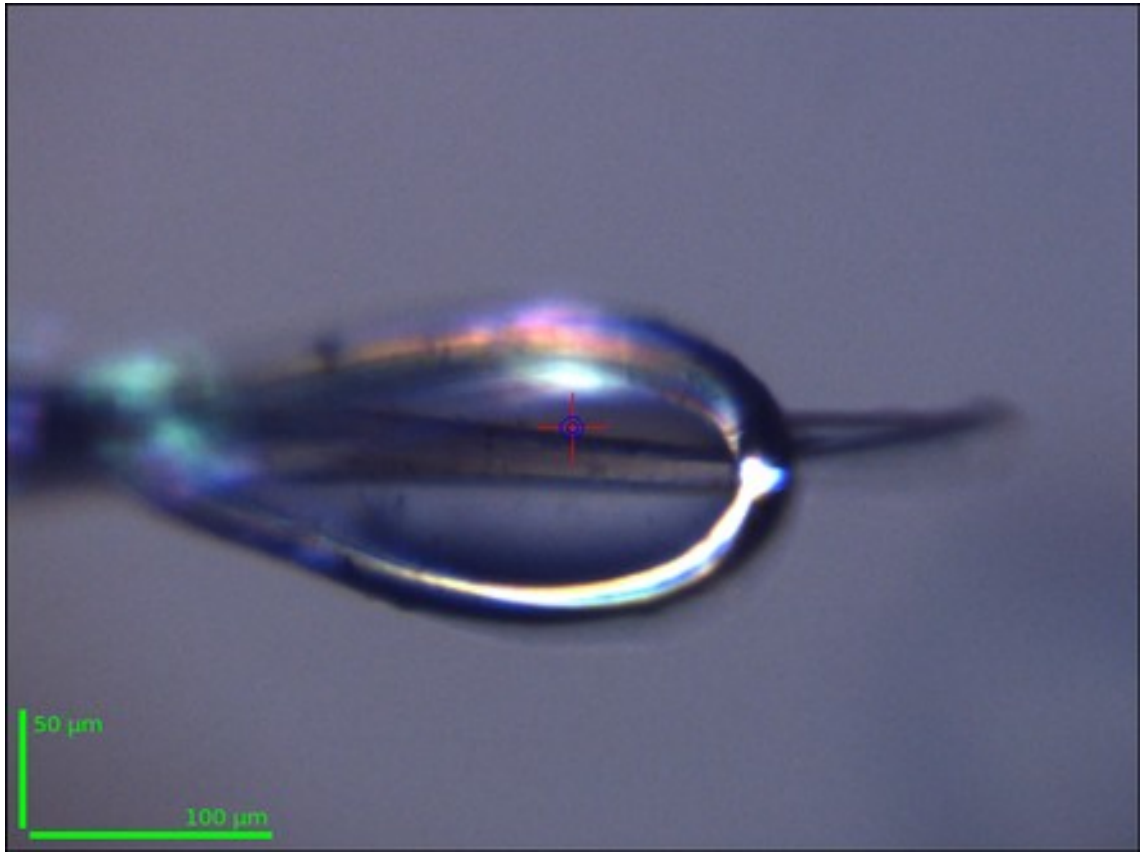
Volume 74 | **Part 5** | May 2018 | Pages 433–440 | 10.1107/S2059798318003728

EMBL

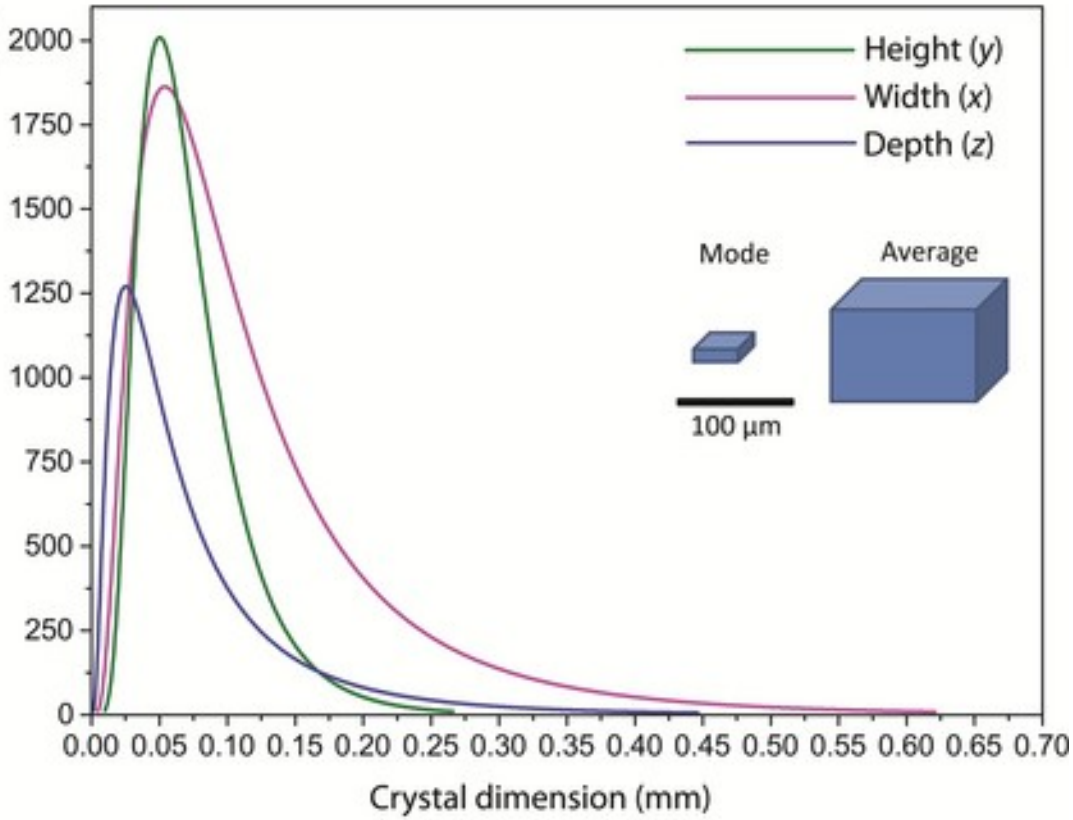
Pseudo-helical data collection – first automated helical workflow that accounts for crystal variability



- Full ‘normal’ data set the subsequent helical
- Auto-peak selection or user defined
- Stringent thresholding – regions within 30%
- SAD option available



EMBL



Bowler M.W**.**, Svensson, O and Nurizzo, D. (2016) *Cryst. Rev.*, **22**, 229-245

EMBL

Molecular-weight dependence of the minimum required crystal size

464 µm edge

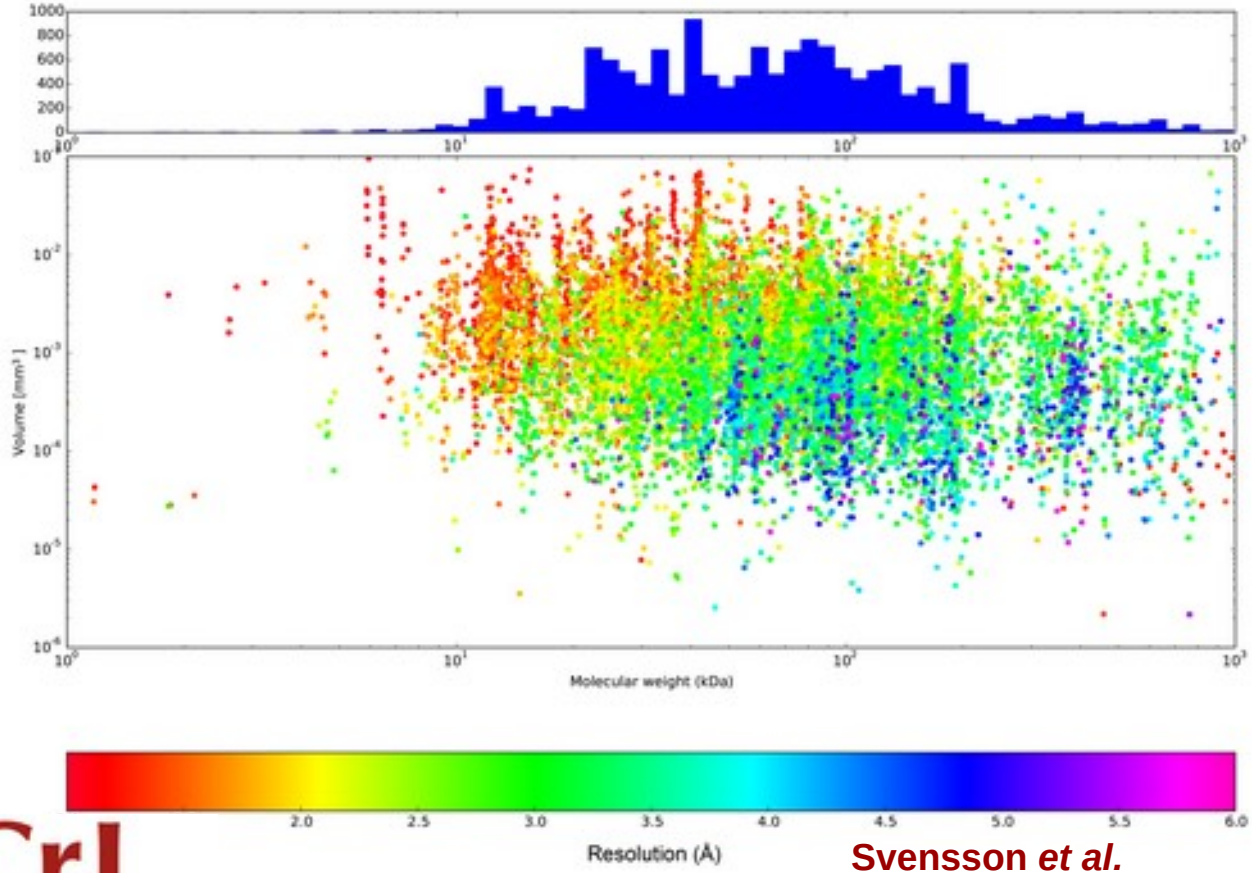
215 µm edge

100 µm edge

46.4 µm edge

21.5 µm edge

10 µm edge

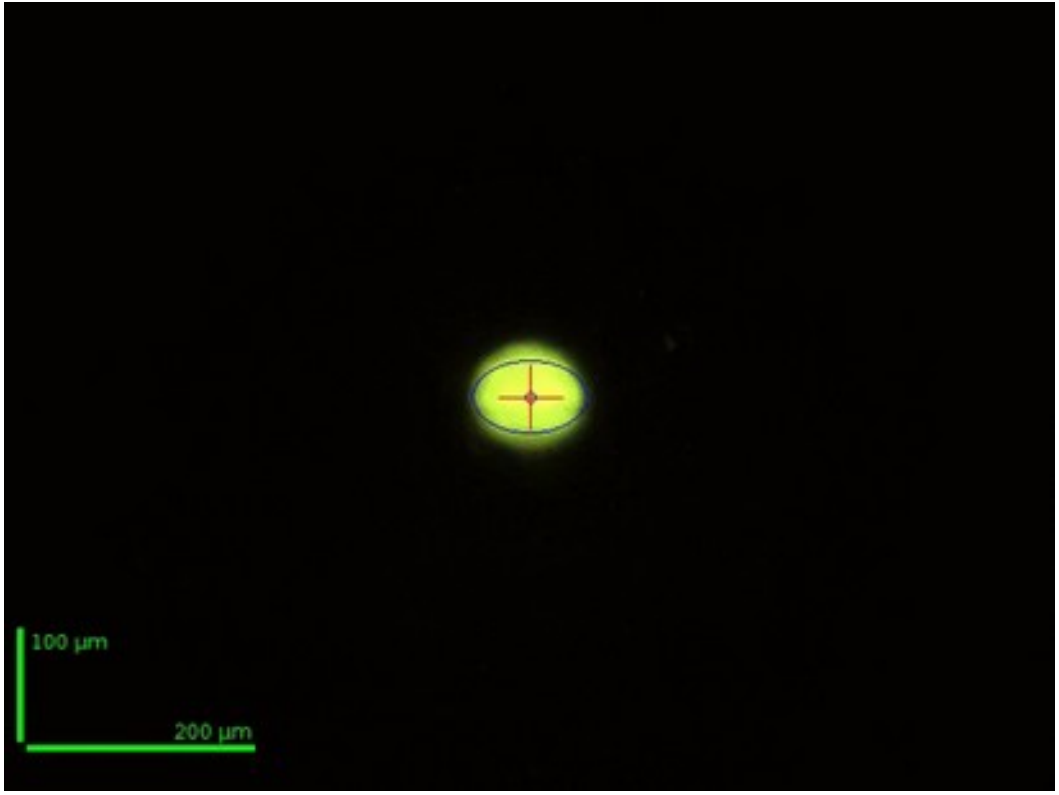


IUCrJ

Svensson *et al.*

Volume 6 | **Part 5** | September 2019 | Pages 822–831 |

EMBL



EMBL

β_1 adrenergic GPCR

Thanks to Tony Warne

Acta Cryst D STRUCTURAL BIOLOGY

Svensson *et al.*

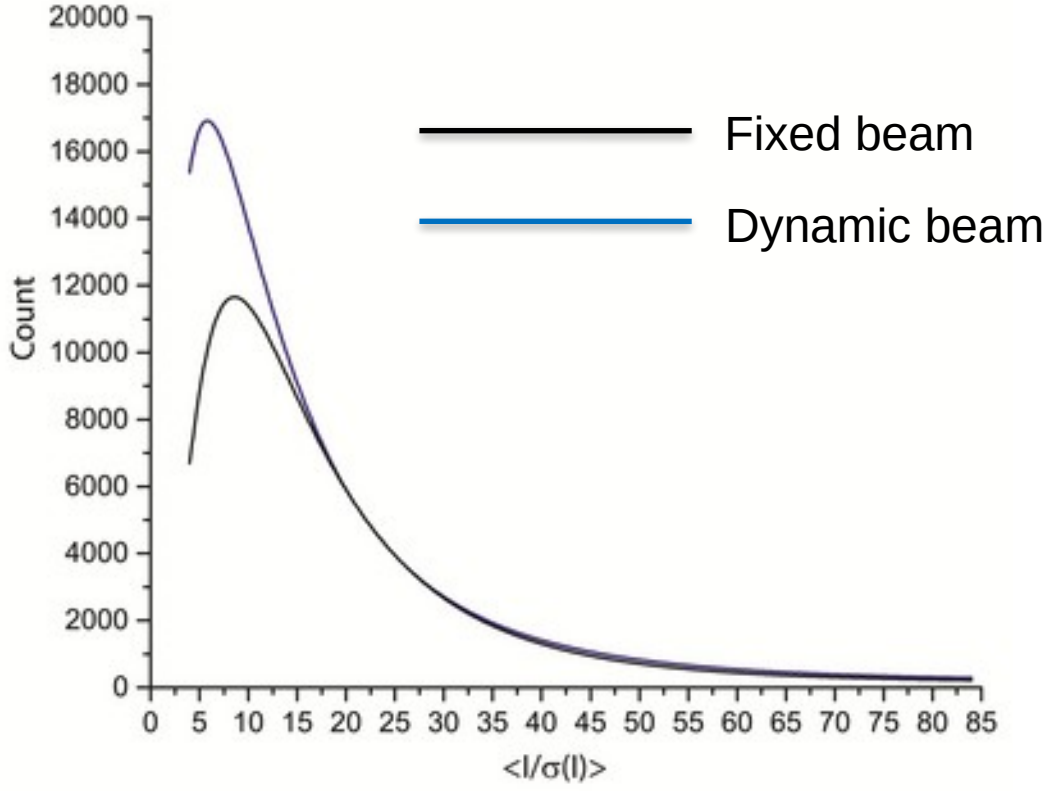
Volume 74 | **Part 5** | May 2018 | Pages 433–440 | 10.1107/S2059798318003728

β_1 adrenergic GPCR

Crystal	Crystal dimensions (x, y, z, mm)	Fixed beam diameter		Adaptable beam diameter	
		Resolution limit (Å)	<I/(I)>	Resolution limit (Å)	<I/(I)>
adrcpt-For41	0.109 x 0.053 x 0.025	3.77	6.7	4.13	4.4
adrcpt-For42	0.084 x 0.025 x 0.025	4.22	4.3	3.53	10.6
adrcpt-For45	0.035 x 0.045 x 0.051	3.95	6.2	-	-
adrcpt-For47	0.105 x 0.061 x 0.051	3.74	4.7	3.72	5.4
adrcpt-For48	0.105 x 0.039 x 0.064	-	-	3.8	5.7
adrcpt-For58	0.169 x 0.050 x 0.061	3.88	6.6	4.11	4.5
adrcpt-For59	0.042 x 0.024 x 0.025	3.25	9.2	3.16	8.3
adrcpt-For67	0.064 x 0.026 x 0.031	-	-	3.8	5.6

β_1 adrenergic GPCR

Crystal	Crystal dimensions (x, y, z, mm)	Fixed beam diameter		Adaptable beam diameter	
		Resolution limit (Å)	<I/▦(I)>	Resolution limit (Å)	<I/▦(I)>
adrcpt-For41	0.109 x 0.053 x 0.025	3.77	6.7	4.13	4.4
adrcpt-For42	**0.084 x 0.025 x 0.025**	**4.22**	**4.3**	**3.53**	**10.6**
adrcpt-For45	0.035 x 0.045 x 0.051	3.95	6.2	-	-
adrcpt-For47	0.105 x 0.061 x 0.051	3.74	4.7	3.72	5.4
adrcpt-For48	0.105 x 0.039 x 0.064	-	-	3.8	5.7
adrcpt-For58	0.169 x 0.050 x 0.061	3.88	6.6	4.11	4.5
adrcpt-For59	**0.042 x 0.024 x 0.025**	**3.25**	**9.2**	**3.16**	**8.3**
adrcpt-For67	**0.064 x 0.026 x 0.031**	**-**	**-**	**3.8**	**5.6**

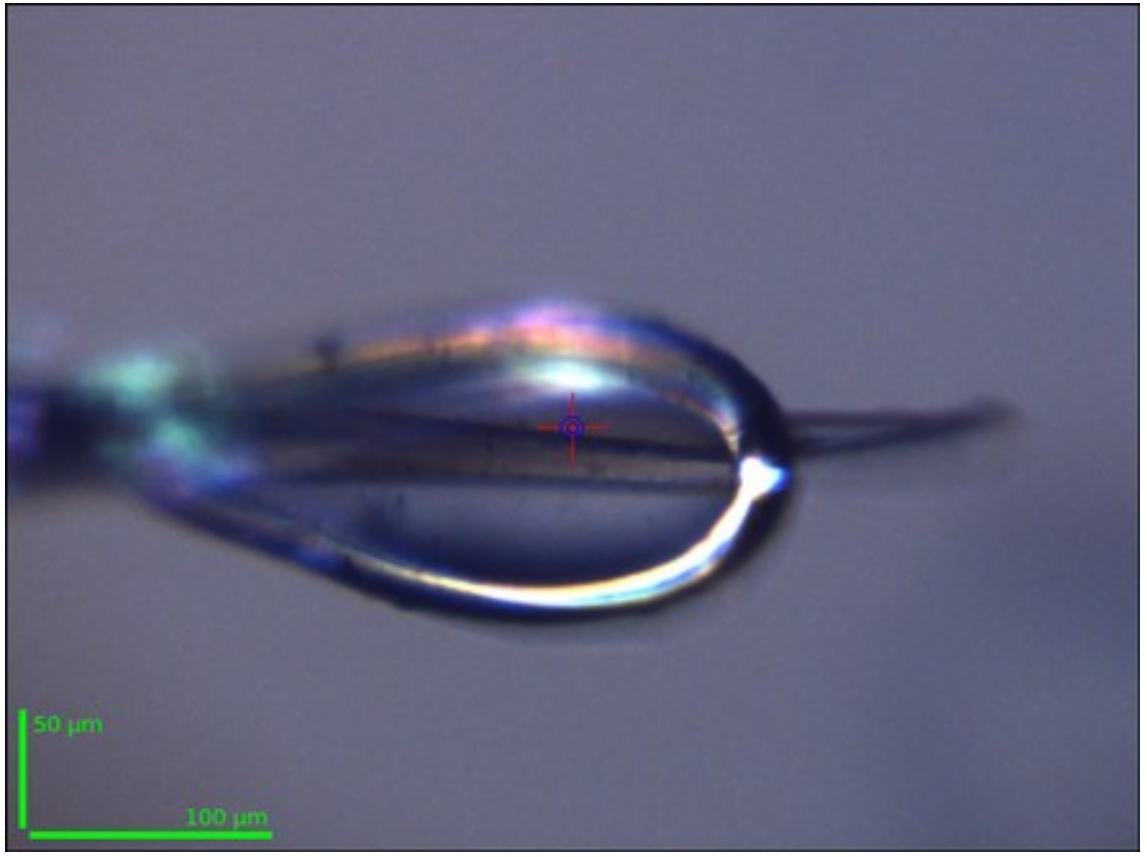


STRUCTURAL BIOLOGY

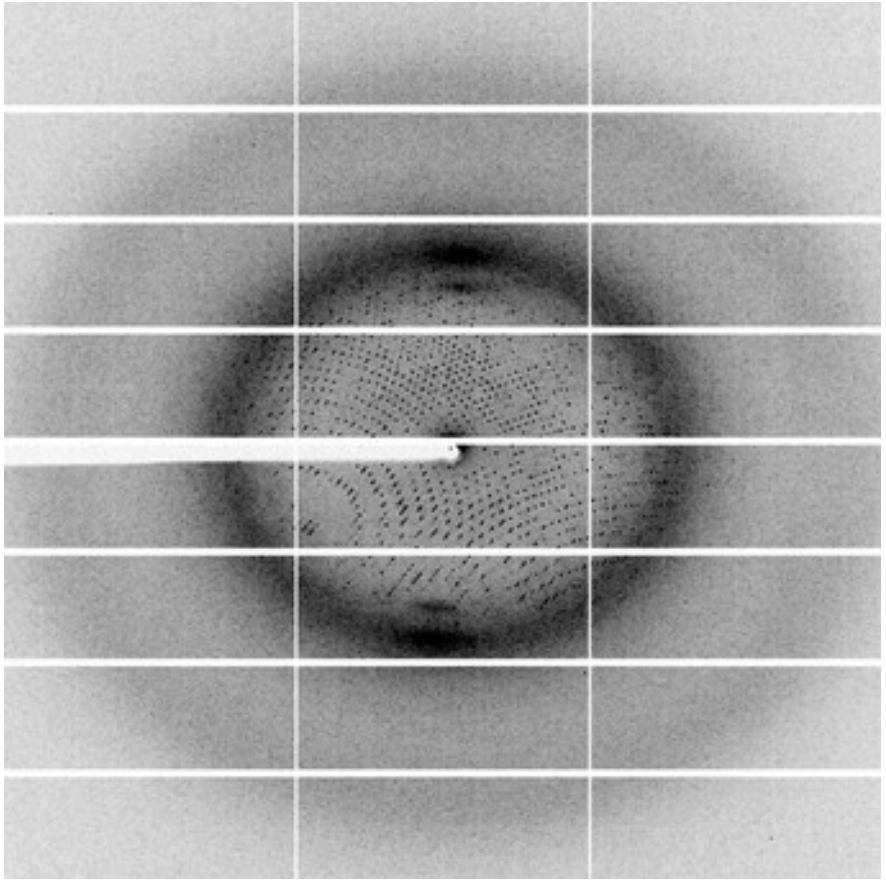
Svensson *et al.*

Volume 74 | **Part 5** | May 2018 | Pages 433–440 | 10.1107/S2059798318003728

EMBL



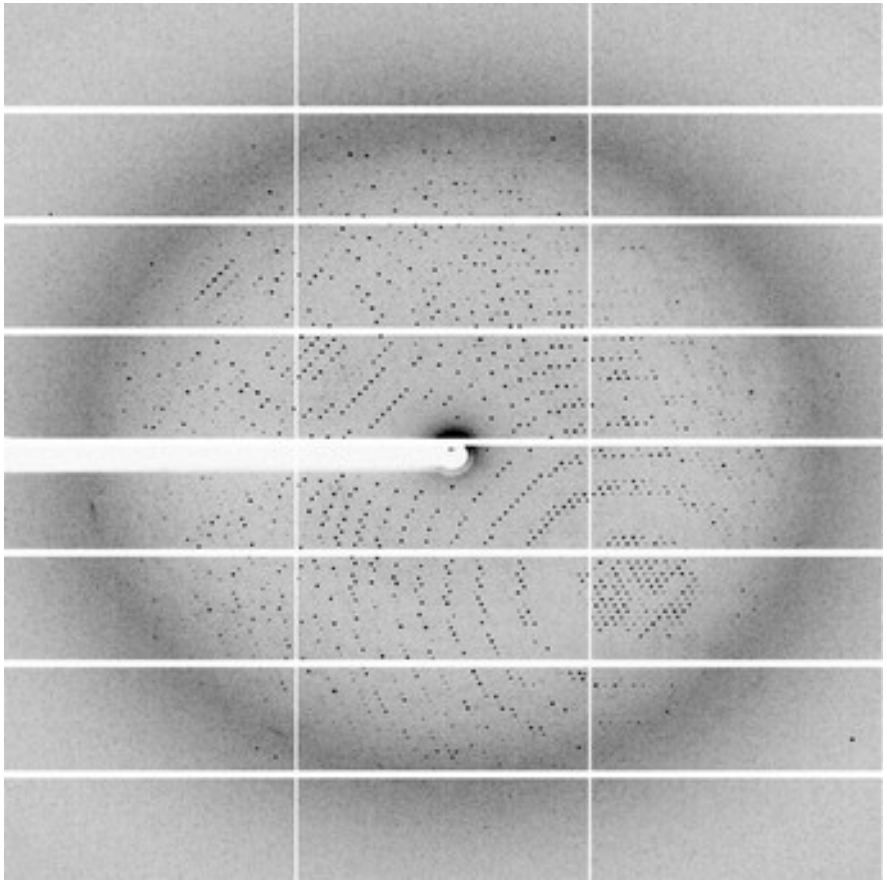
EMBL



2 Å

Flu polymerase

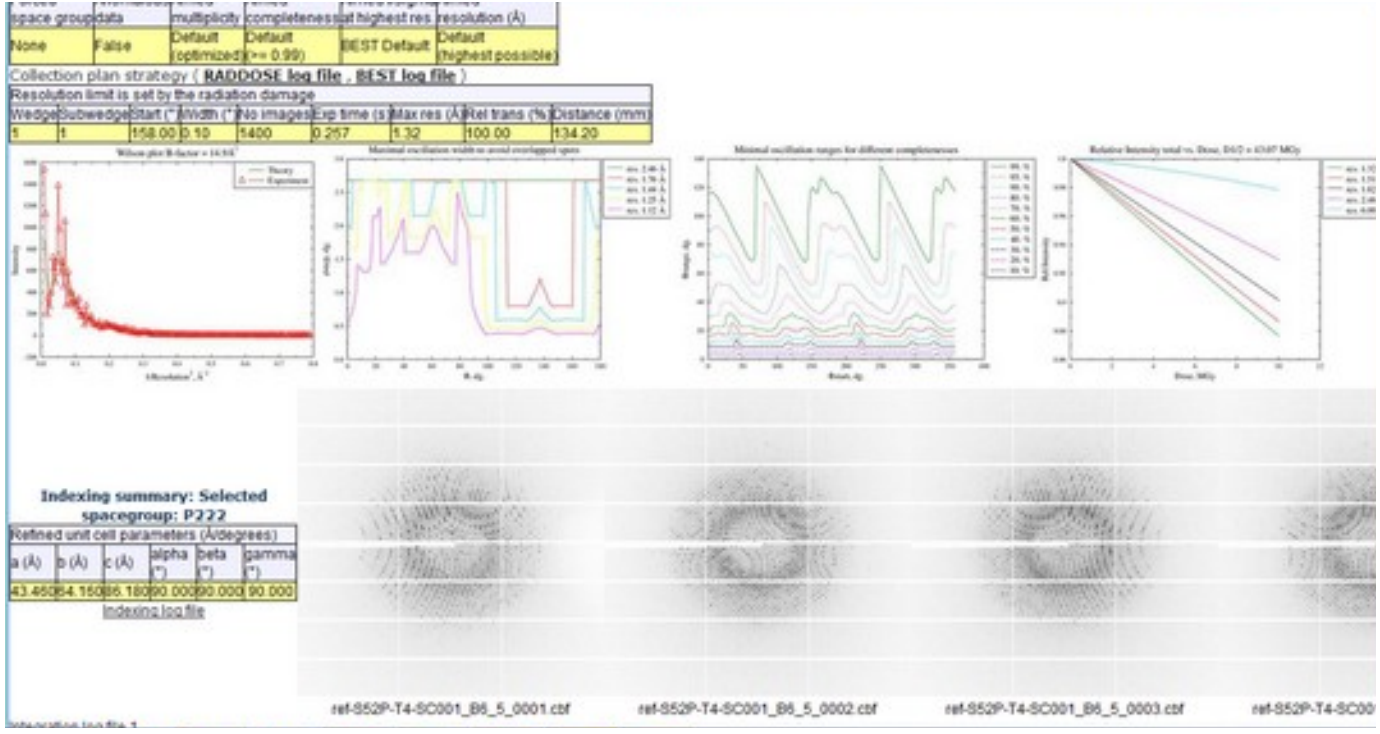
EMBL



3 Å

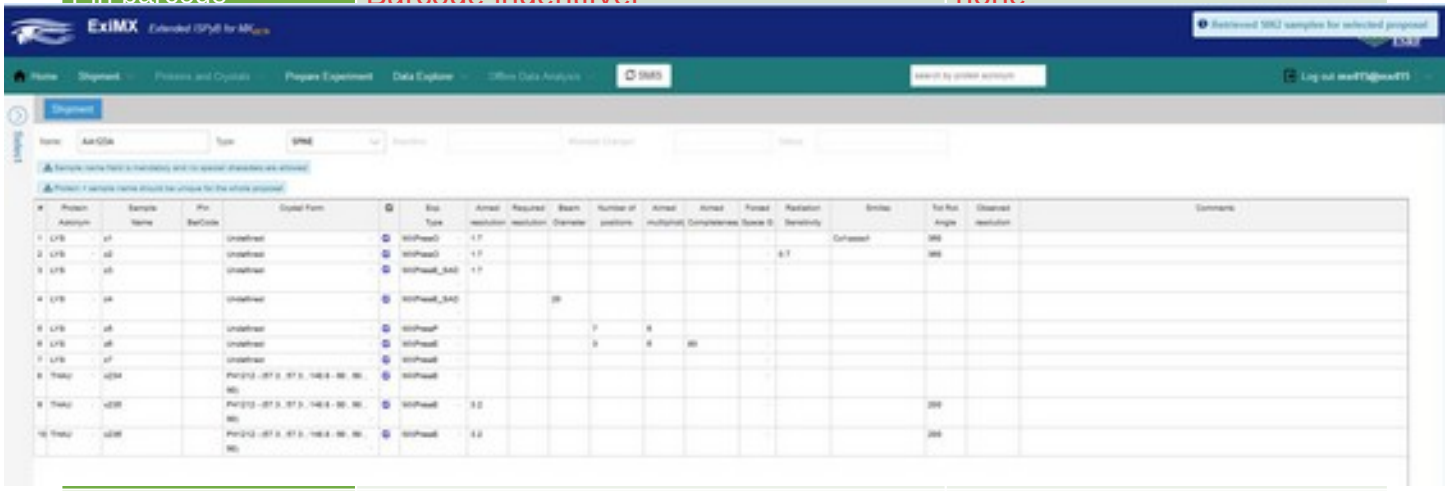
Flu polymerase

EMBL


space group
data
multiplicity
completeness
at highest res
resolution (Å)
None
False
Default (optimized)
Default (>= 0.99)
BEST Default
Default (highest possible)
Collection plan strategy (RADDOSE log file , BEST log file)
Resolution limit is set by the radiation damage
Wedge
Subwedge
Start (°)
Width (°)
No images
Exp time (s)
Max res (Å)
Xtal trans (%)
Distance (mm)
1
1
158.00
0.10
1400
0.257
1.32
100.00
134.20
Indexing summary: Selected spacegroup: P222
Refined unit cell parameters (Å/degrees)
a (Å)
b (Å)
c (Å)
alpha (°)
beta (°)
gamma (°)
43.465
84.150
86.180
90.000
90.000
90.000
Indexing log file
ref-S52P-T4-SC001_B6_5_0001.cbf
ref-S52P-T4-SC001_B6_5_0002.cbf
ref-S52P-T4-SC001_B6_5_0003.cbf

Diffraction Plan entry	Definition	Default value
Protein acronym	Defines the protein that is registered with the ESRF safety group	Required field
Sample name	User defined unique identifier	Required field
Pin barcode	Barcode indentifyer	none
Experiment type	Define MXPressE / O / SAD / Score	MXPressE
Space Group	If present used for strategy calculation and autoprocessing	none
Pre-observed resolution	Resolution that the detector will be set to for mesh scans, characterisation images and default data collection	2.0 Å
Required resolution	Threshold resolution, samples below cutoff will not be collected	none
Radiation sensitivity	BEST input in case of highly radiation sensitive crystals	1
Total rotation range	Total required rotation for data set	minimum
Required completeness	-	99%
Required multiplicity	-	4
Number of positions	For multiple crystals	1
Preferred beamsize	Select appropriate beamsize for crystals	Adapted to crystal volume

Diffraction Plan entry	Definition	Default value
Protein acronym	Defines the protein that is registed with the ESRF safety group	Required field
Sample name	User defined unique identifier	Required field
Pin barcode	Barcode indentifyer	none



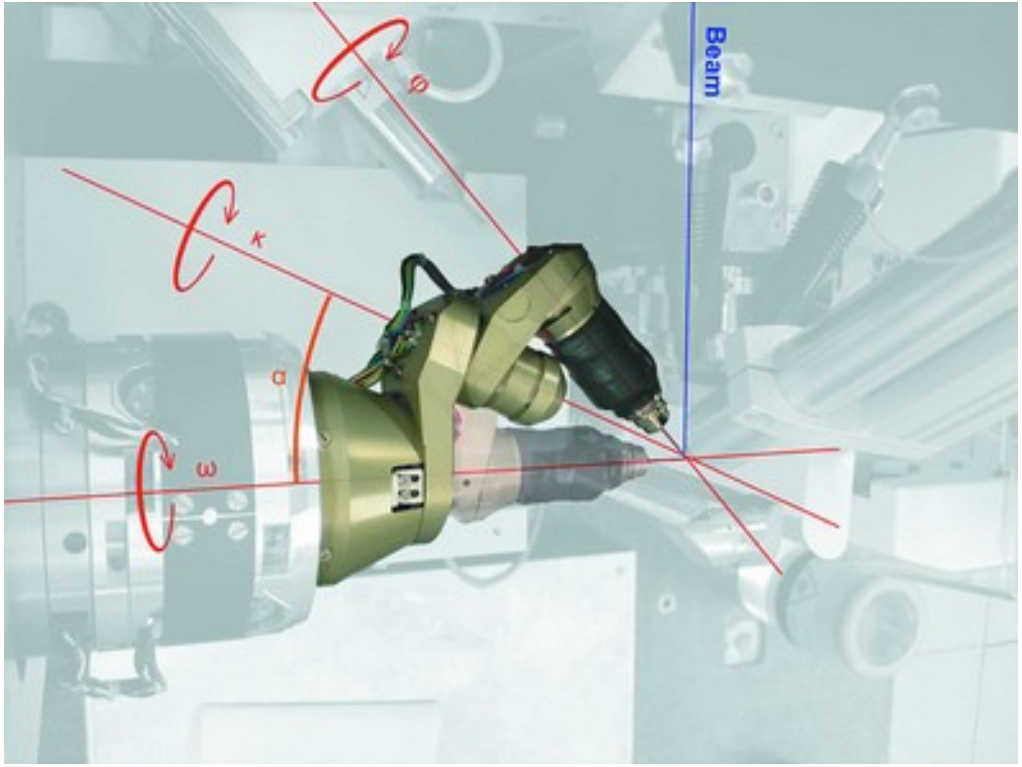
Required completeness	-	99%
Required multiplicity	-	4
Number of positions	For multiple crystals	1
Preferred beamsize	Select appropriate beamsize for crystals	50 µm

Acta Cryst D STRUCTURAL BIOLOGY

Svensson *et al.*

Volume 71 | **Part 8** | August 2015 | Pages 1757–1767

EMBL



Acta Cryst D STRUCTURAL BIOLOGY

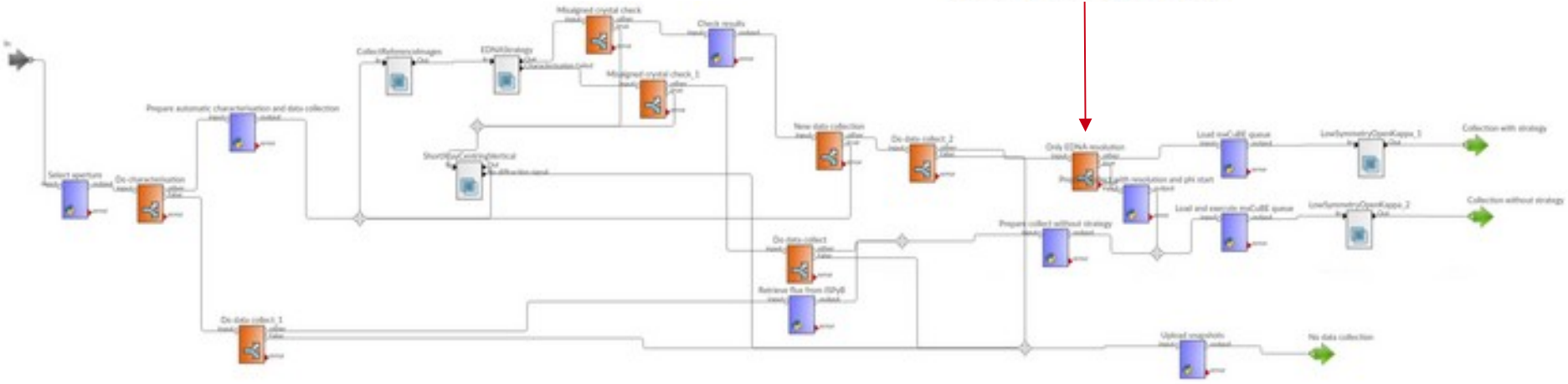
Brockhauser *et al.*

Volume 69 | **Part 7** | July 2013 | Pages 1241–1251 | 10.1107/S0907444913003880

EMBL

GΦL

Global Phasing Limited



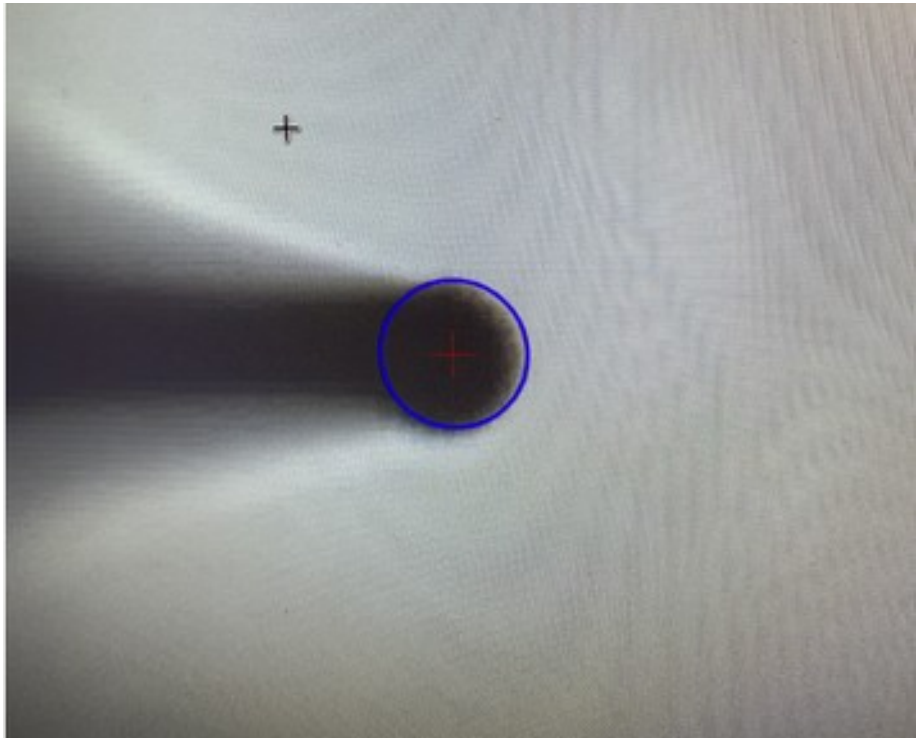
STRUCTURAL BIOLOGY

Svensson *et al.*

Volume 71 | **Part 8** | August 2015 | Pages 1757–1767

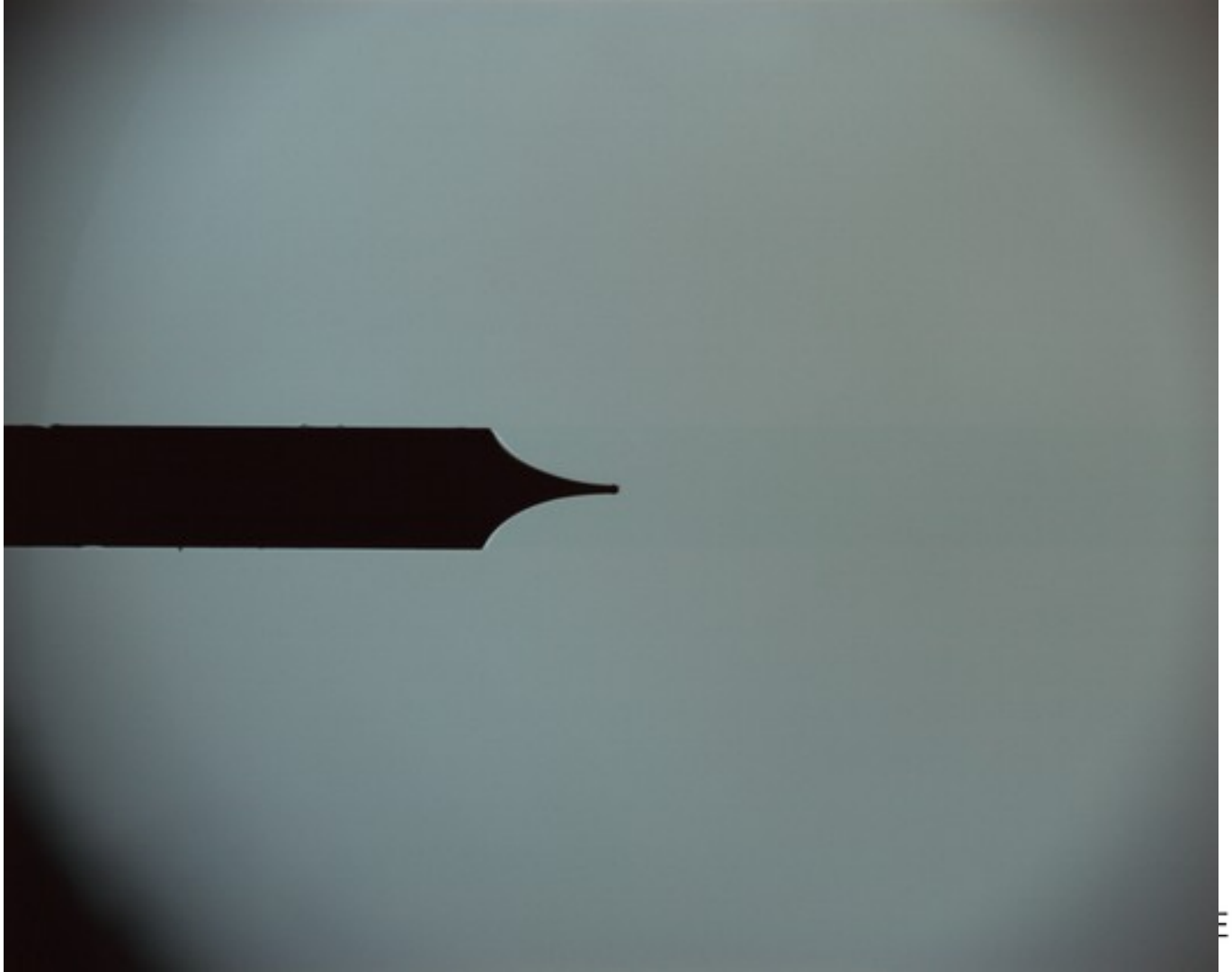
EMBL

Translation and rotation calibration using tungsten balls

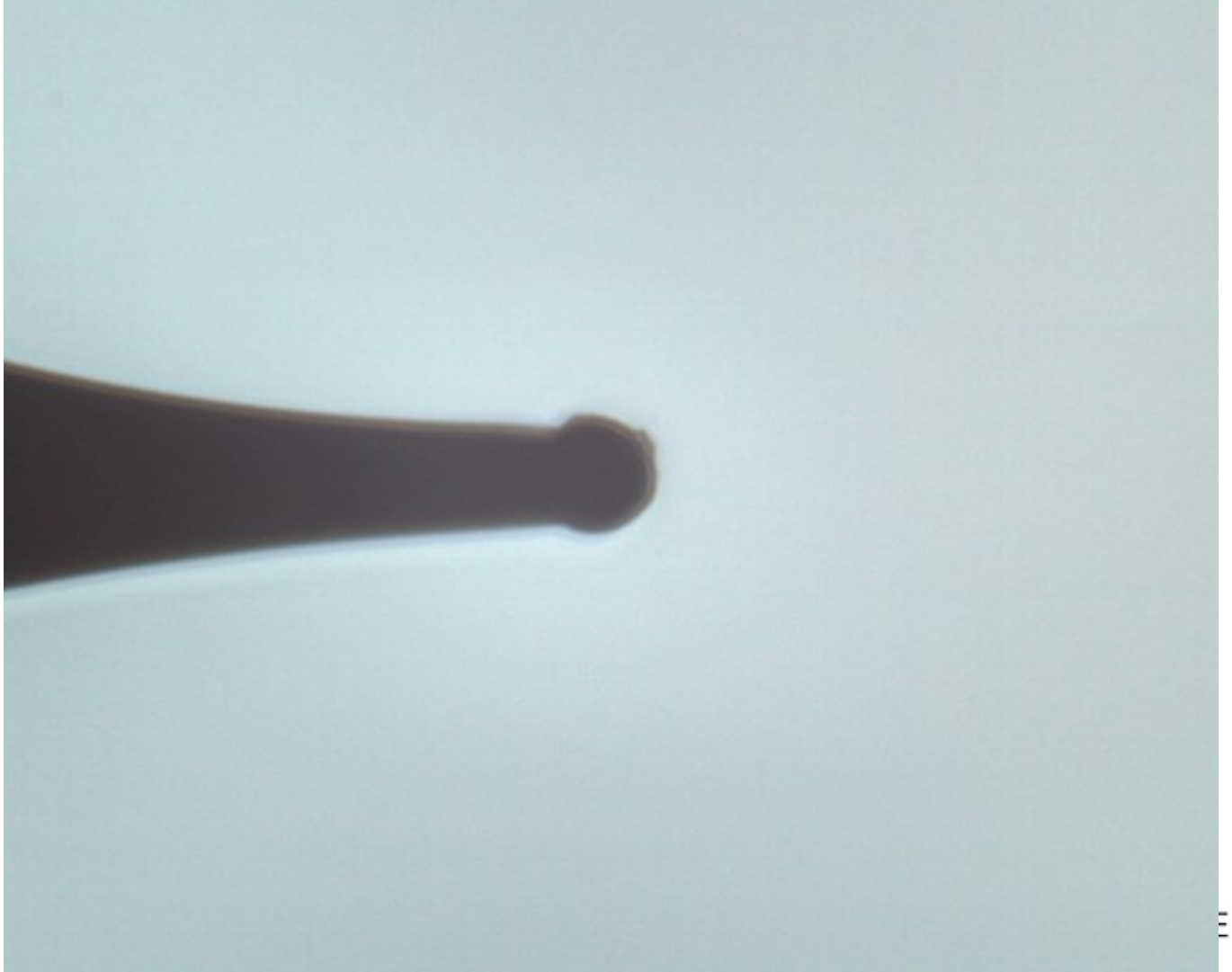


Thanks to Robin Lener, Stefan Kubsky *et al.* Synchrotron Soleil

EMBL

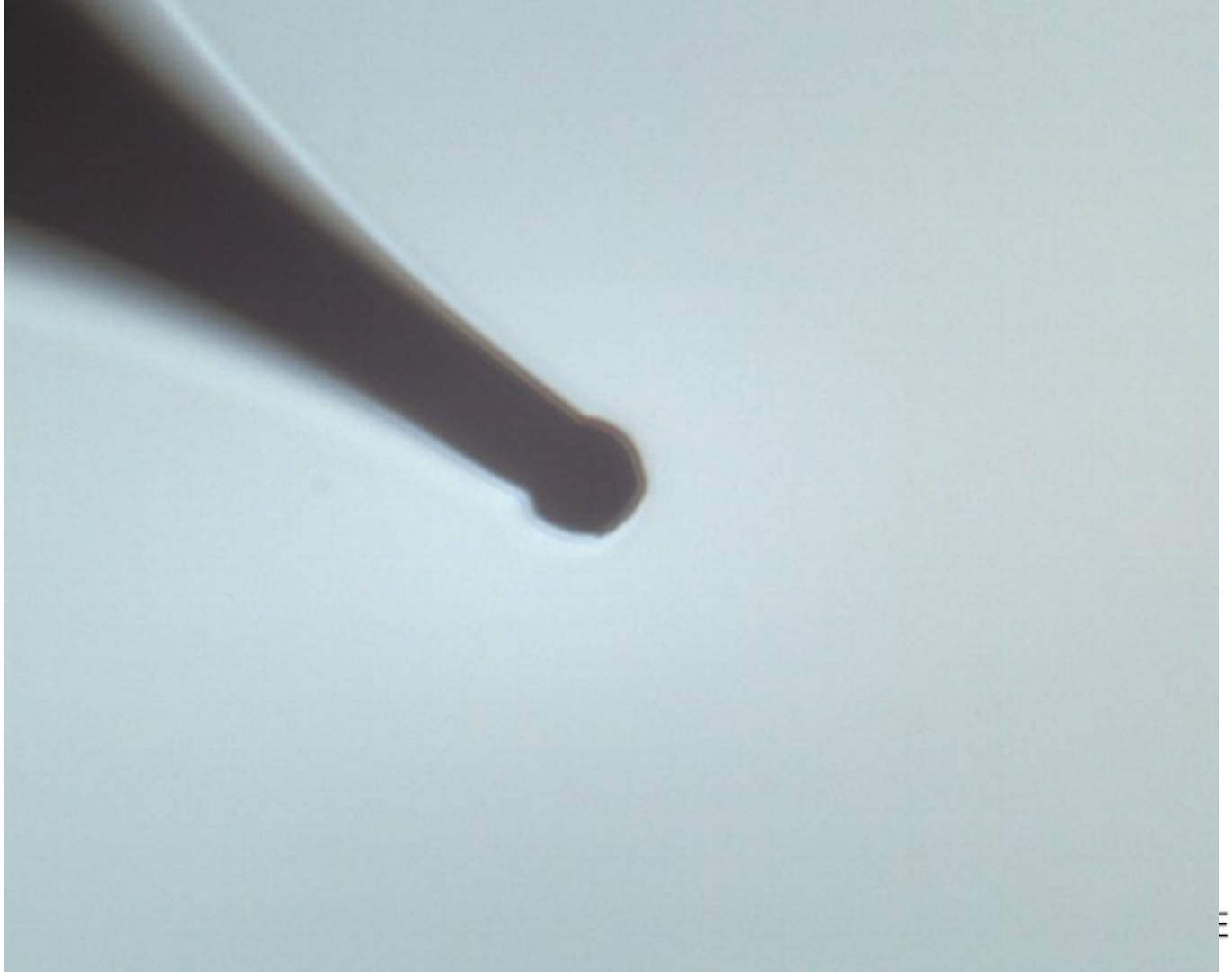


EMBL



EMBL

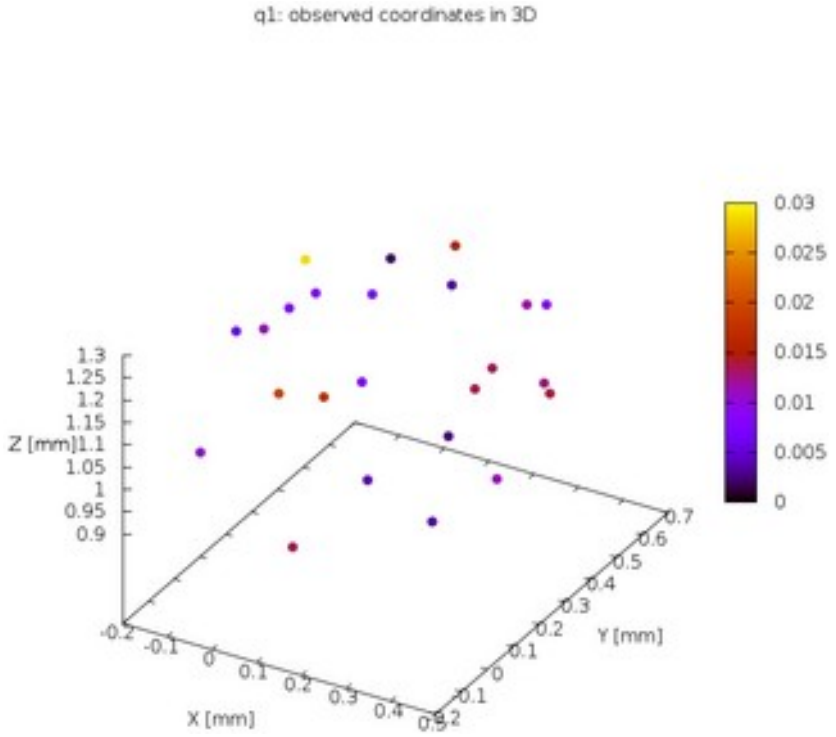
EMBL



EMBL

GΦL

Global Phasing Limited



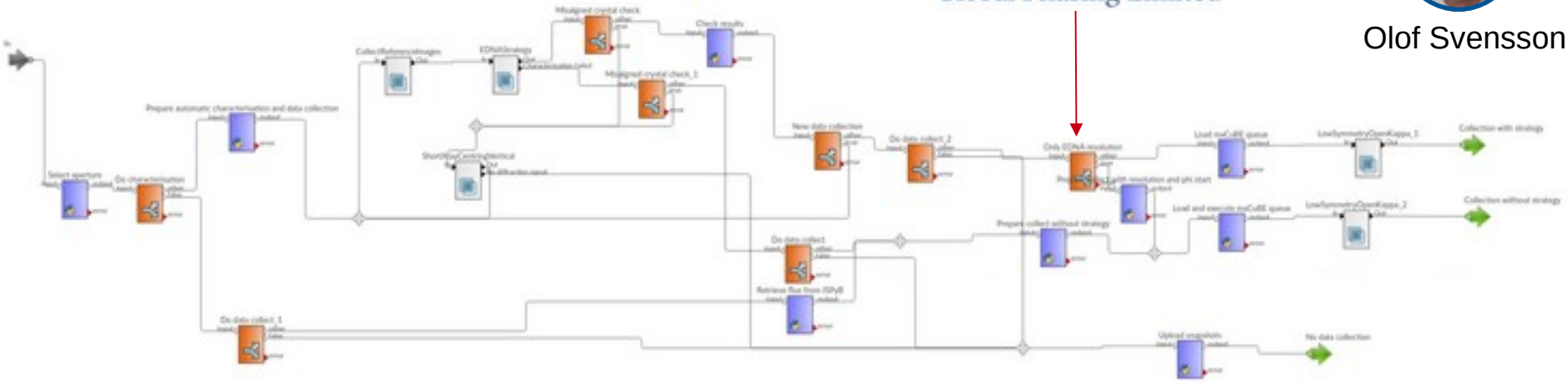
Automatic workflow centres ball over full kappa and phi range and runs GPhL programs recen
and updates calibration values

EMBL

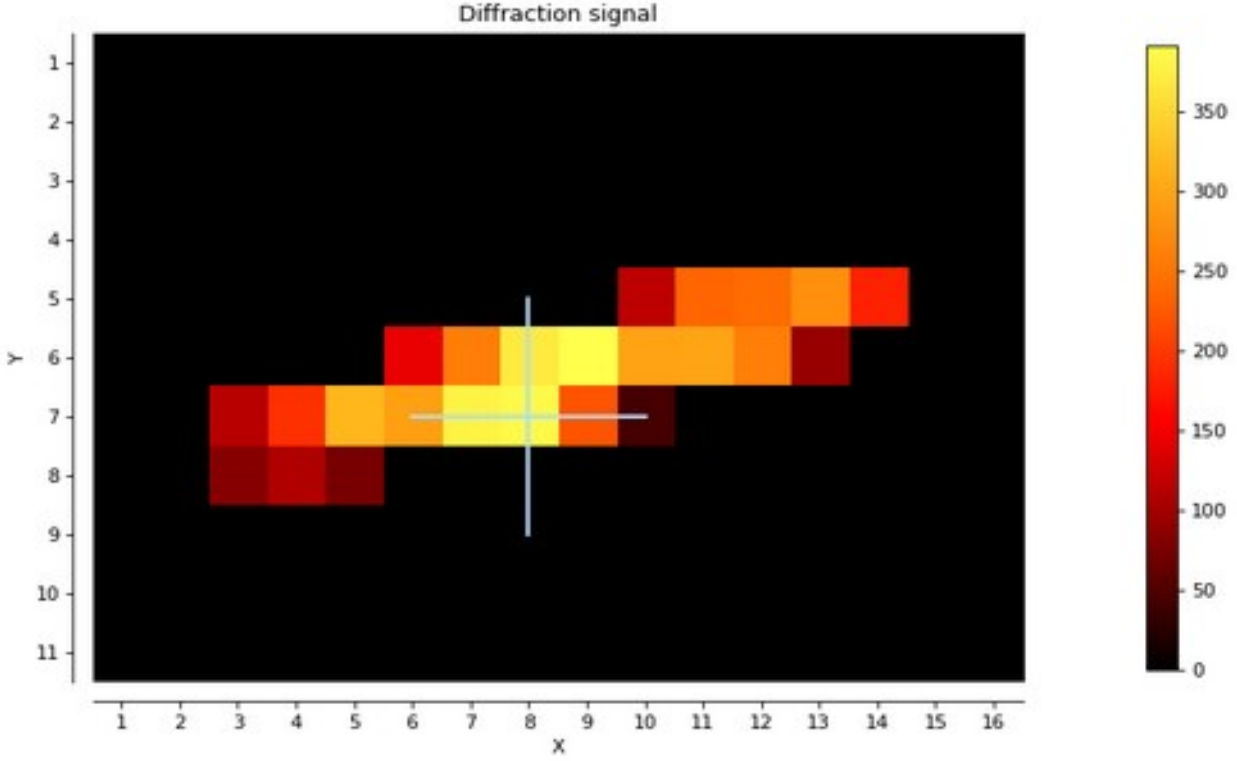
GΦL

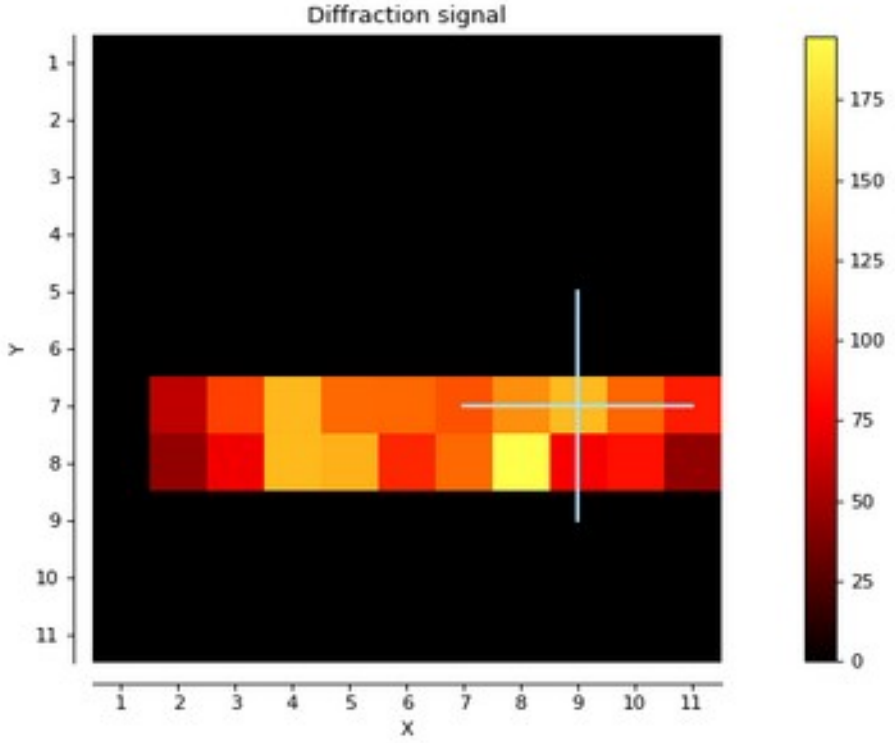
Global Phasing Limited

Olof Svensson

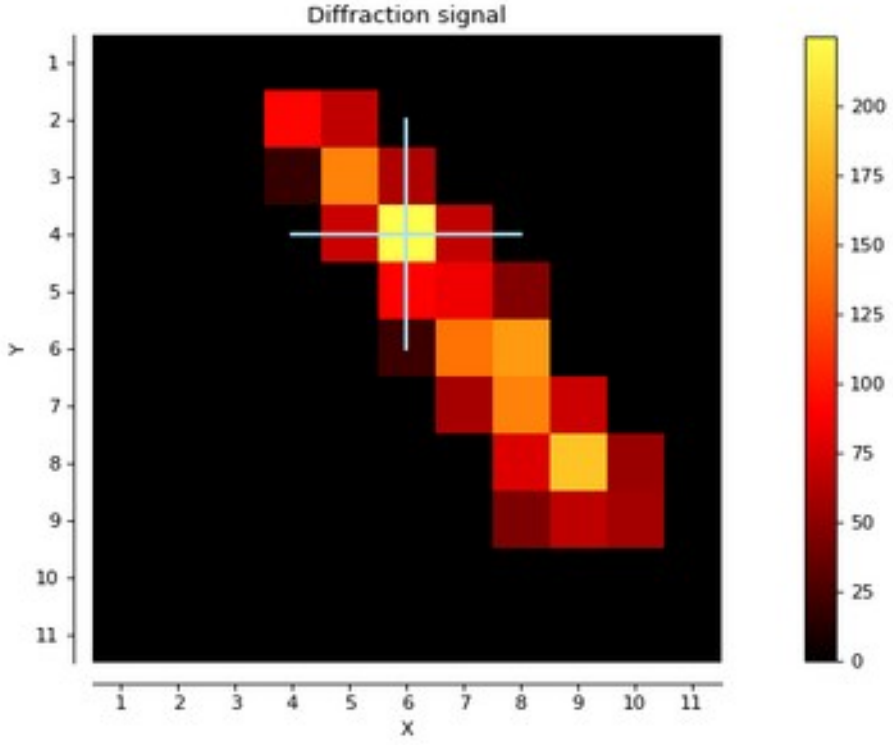


ESRF/EMBL workflow runs to characterisation
Passes over: **resolution**, **SG**, **dose budget**, **user requirements**

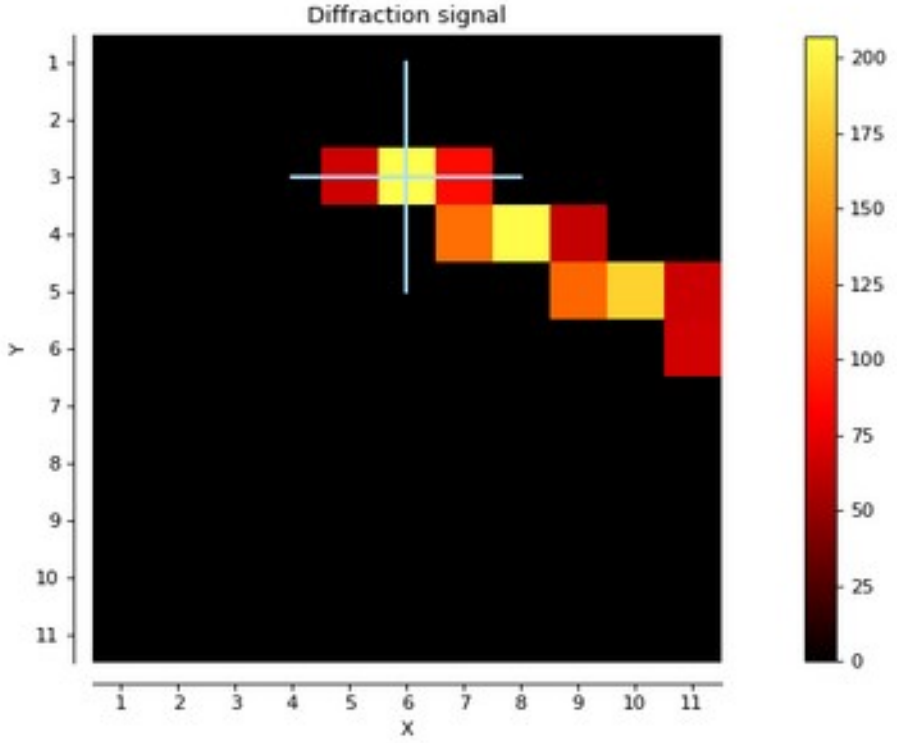

Diffraction signal
Y
X
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
350
300
250
200
150
100
50
0


Diffraction signal
Y
X
1
2
3
4
5
6
7
8
9
10
11
175
150
125
100
75
50
25
0

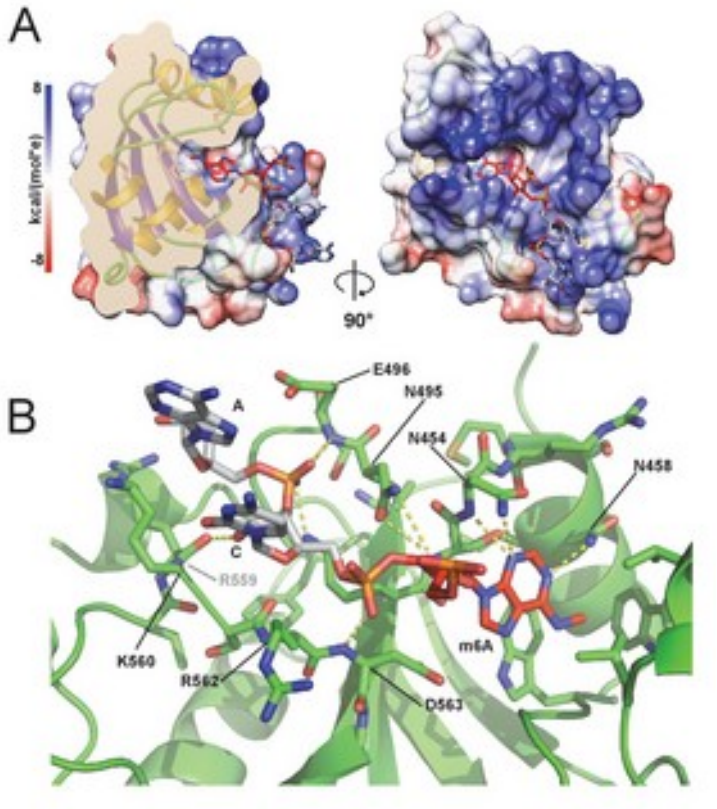
EMBL


Diffraction signal
Y
X
1
2
3
4
5
6
7
8
9
10
11
200
175
150
125
100
75
50
25
0

EMBL


Diffraction signal
Y
X
1
2
3
4
5
6
7
8
9
10
11
200
175
150
125
100
75
50
25
0

EMBL



YTH domain of CPCF4 from *T. gondii* – an M6A reader involved in gene partitioning

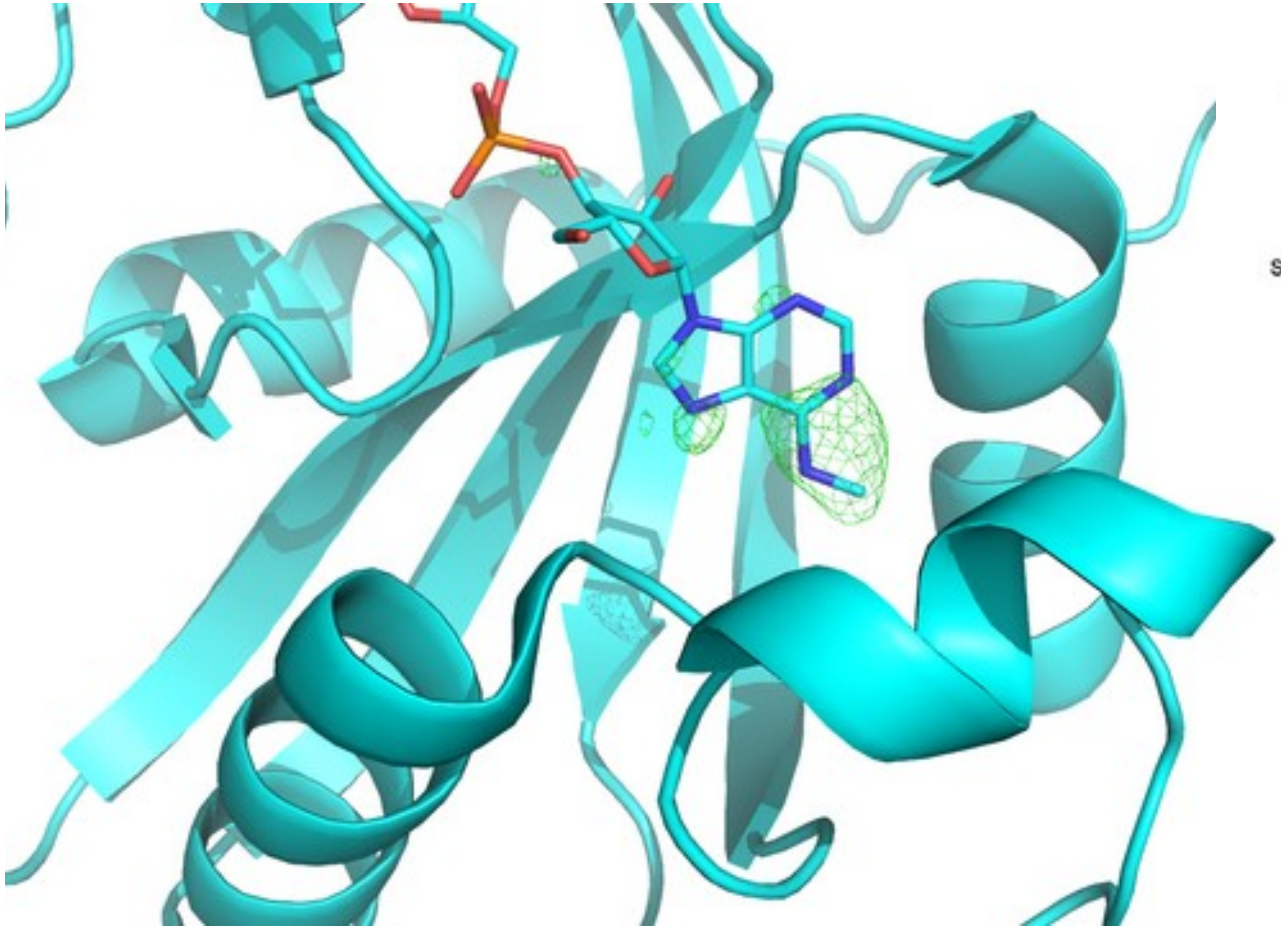
*P*1

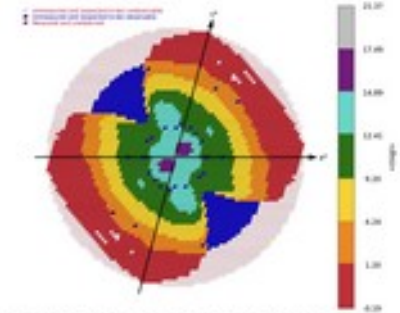
a 32.45 Å, b 35.14 Å, c 38.36 Å

α 114.31 °β 101.28 °γ 97.34 °

D. C. Farhat, M. W. Bowler, G. Communie, D. Pontier, L. Belmudes, C. Mas, C. Corrao, Y. Couté, A. Bougdour, T. Lagrange, M. A. Hakimi, C. Swale, A plant-like mechanism coupling m6A reading to polyadenylation safeguards transcriptome integrity and developmental gene partitioning in Toxoplasma. *eLife* **10**, e68312 (2021).

EMBL

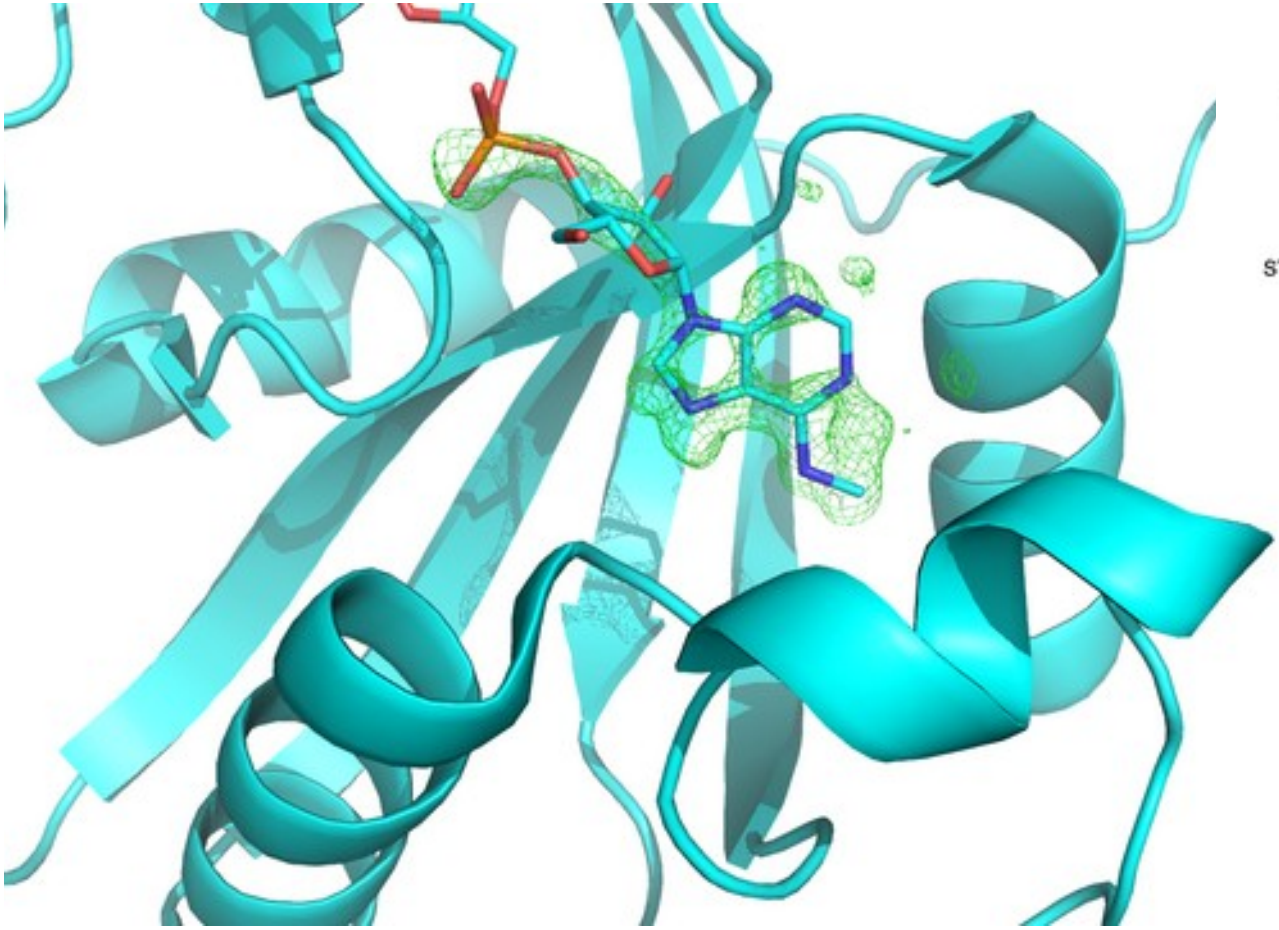


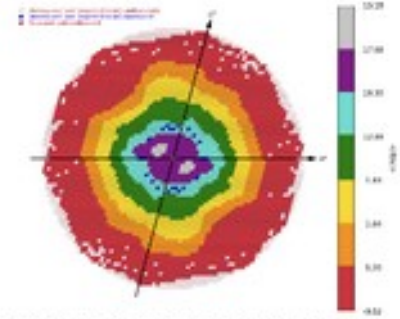


GΦL

Global Phasing Limited

EMBL



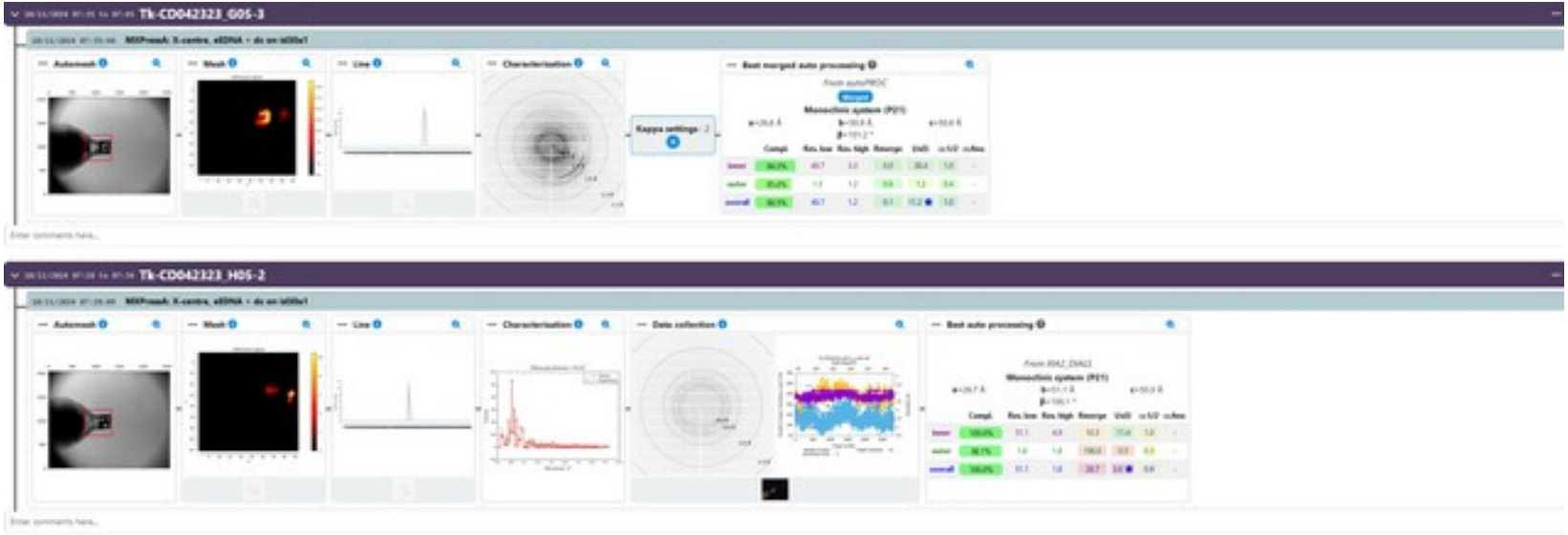


GΦL

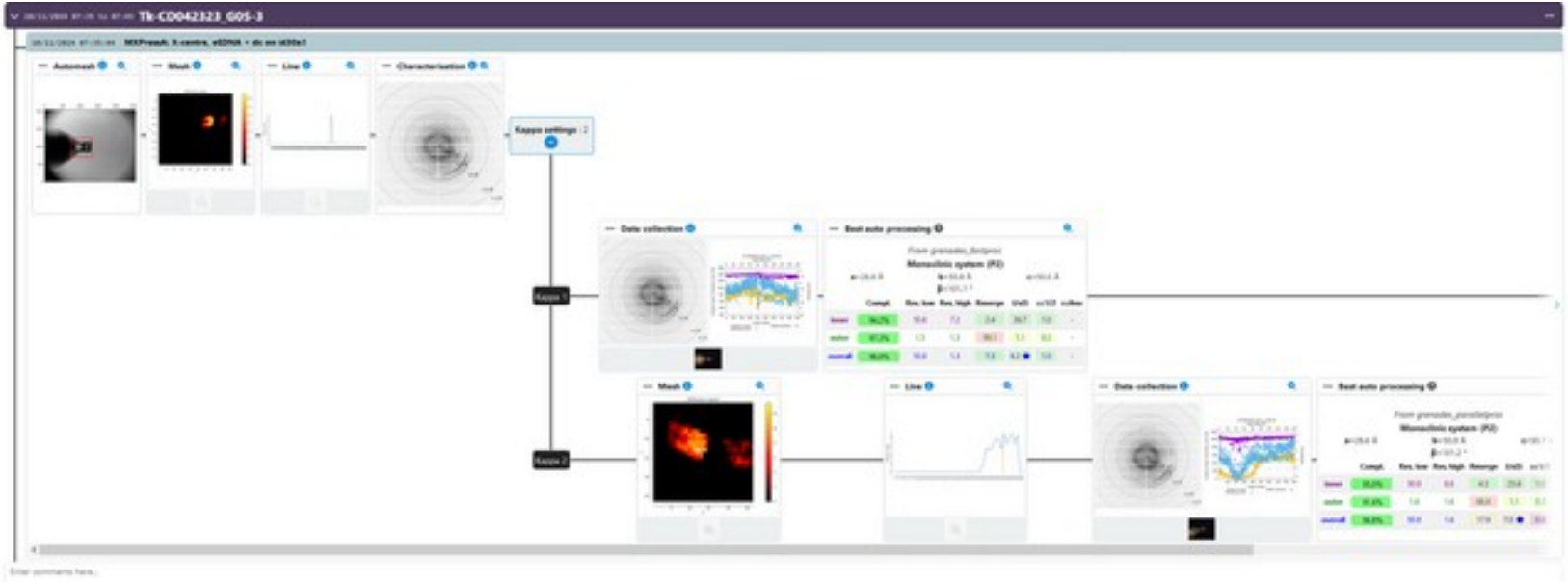
Global Phasing Limited

EMBL

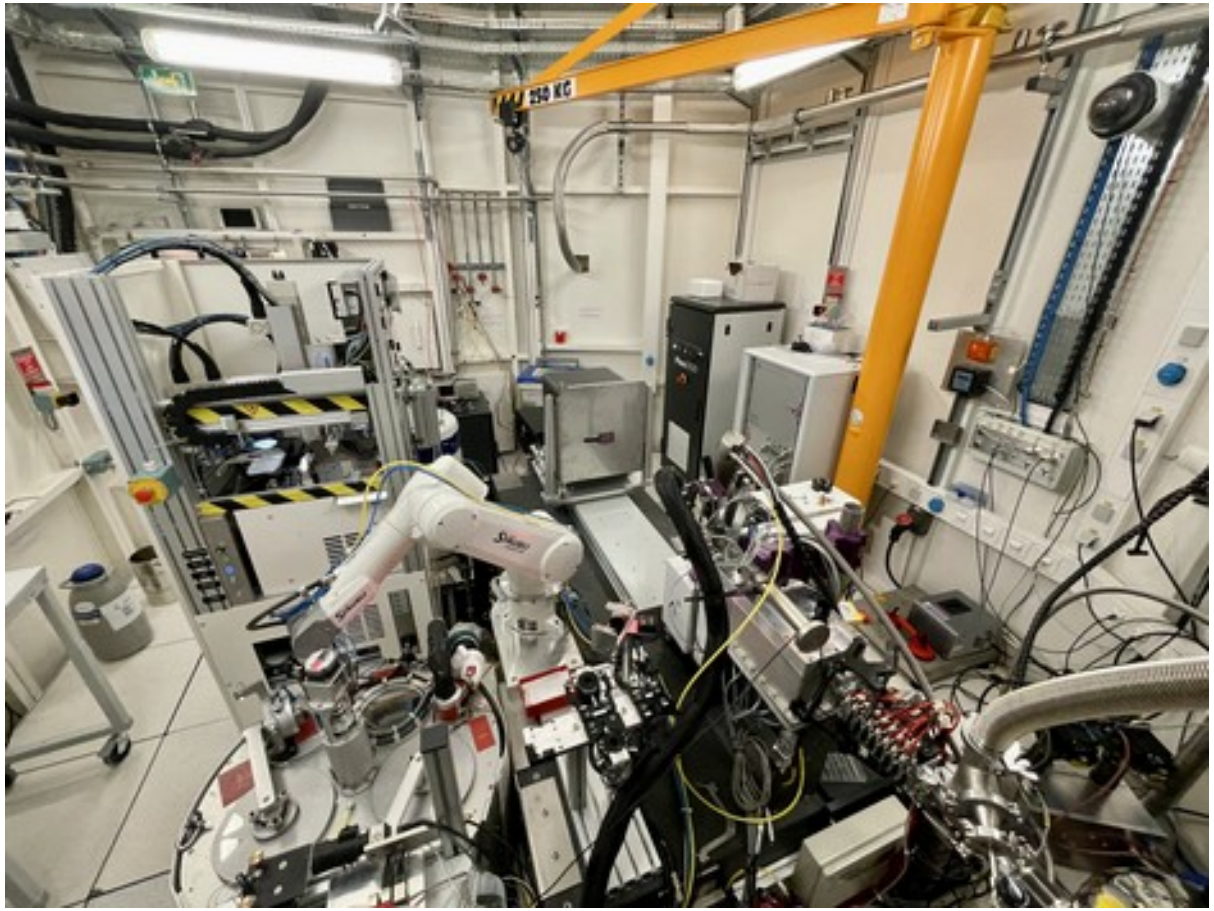
ISPyB-DRAC



ISPyB-DRAC


Automesh
Mesh
Line
Characterisation
Kappa settings
Data collection
Best auto processing
Monoclinic system (P2)
Mesh
Line
Data collection
Best auto processing
Monoclinic system (P2)

CrystalDirect harvester at MASSIF-1



Serena Rocchio

Automated crystal harvesting and data collection pipeline

Software integration

Crystal harvesting

Workflow

CrystalDirect harvester

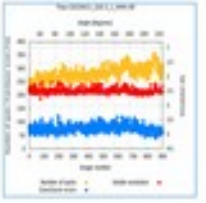
MX3

Harvesting Plan

CRIMS

Crystallization Information Management System

ISPyB - ExiMX



Data collection

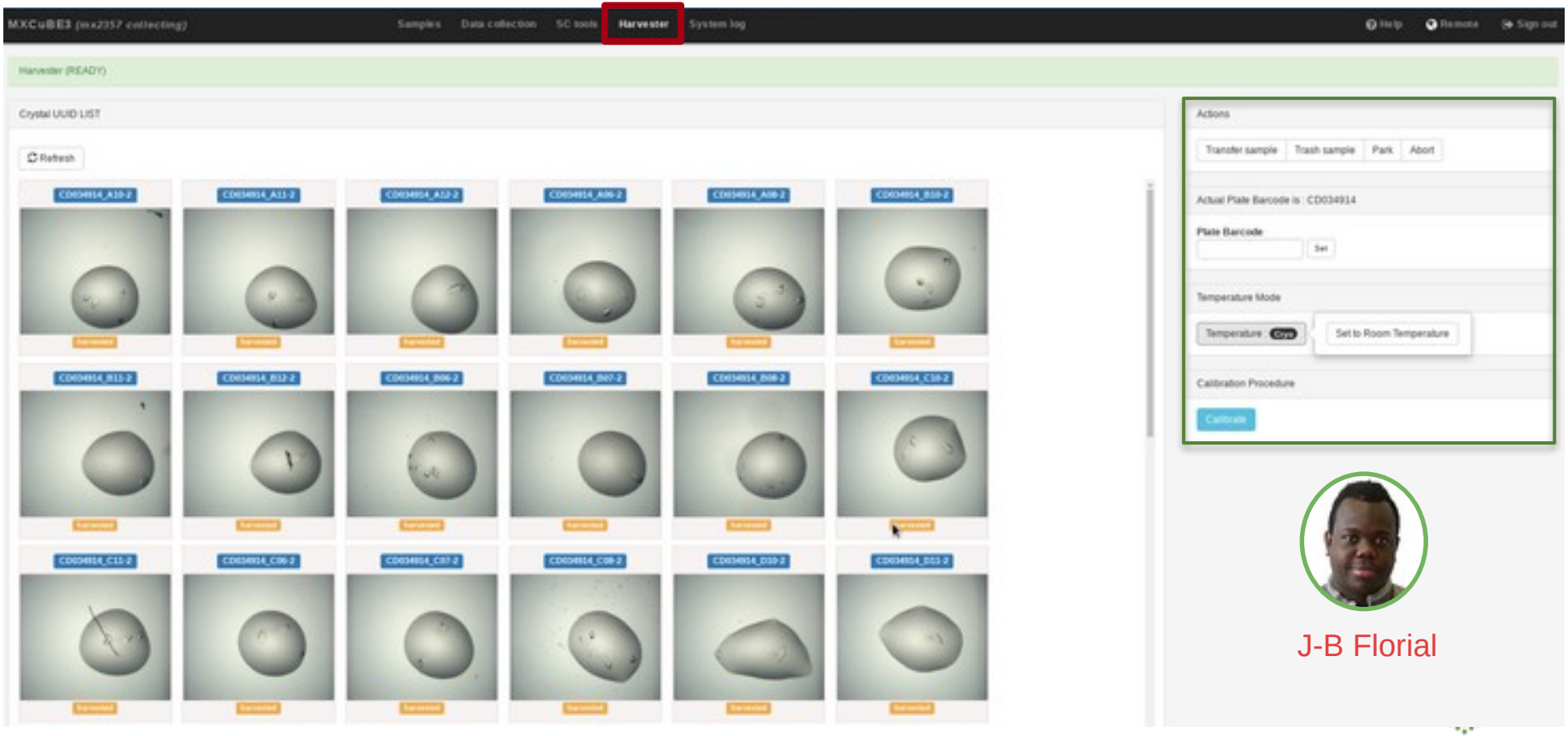
ESRF

Protein crystallisation

EMBL

Automated crystal harvesting and data collection pipeline

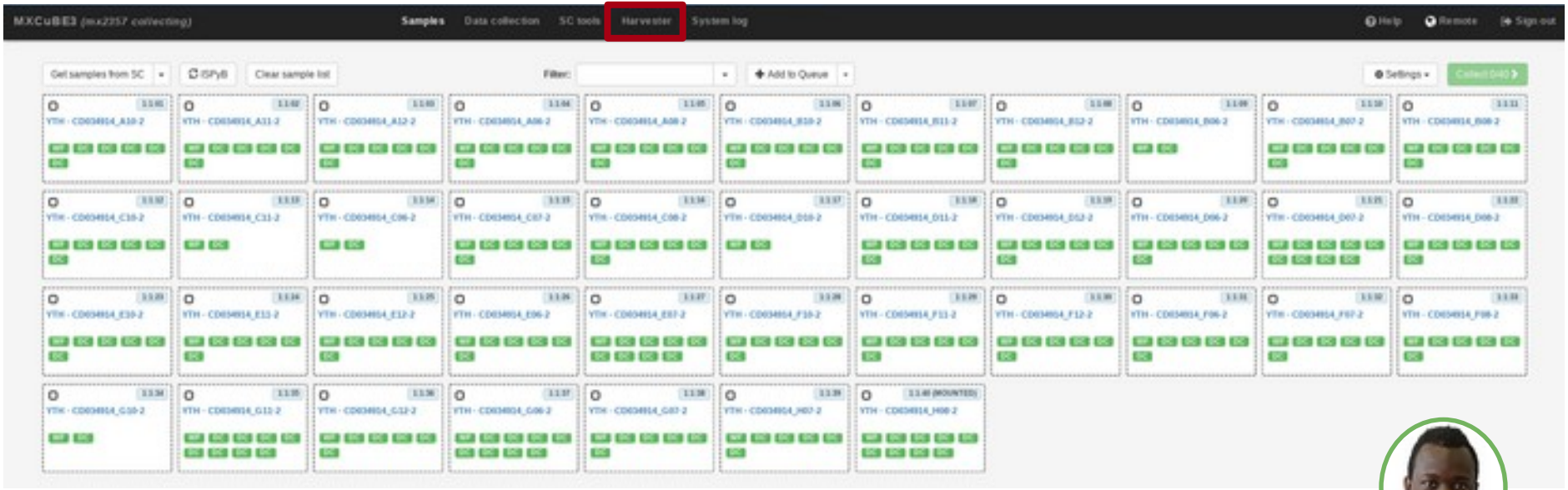
Software integration: MxCuBE^3 hardware object



J-B Florial

Automated crystal harvesting and data collection pipeline

Software integration: MxCuBE^3 hardware object



Samples harvested and collected in automated mode without user intervention

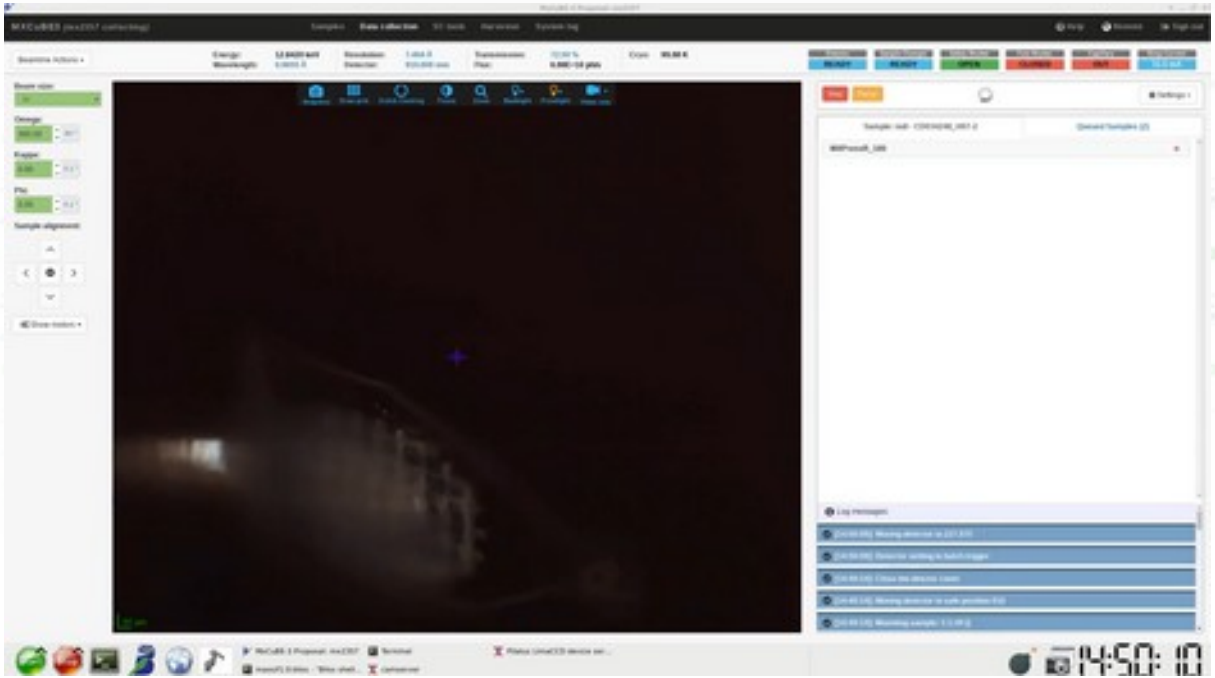
J-B Florial

ESRF

EMBL

Automated crystal harvesting and data collection pipeline

Software integration: MxCuBE^3 hardware object



Jeremy Sinoir

J-B Florial

ESRF

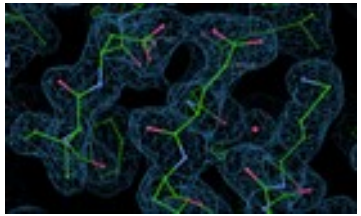
EMBL

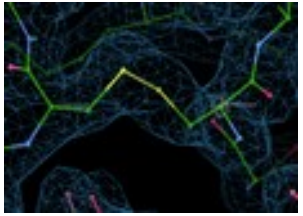
Automated crystal harvesting and data collection pipeline

Data Collection at Room Temperature - Results

Pipeline validation with model system
Workflow optimization

Thau

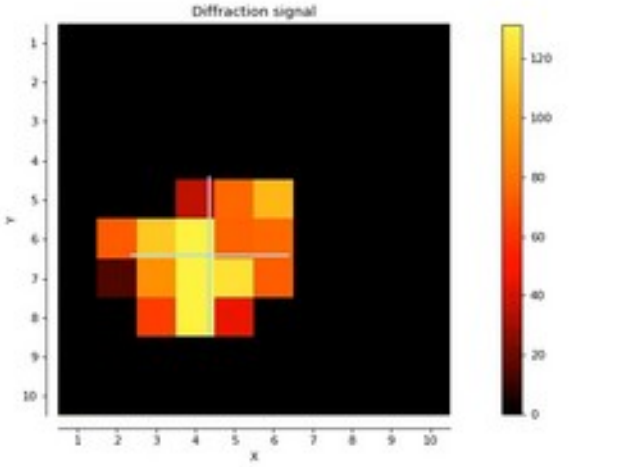




Space group: P 41 21 2
Unit cell: 58.2 58.2 151.2

R: 1.97 Å

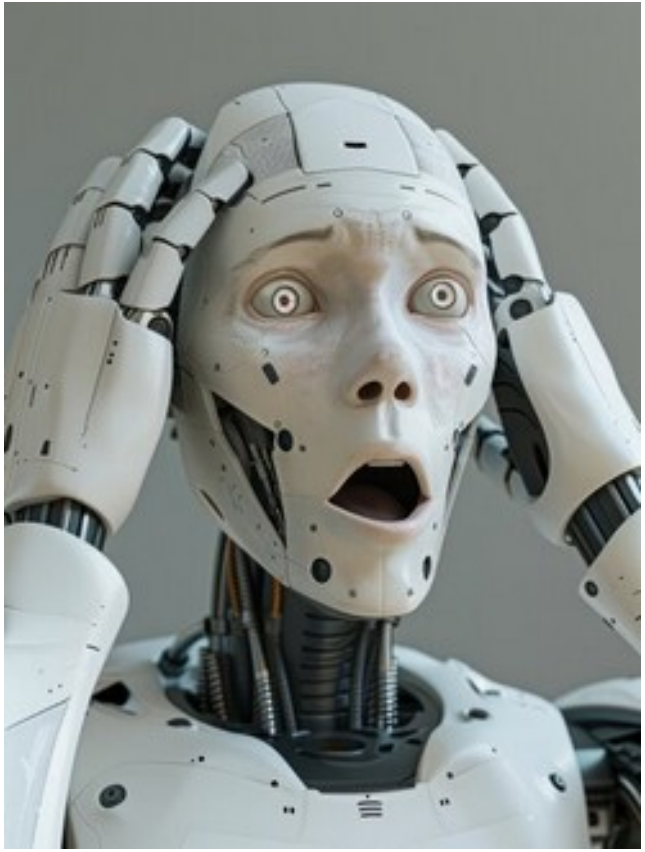
Mean B value: 36.12



Complete dataset collected from single crystal without impacting crystal propriety and no evident sign of radiation damage

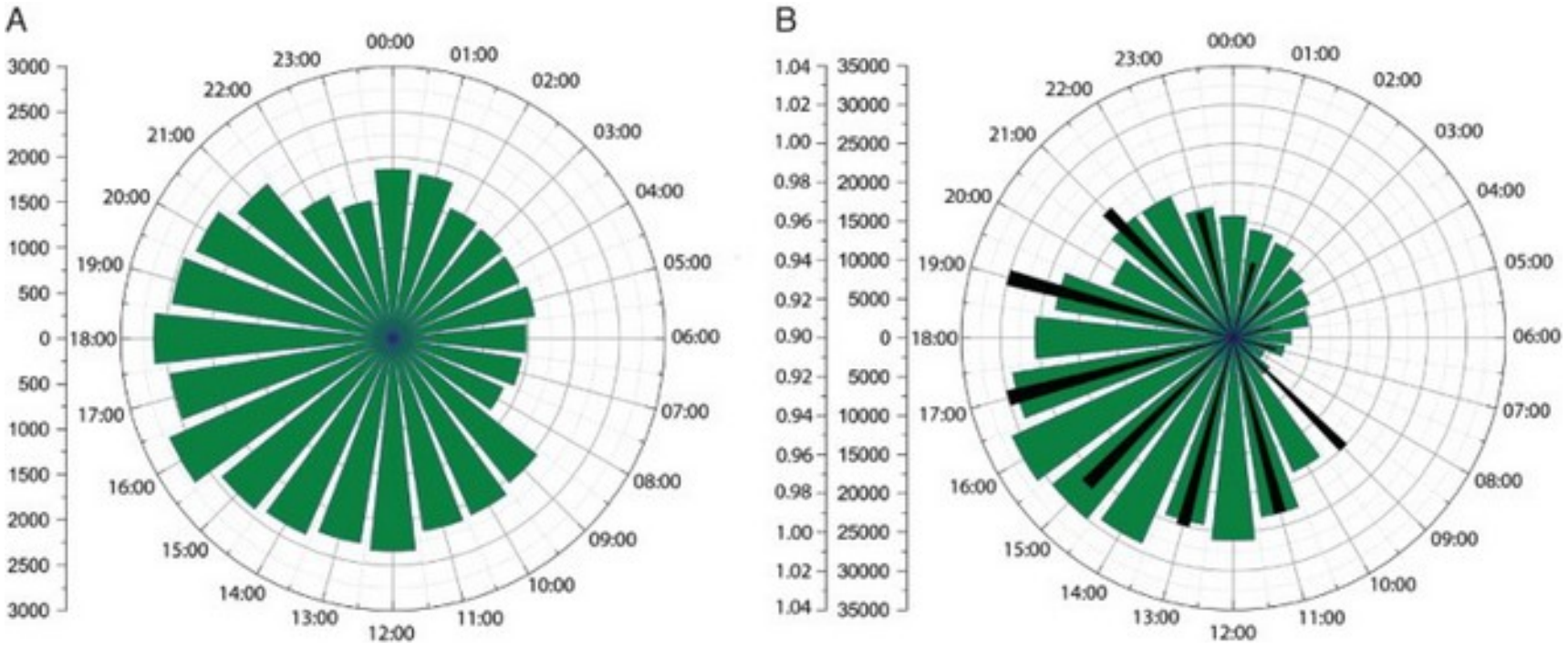
Reproducible results (1.7 Å – 2.1 Å)

2Fo-Fc *contoured* at 1.0 σ, Fo-Fc *contoured* at 3.0 σ



What do we gain from automation?

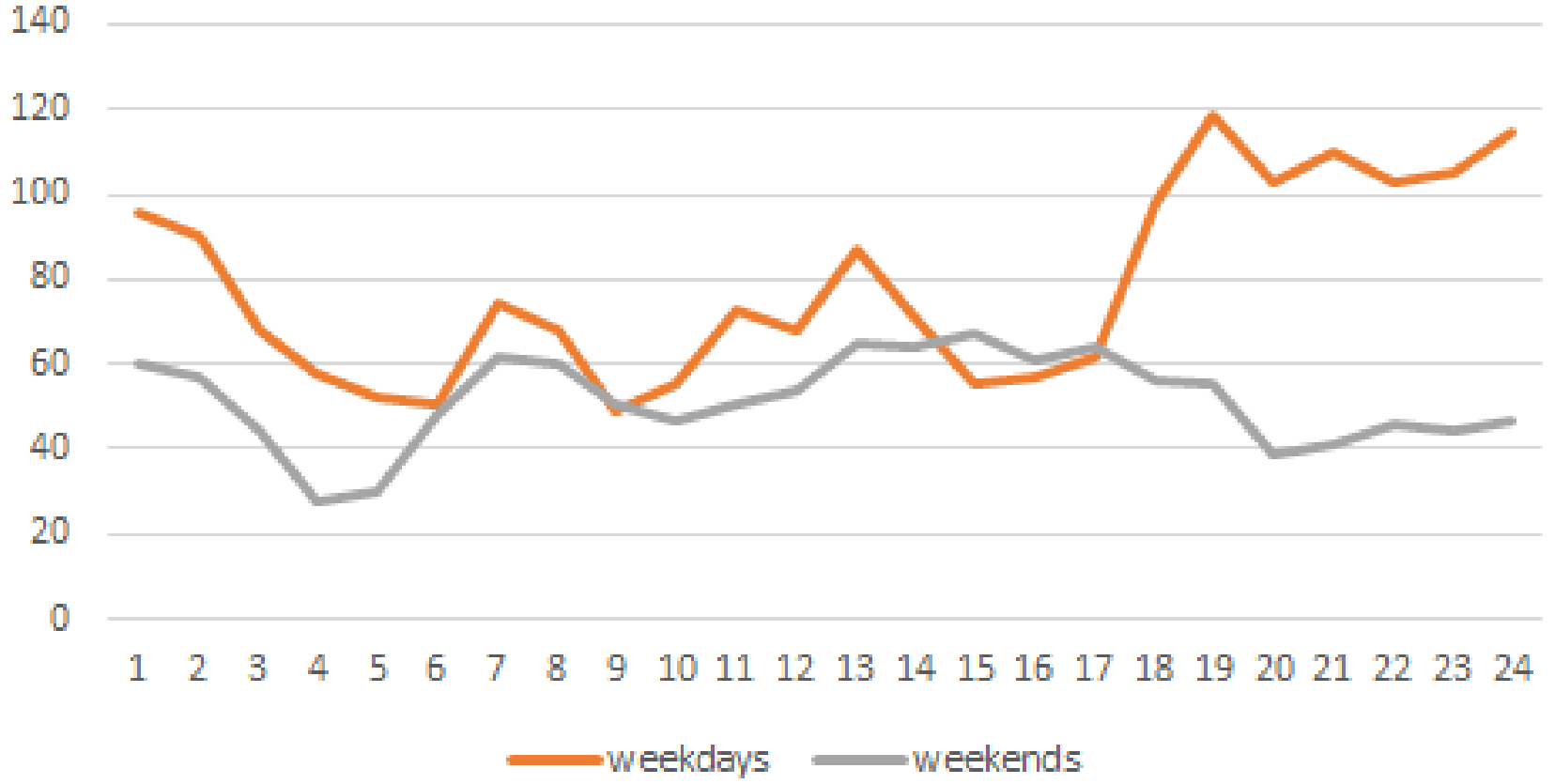
EMBL



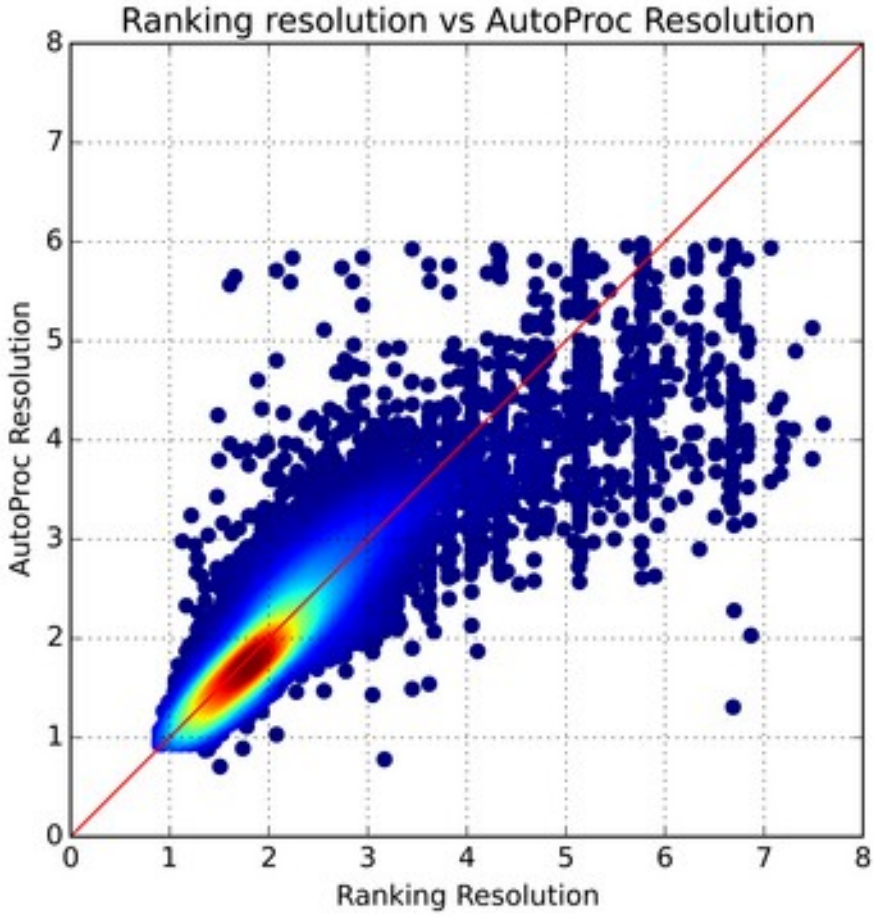
Bowler M.W**.**, Svensson, O and Nurizzo, D. (2016) *Cryst. Rev.*, **22**, 229-245

EMBL

MASSIF-1



EMBL

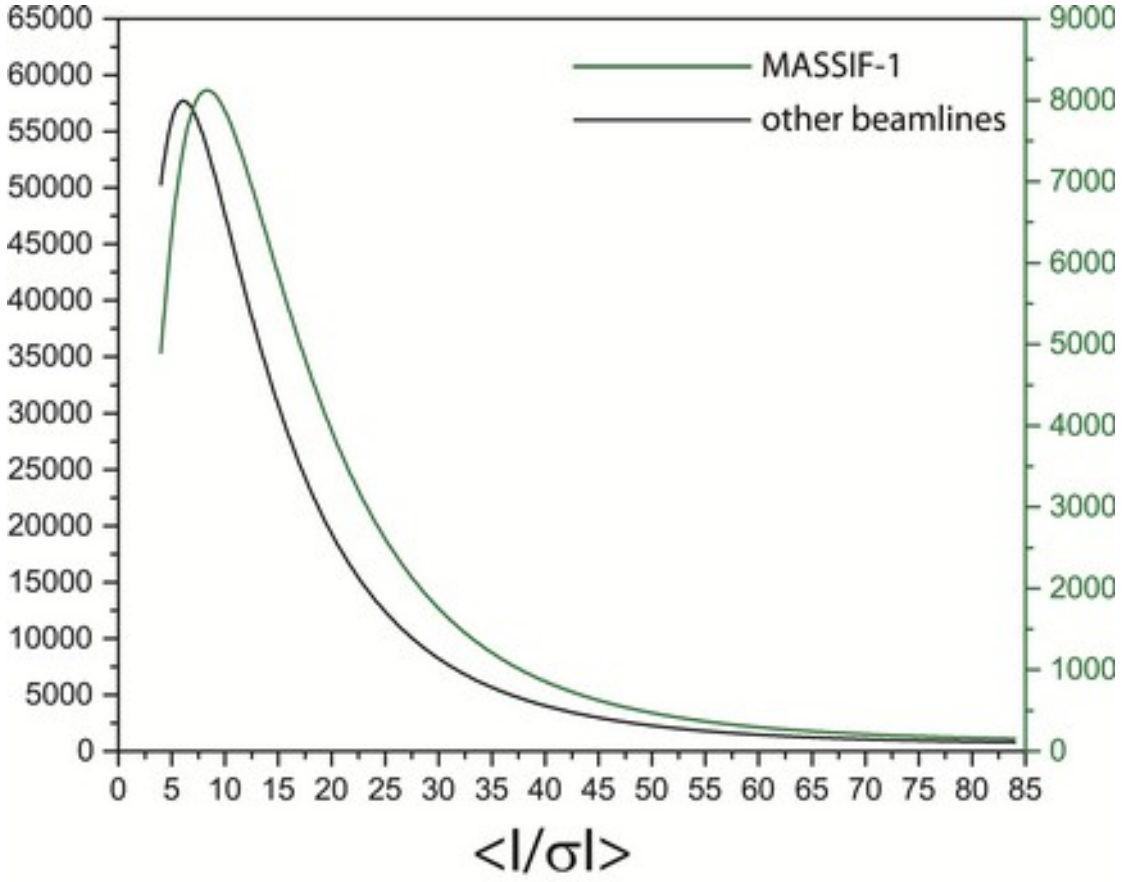


STRUCTURAL BIOLOGY

Svensson *et al.*

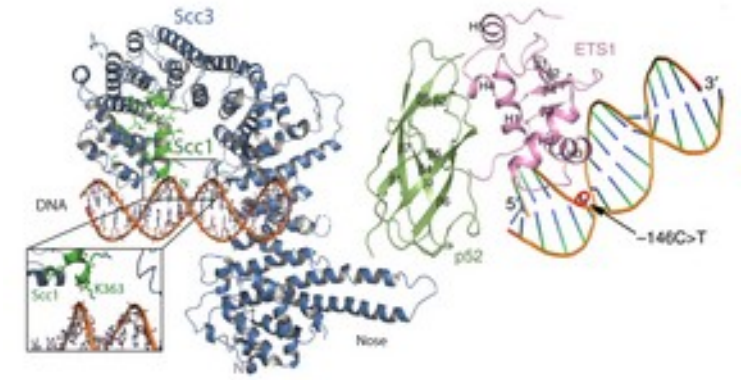
Volume 74 | **Part 5** | May 2018 | Pages 433–440 | 10.1107/S2059798318003728

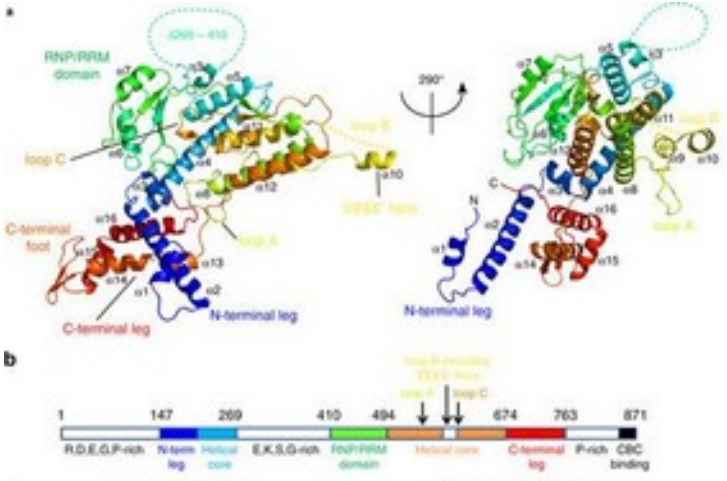
EMBL



Bowler M.W., Svensson, O and Nurizzo, D. (2016) *Cryst. Rev.*, **22**, 229-245

EMBL





Not just for lysozyme!

Scc3/Scc1 complex with DNA: – 600 crystals final 3.8 Å data set

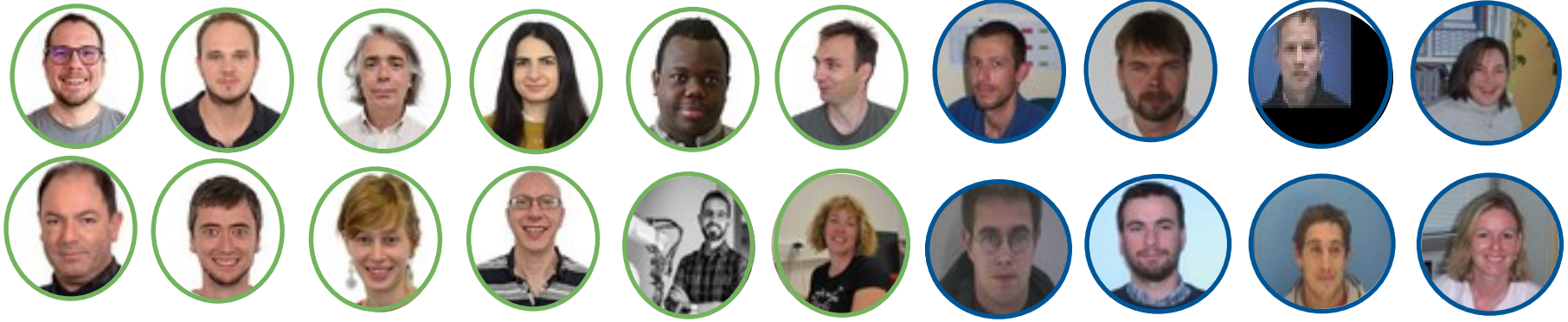
p52 and ETS1: 300 crystals final 3.0 Å data set

ARS2: 300 crystals Se-Met phasing at 4.0 Å

EMBL

EMBL

ESRF



ARINAX

GΦL

Global Phasing Limited

EMBL


EMBL
ESRF
A decade of fully automatic data collection at ESRF
5 December 2024
EPN Campus, Grenoble, France
Confirmed Speakers
Gerard Bricogne
Global Phasing Ltd., UK
Matthew Bowler
Katherine McAuley
Olof Svensson
Didier Nurizzo
Gilles Labesse
Serena Ricchio
Stephen Cusack
European Molecular Biology Laboratory, FR
Frank Kosielski
University College London, UK
Organisers
Matthew Bowler (EMBL)
Didier Nurizzo (ESRF)
Eleanor Ryan (ESRF)
Contact: 10symposium2024@esrf.fr