HKL-3000

Toward the Future of Structural Biology

Wladek Minor

Structural Biology
for the ~~high-throughput~~ high-output era

High throughput SB

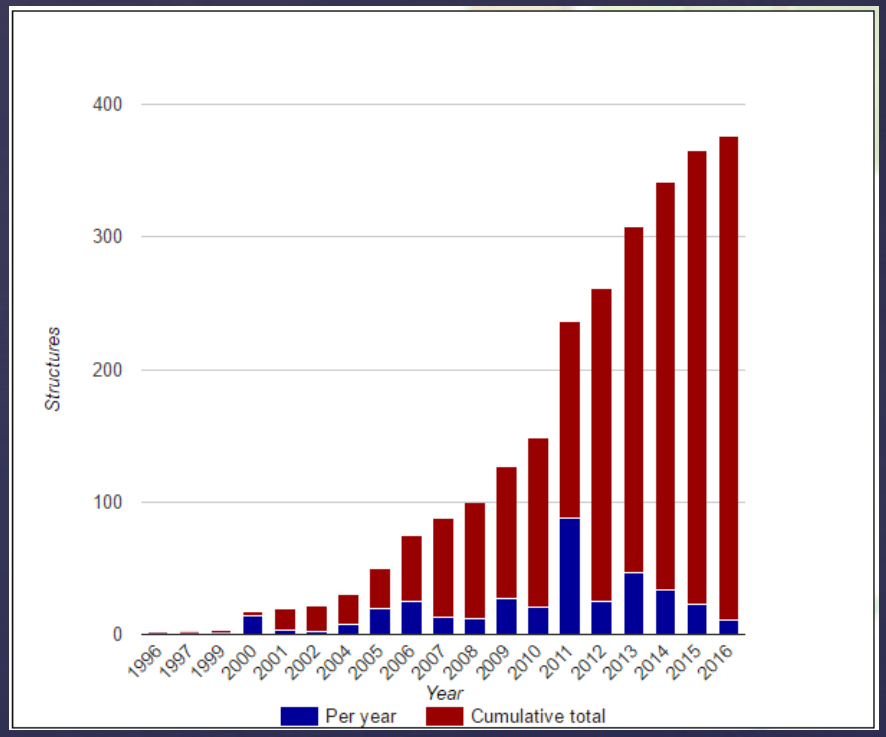
- Automatic cloning
- HT automatic expression
- HT automatic purification
- HT automatic crystallization
- HT automatic data collection
- HT automatic structure solution/refinement

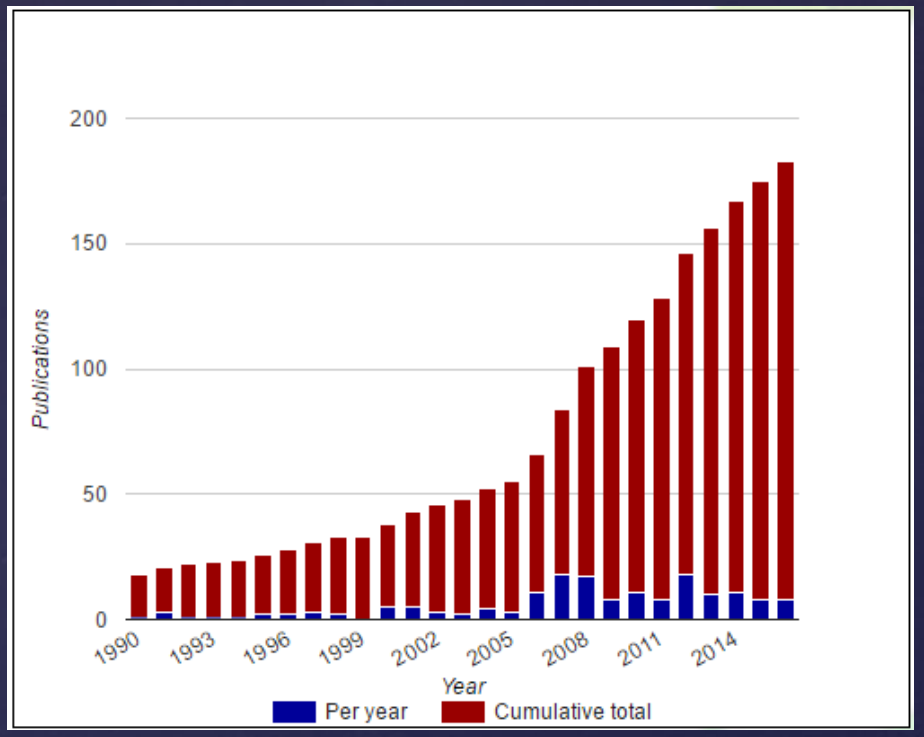
High throughput SB

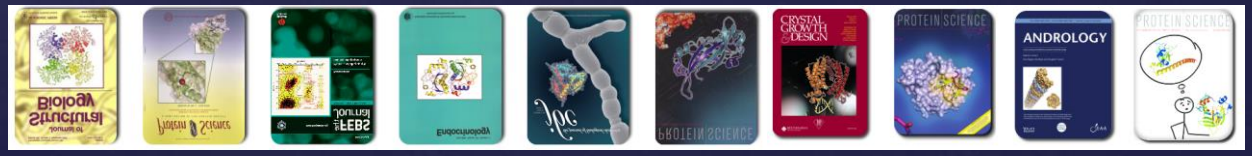
- Automatic cloning
- HT automatic expression
- HT automatic purification
- HT automatic crystallization
- HT automatic data collection
- HT automatic structure solution/refinement

- **Automatic paper writing**

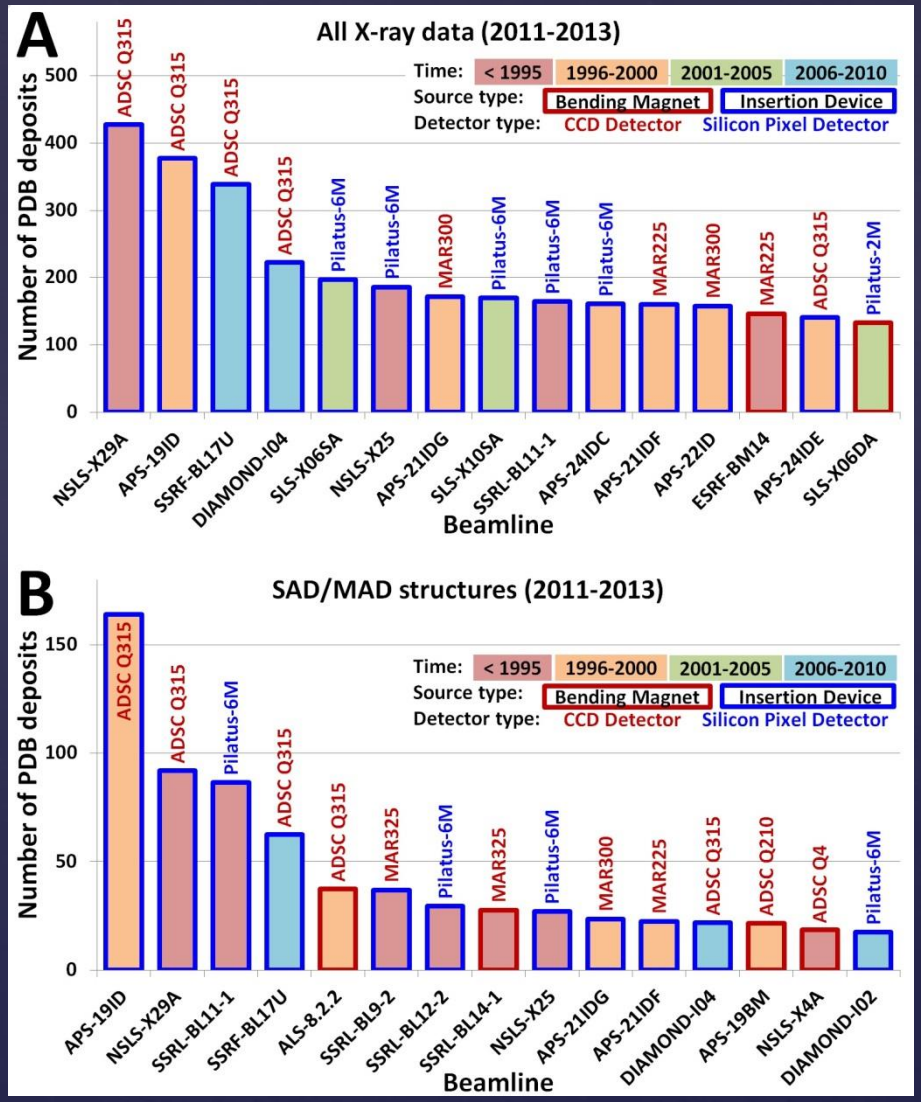
User perspective







Where we should collect data ?



Zheng *et al*, *Expert Opinion on Drug Discovery(2014) 9: 125-37*

Diffraction experiment - the last experiment before deposition to PDB

Dataset – 2minutes, sample change 2minutes -> 10minutes

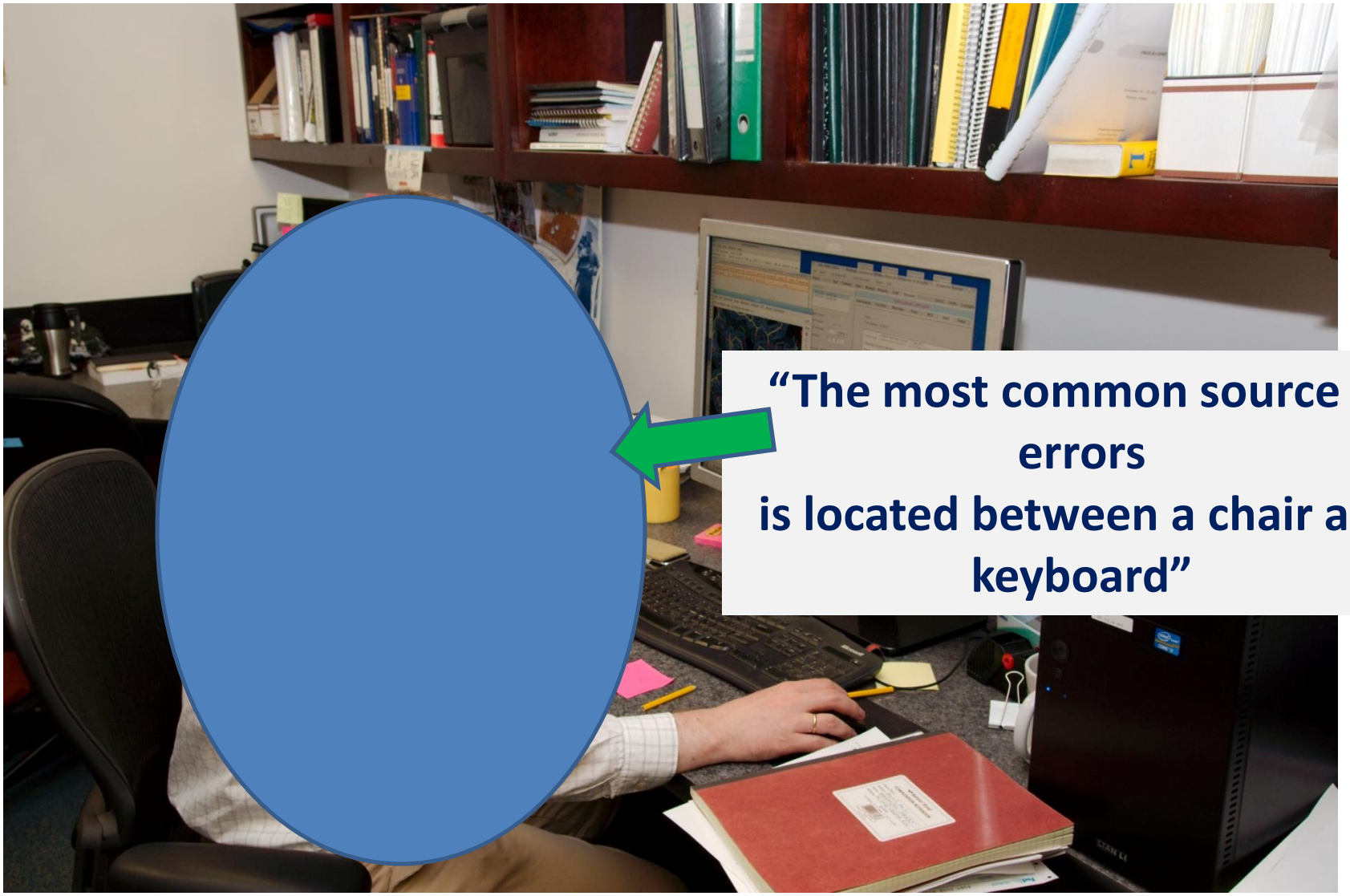
6 datasets/hour -> 144 datasets/day

180 days -> 25920 datasets/day -> 2.5 PDB

125 synchrotron stations -> 324 PDB

Efficiency -> 0.3%

User from a Programmer Point of View

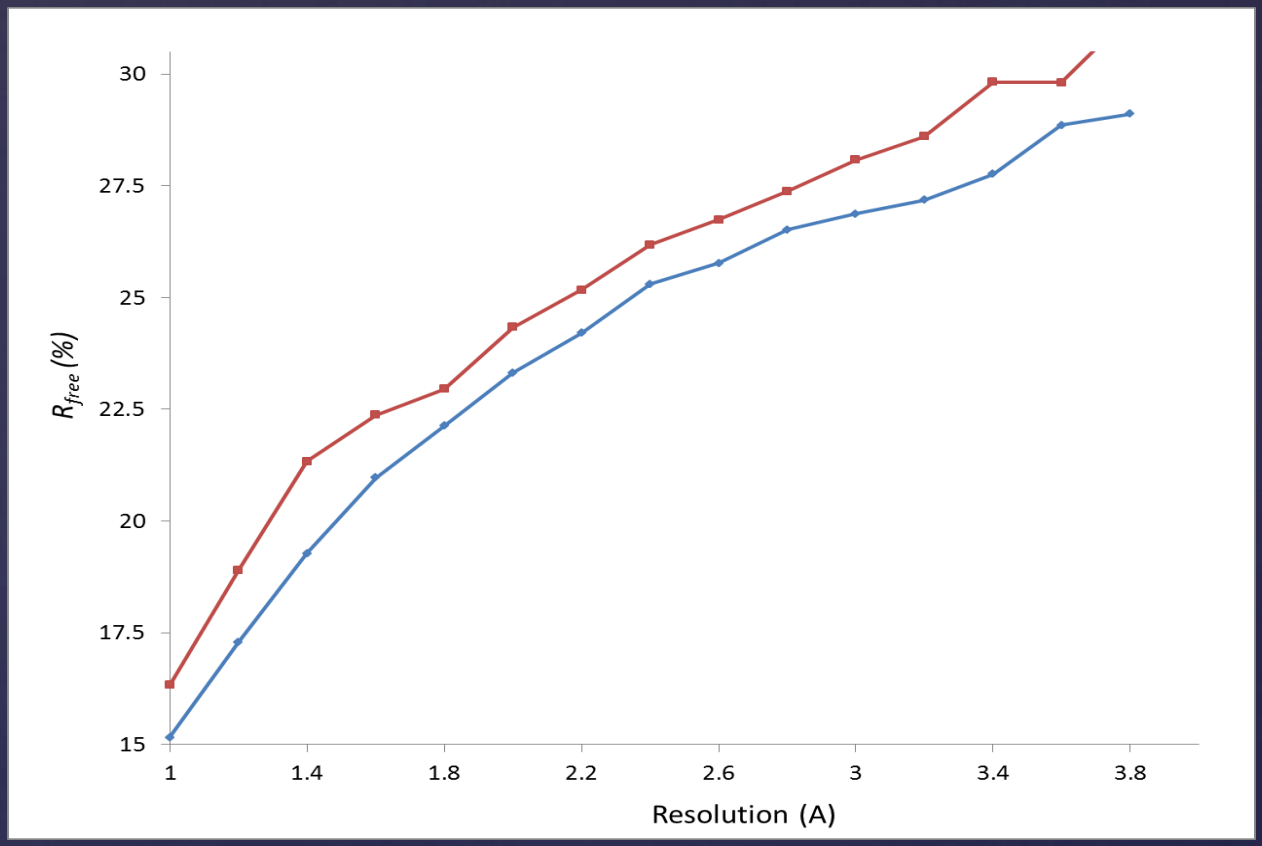


**“The most common source of errors
is located between a chair and keyboard”**

What experimenters know about data collection ?

```
REMARK   3  ESTIMATED OVERALL COORDINATE ERROR.
REMARK   3   ESU BASED ON R VALUE                            (A): NULL
REMARK   3   ESU BASED ON FREE R VALUE                       (A): NULL
REMARK   3   ESU BASED ON MAXIMUM LIKELIHOOD                 (A): NULL
REMARK   3   ESU FOR B VALUES BASED ON MAXIMUM LIKELIHOOD (A**2): NULL
REMARK   3
REMARK   3  RMS DEVIATIONS FROM IDEAL VALUES.
REMARK   3   DISTANCE RESTRAINTS.                      RMS    SIGMA
REMARK   3    BOND LENGTH                     (A) : NULL  ; NULL
REMARK   3    ANGLE DISTANCE                  (A) : NULL  ; NULL
REMARK   3    INTRAPLANAR 1-4 DISTANCE        (A) : NULL  ; NULL
REMARK   3    H-BOND OR METAL COORDINATION    (A) : NULL  ; NULL
REMARK   3
REMARK   3   PLANE RESTRAINT                  (A) : NULL  ; NULL
REMARK   3   CHIRAL-CENTER RESTRAINT       (A**3) : NULL  ; NULL
REMARK   3
REMARK   3   NON-BONDED CONTACT RESTRAINTS.
REMARK   3    SINGLE TORSION                  (A) : NULL  ; NULL
REMARK   3    MULTIPLE TORSION                (A) : NULL  ; NULL
REMARK   3    H-BOND (X...Y)                  (A) : NULL  ; NULL
REMARK   3    H-BOND (X-H...Y)                (A) : NULL  ; NULL
REMARK   3
REMARK   3   CONFORMATIONAL TORSION ANGLE RESTRAINTS.
REMARK   3    SPECIFIED                 (DEGREES) : NULL  ; NULL
REMARK   3    PLANAR                    (DEGREES) : NULL  ; NULL
REMARK   3    STAGGERED                 (DEGREES) : NULL  ; NULL
REMARK   3    TRANSVERSE                (DEGREES) : NULL  ; NULL
REMARK   3
REMARK   3  ISOTROPIC THERMAL FACTOR RESTRAINTS.     RMS    SIGMA
REMARK   3   MAIN-CHAIN BOND                (A**2) : NULL  ; NULL
REMARK   3   MAIN-CHAIN ANGLE               (A**2) : NULL  ; NULL
REMARK   3   SIDE-CHAIN BOND                (A**2) : NULL  ; NULL
```

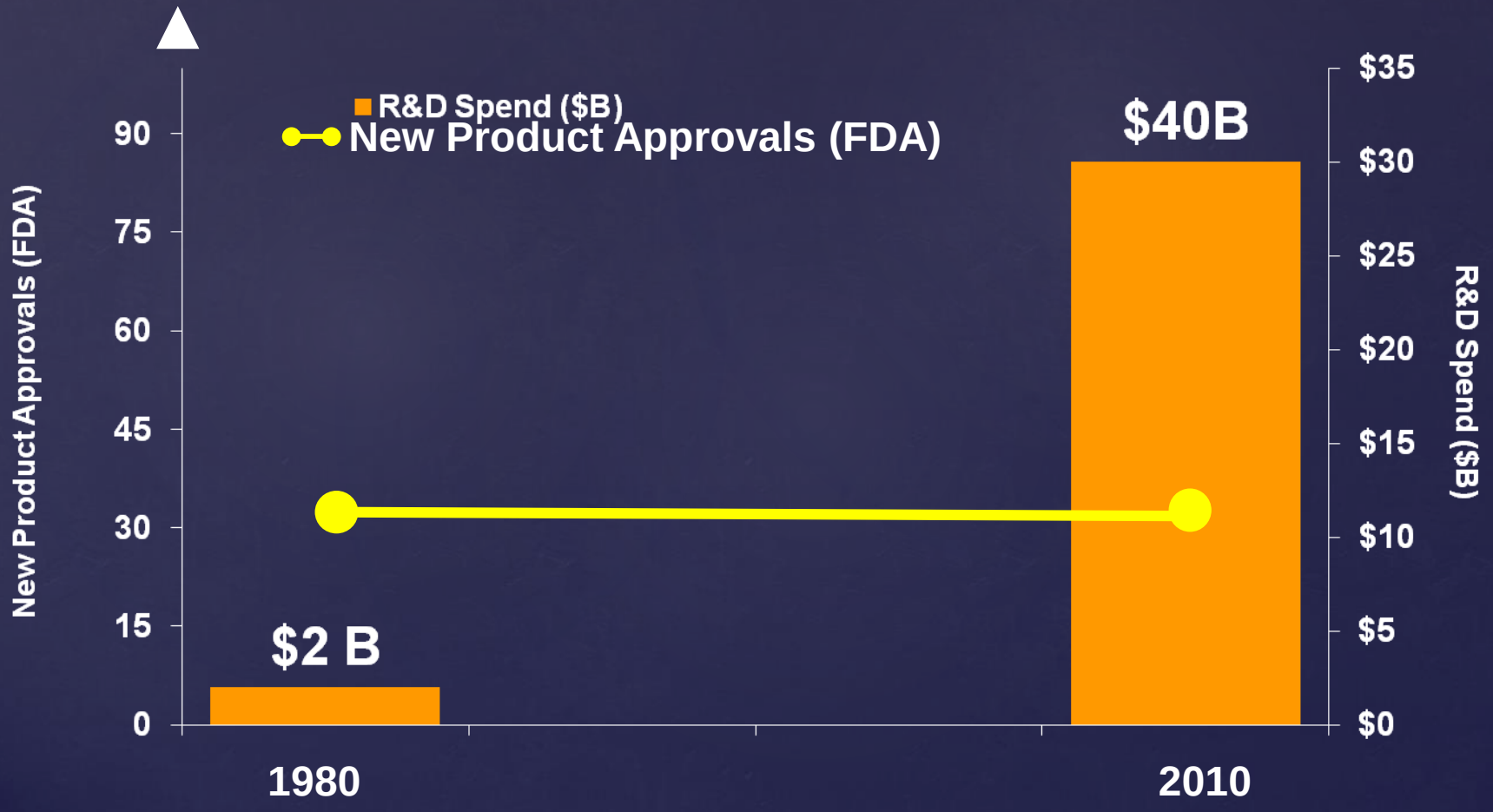
Unexpected correlation?



Average R_{free} by resolution bin (with a width of 0.2 Å for X-ray crystallography PDB structures deposited after January 1, 2001, divided into two groups by the number of missing data items ("NULLs") in the PDB file. The means for "high-completion" deposits (20 NULLs or less) are shown in blue, and the means for "low-completion" deposits (50 or more NULLs) are shown in red.

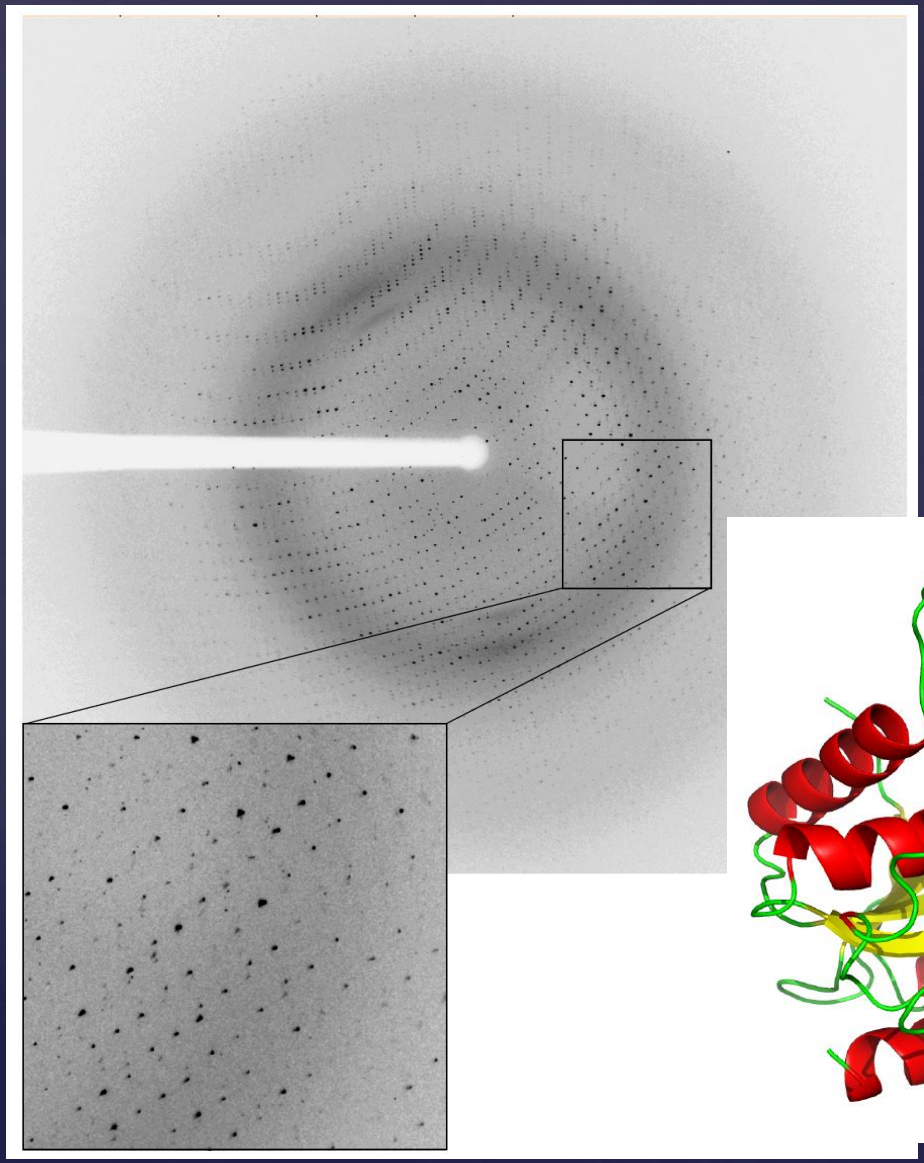
J Struct Funct Genomics. 2016 Mar;17(1):1-16.

Industry R&D Spend Compared to New Product Approvals

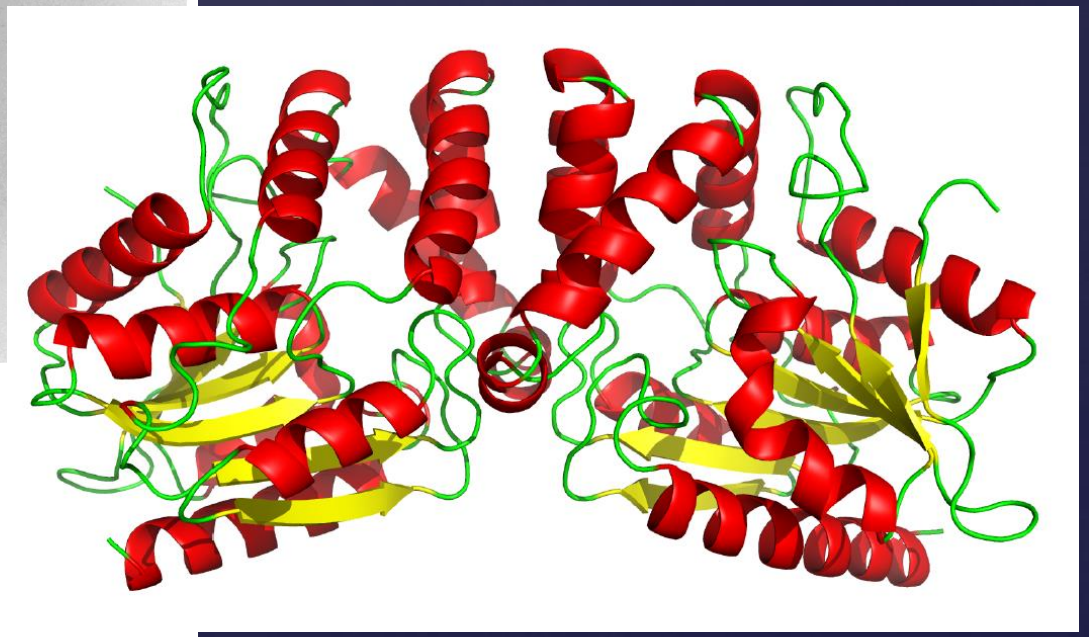


T.Skarzynski – Agilent, *Glaxo*

… number of crystals



RutD from *E. coli*



Data collection

- **What is beneficial for particular type of experiment (MR, SAD, MAD, Ligand screening) ?**
- **What is possible ?**

Examples:

- **Wide sectors *vs.* narrow sectors**
- **Inverse beam experiment**
- **Crystal orientation**

Theory *vs.* Experiment

In theory,

there is no difference between
theory and practice.

Theory *vs.* Experiment

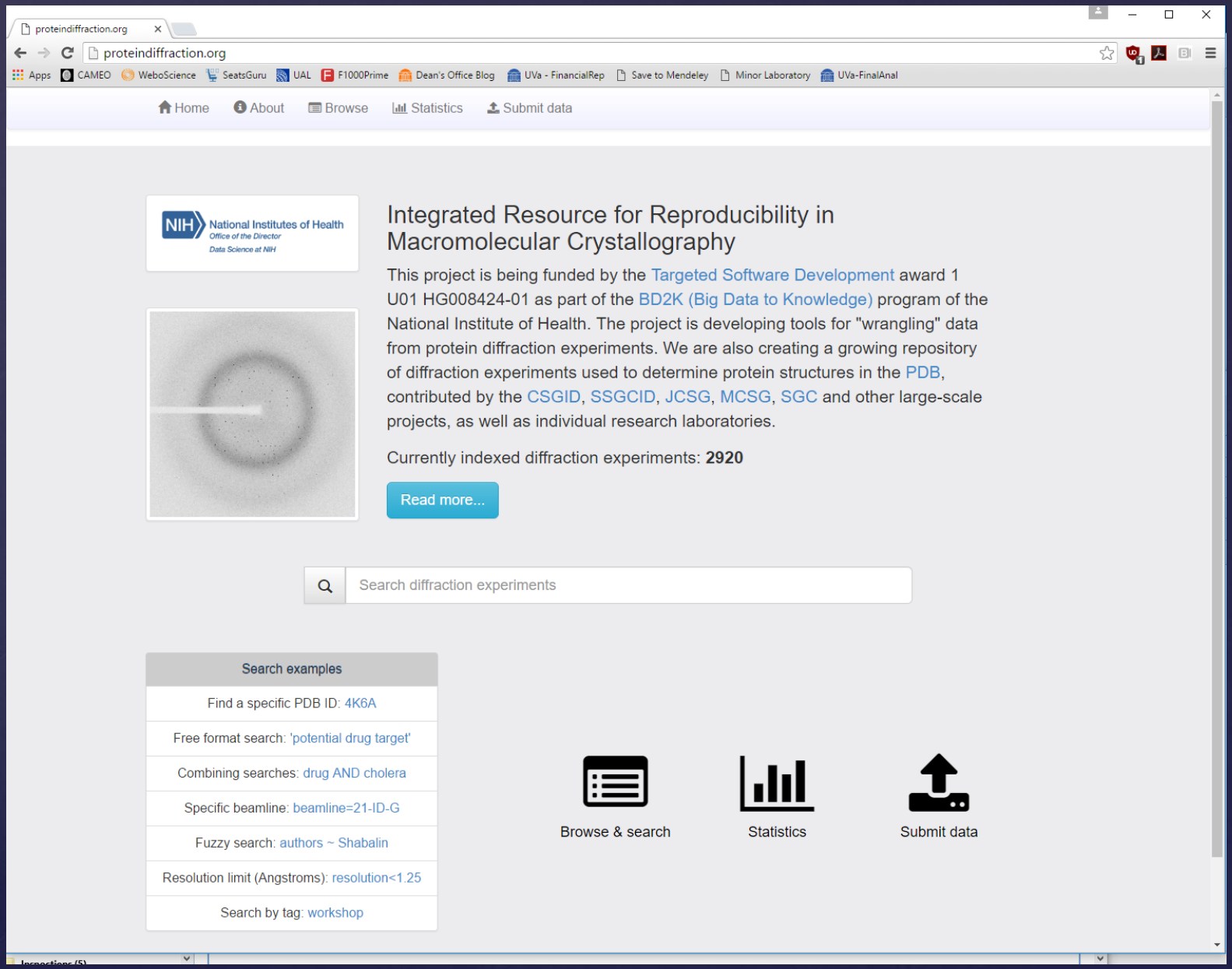
In theory,

there is no difference between
theory and practice.

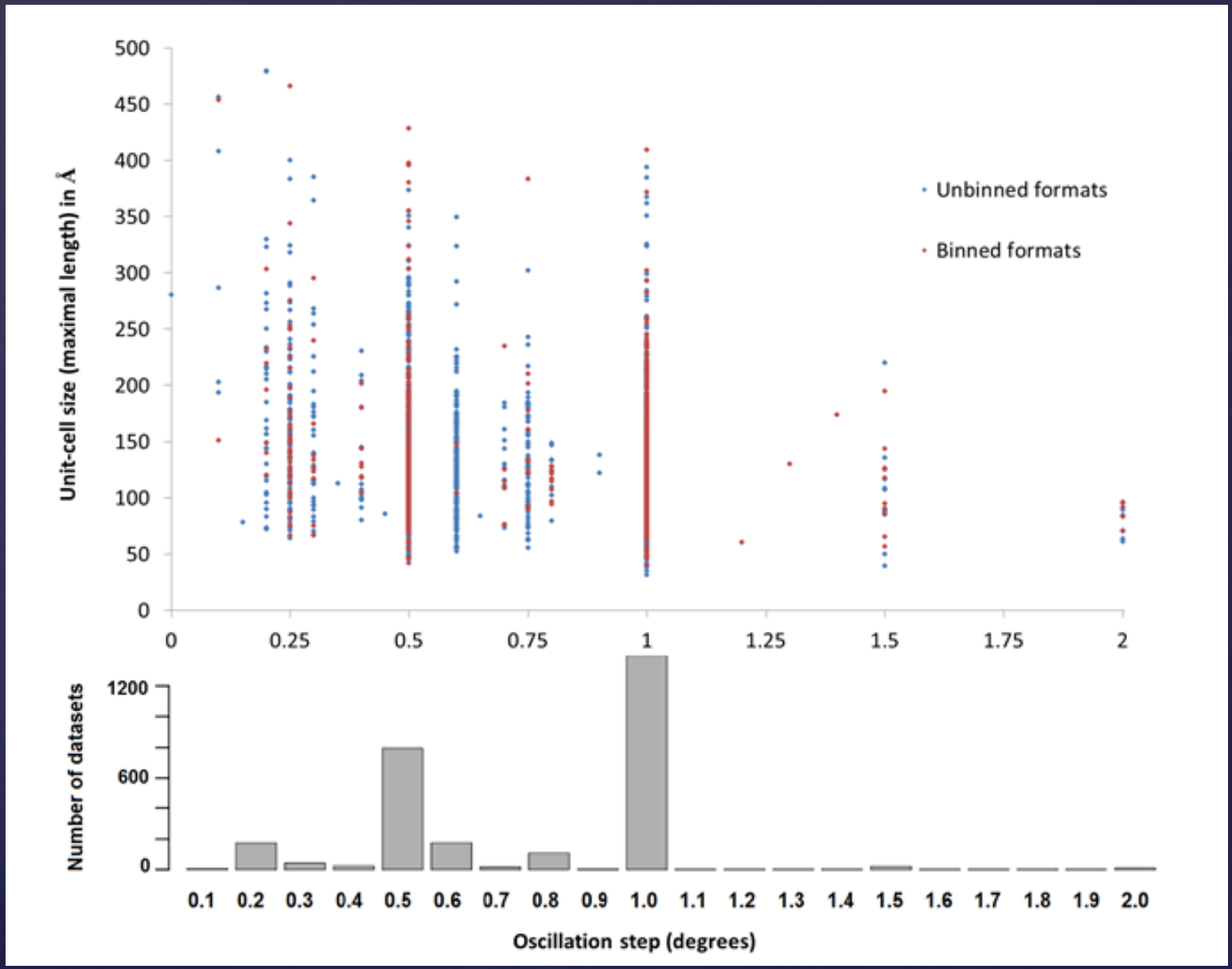
But, in practice, there is.

Jan L.A. van de Snepscheut

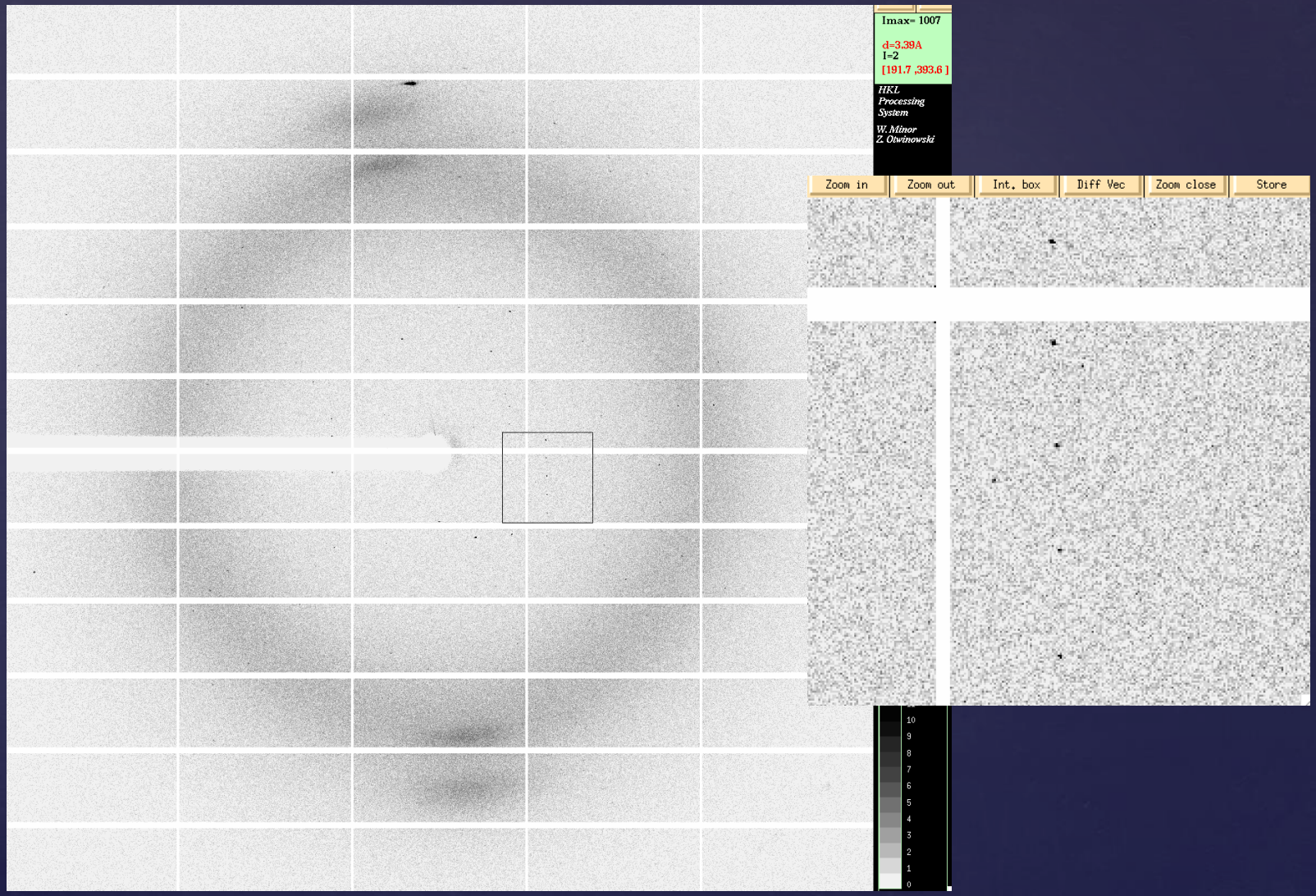
http://www.proteindiffraction.org



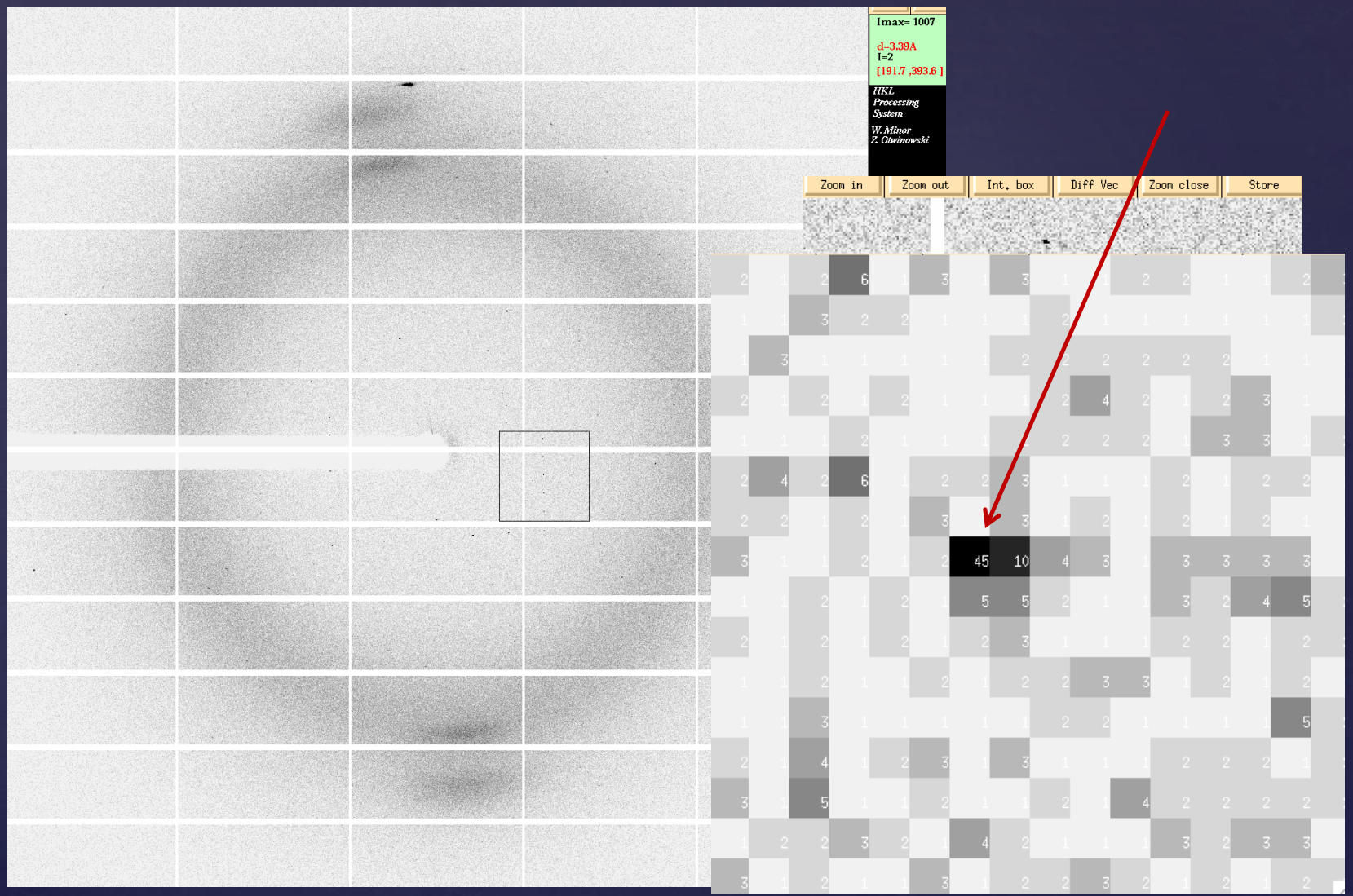
Data collection strategy



Optimal data collection ?



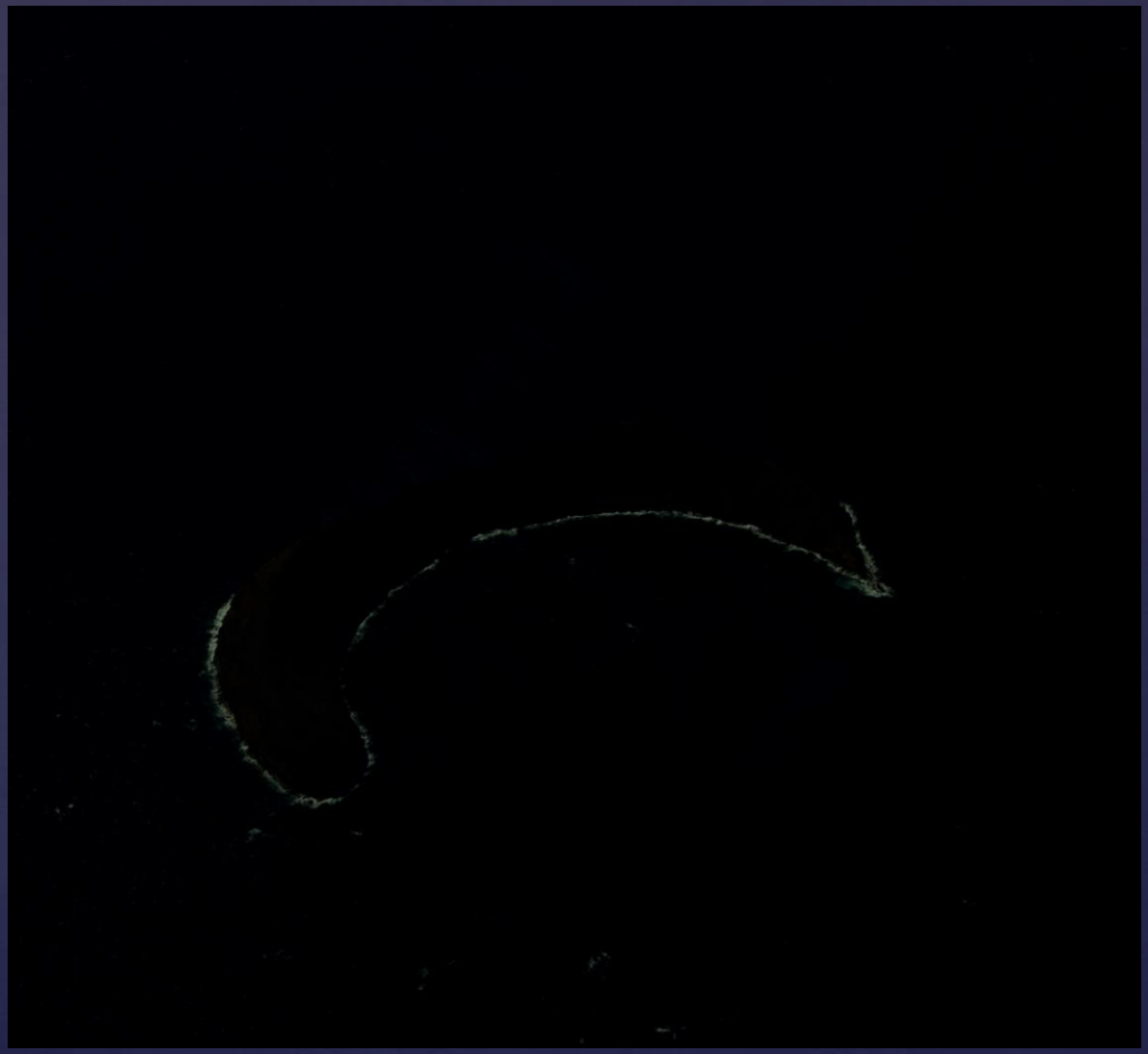
Optimal data collection ?



Header – is CBF header a MAH ?

```
# 2015/May/06 10:30:40
# Pixel_size 172e-6 m x 172e-6 m
# Silicon sensor, thickness 0.001 m
# Oscillation_axis omega
# Excluded_pixels:  badpix_mask.tif
# Chi 0.0000 deg.
# Angle_increment 0.1000 deg.
# Polarization 0.99
# file_comments
# N_oscillations 2500
# Beam_xy (1223.03, 1256.56) pixels
# Exposure_time 0.020000 s
# Phi 0.0020 deg.
# Energy_range (0, 0) eV
# Start_angle 160.6000 deg.
# Detector_distance 0.617619 m
# Detector_Voffset 0.0000 m
# Alpha 0.0000 deg.
# Flat_field: (nil)
# Threshold_setting 7619 eV
# Exposure_period 0.020950 s
# N_excluded_pixels: = 321
# Kappa 0.0020 deg.
# Tau = 0 s
```

Do you like this image?



Do you like this image ?

How expensive is bright lens ?

See more choices

Canon EF 85mm f1.2L II USM Lens for Canon DSLR Cameras - Fixed
by Canon

$1,999.00 Prime
Get it by **Monday, Aug 24**

More Buying Choices
$1,999.00 new (22 offers)
$1,499.99 used (24 offers)

Trade-in eligible for an Amazon gift card

★★★★★ ▾ 159

See Style Options

Canon EF 85mm f/1.8 USM Medium Telephoto Lens for Canon SLR Cameras - Fixed
by Canon

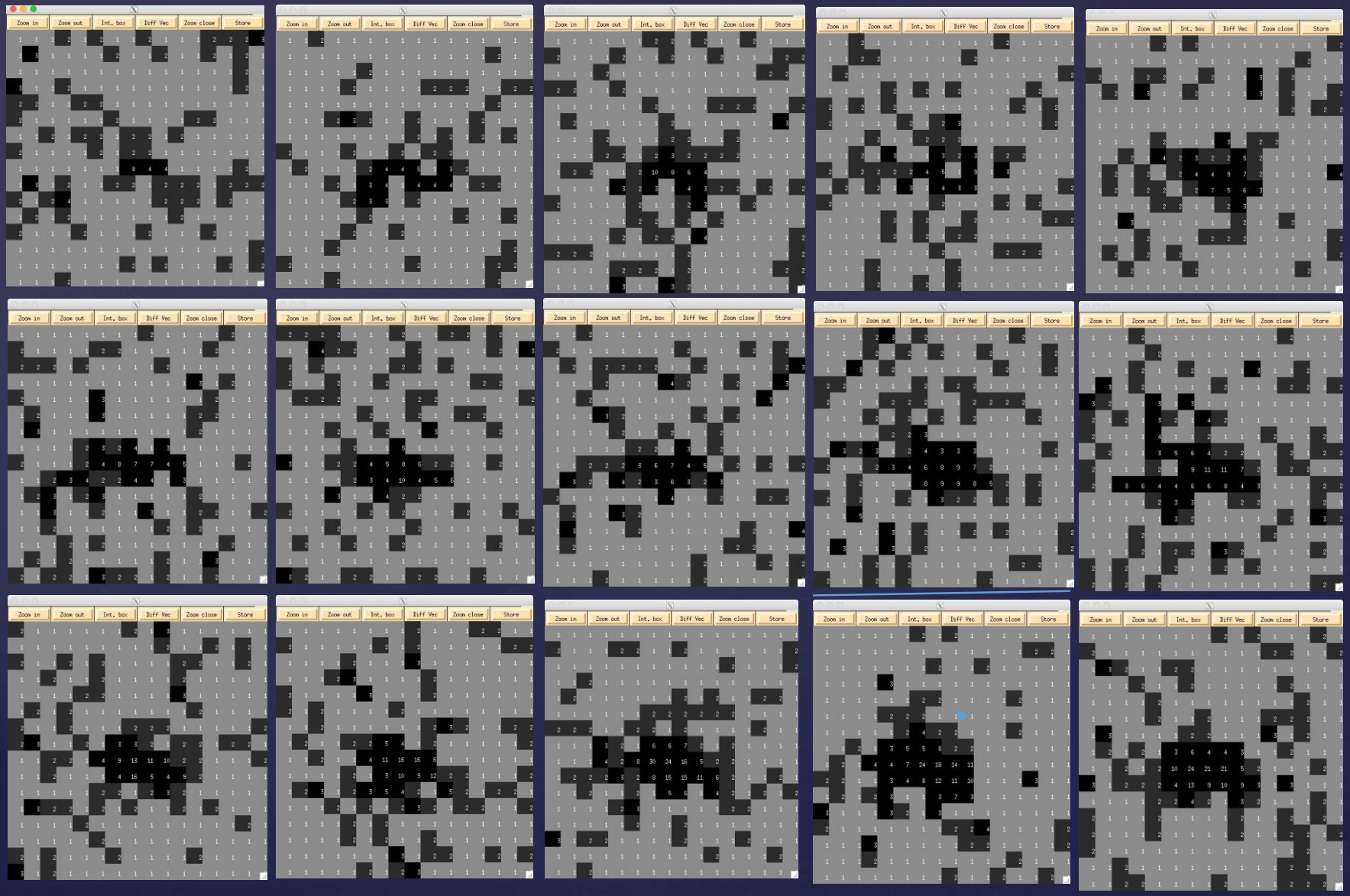
$369.00 Prime
Get it by **Monday, Aug 24**

More Buying Choices
$369.00 new (27 offers)
$298.00 used (26 offers)

Trade-in eligible for an Amazon gift card

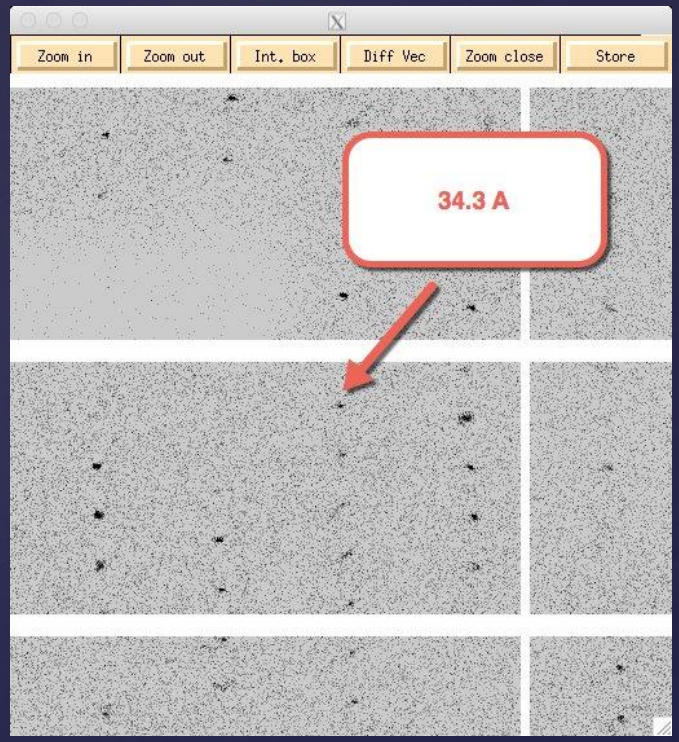
★★★★★ ▾ 770

Thin slicing, 0.01s/frame



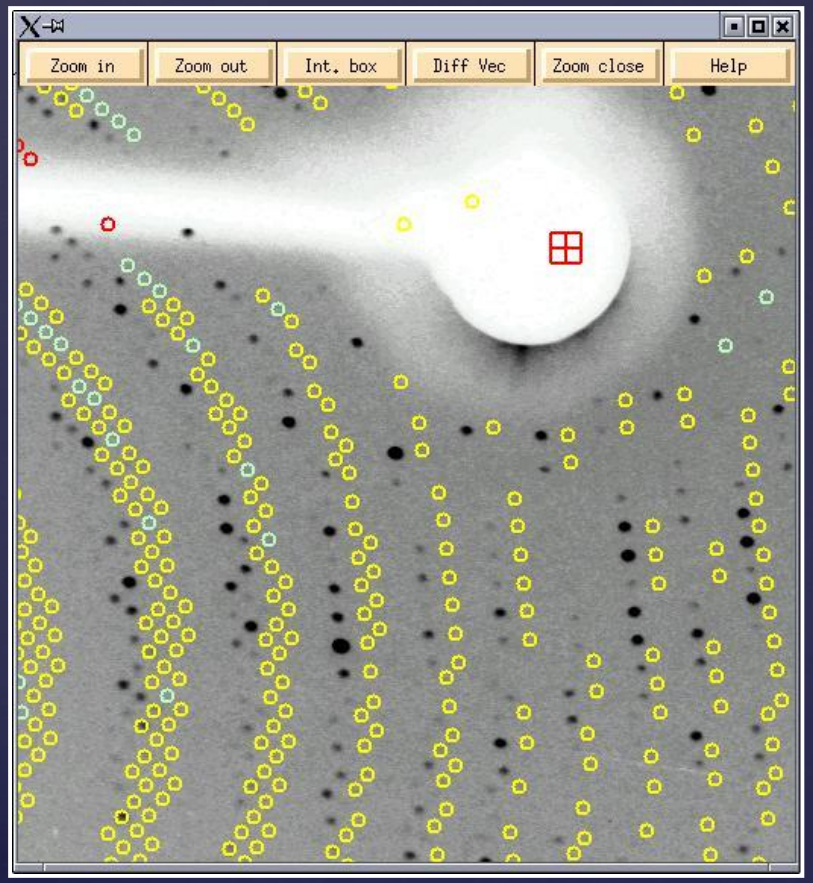
VS, S, W, VW


/Users/wladek/Desktop/Zak1/L242/Lo8242_2_00101.cbf
Prof fit R
Zoom wind
Write/Print
A/D test
Floor Up
Floor Down
reverse
dim
bright
color
Update pred
Full scale
GO
Show Overfl
Peak Sear
Edit P.S.
close
Frame
Dead pixel
Save Dead pixel
Help
Imax=53012
d=2.75A
I=1
[3.2 ,379.9]
HKL
Processing
System
W. Minor
Z. Otwinowski
7
6
5
4
3
2
1
0

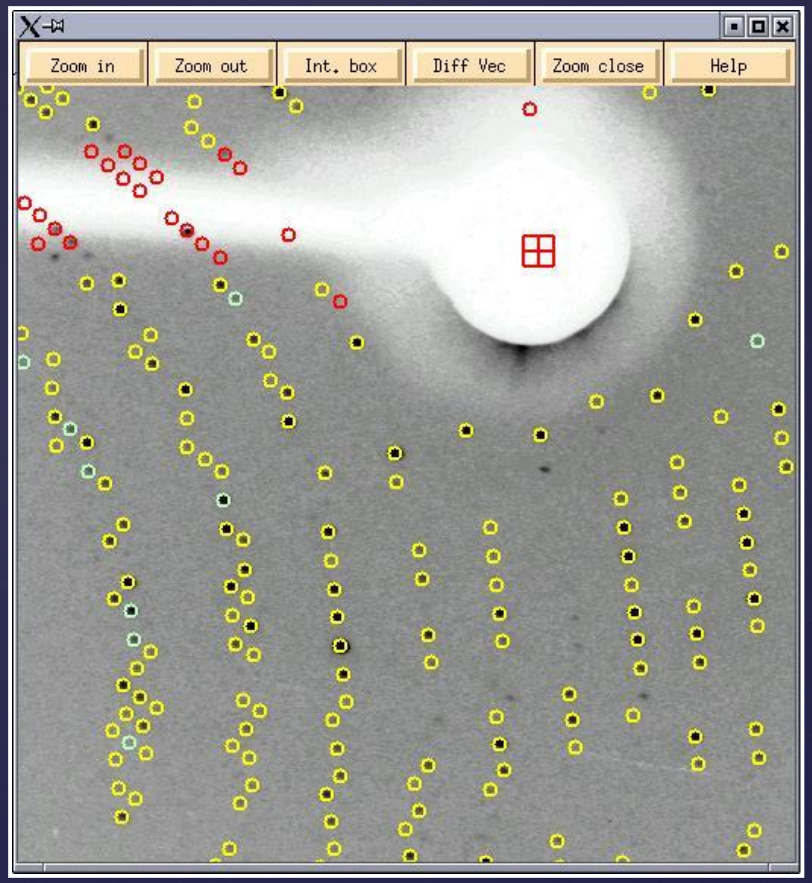

Zoom in
Zoom out
Int. box
Diff Vec
Zoom close
Store
34.3 A

Xbeam, Ybeam

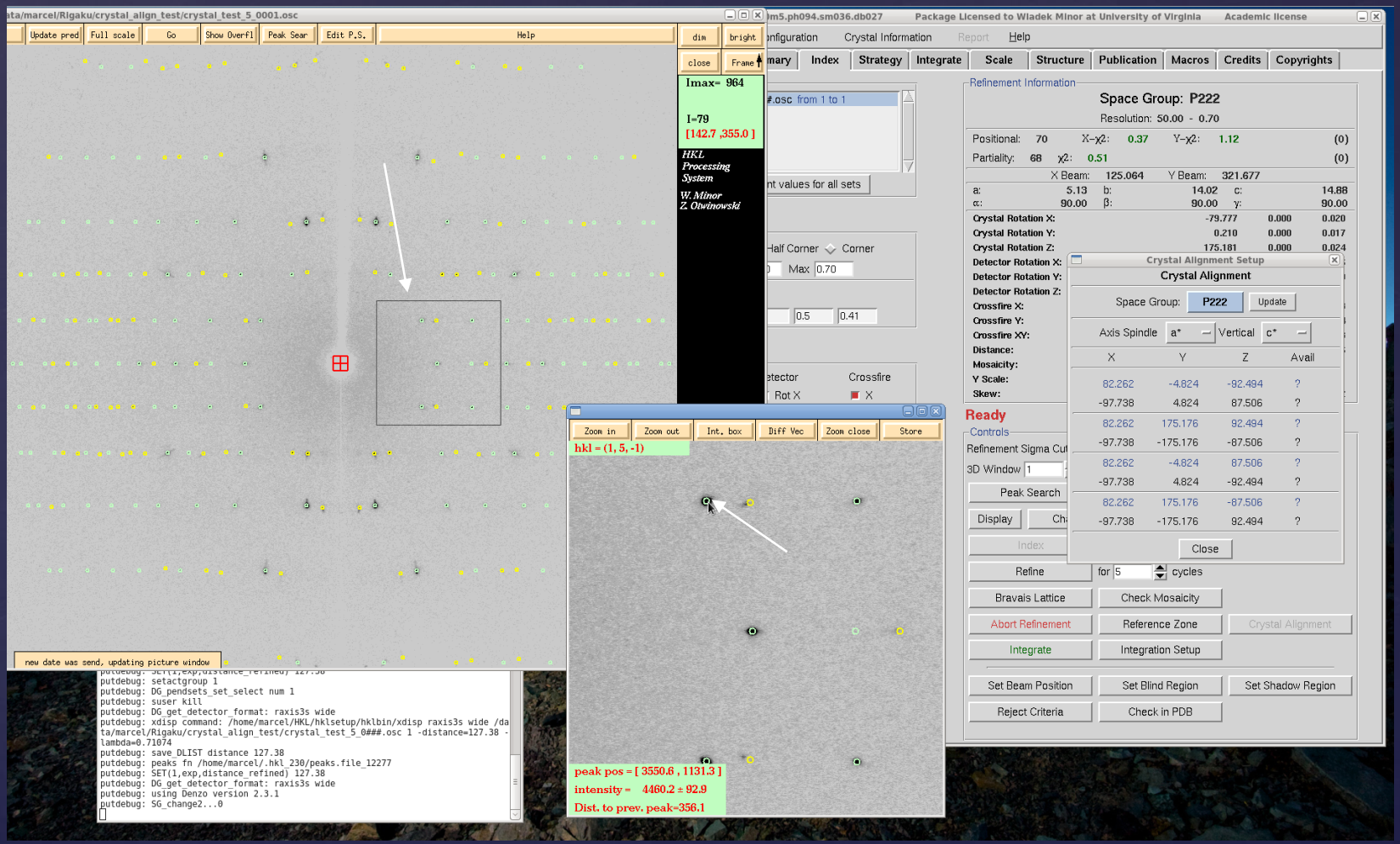
150.6 151.2



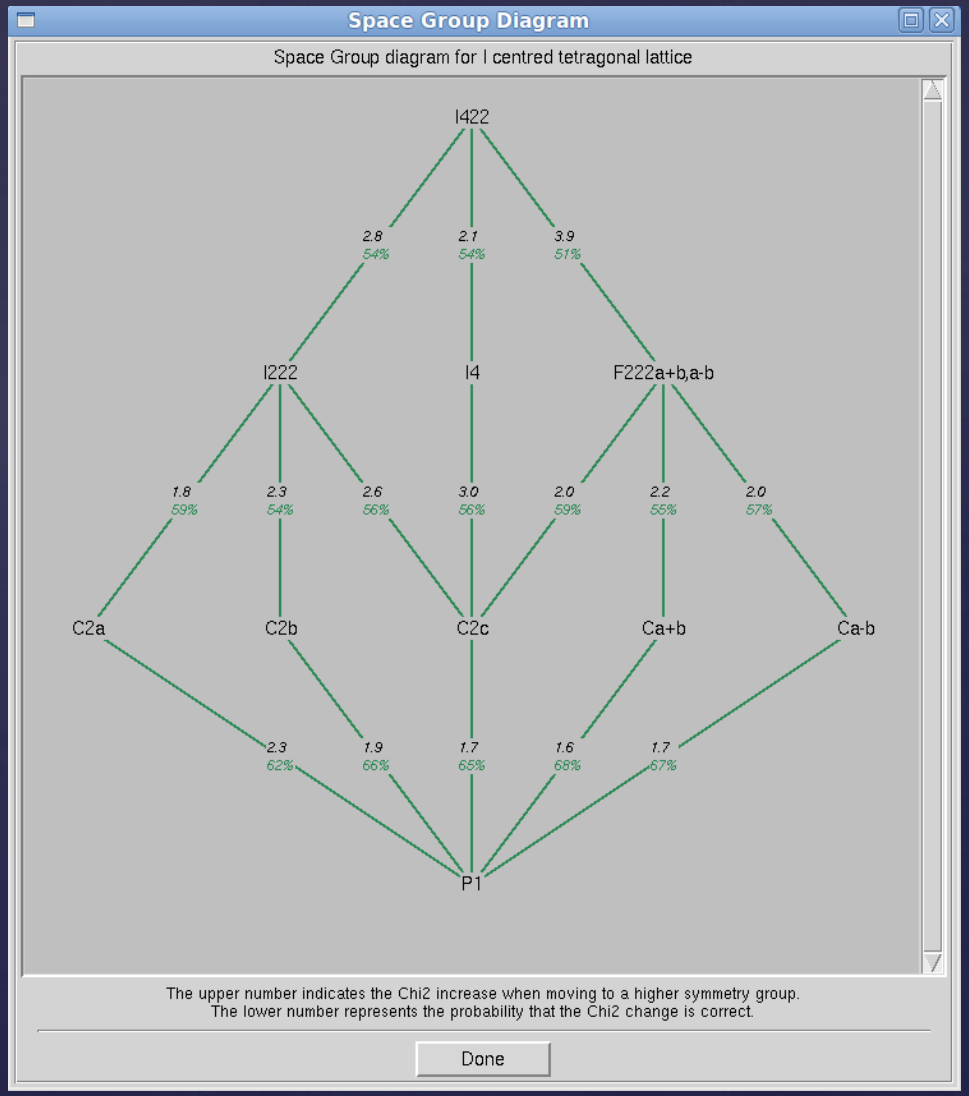
150.0 150.0



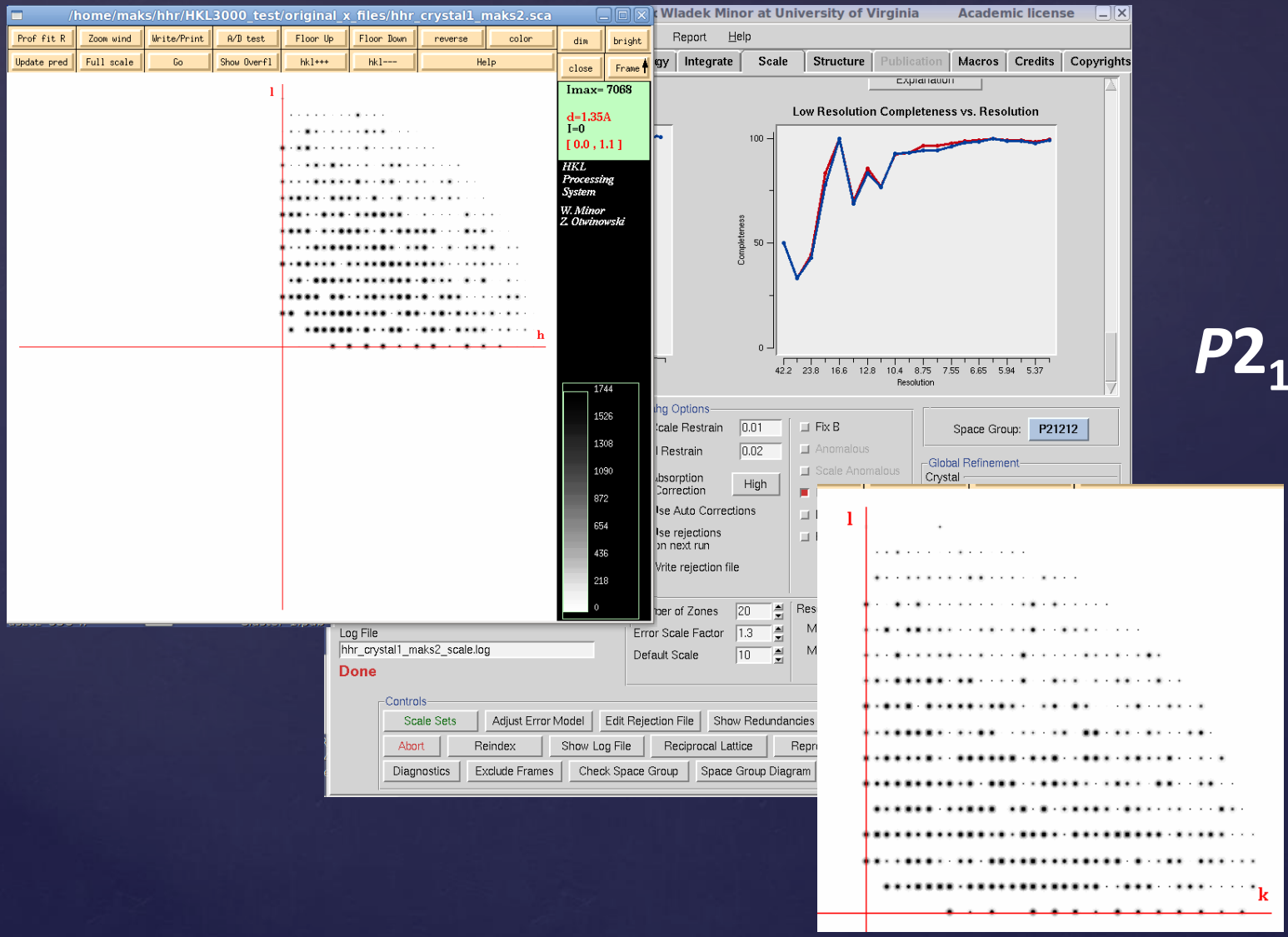
Crystal orientation



Space Group Determination
Laue class diagram

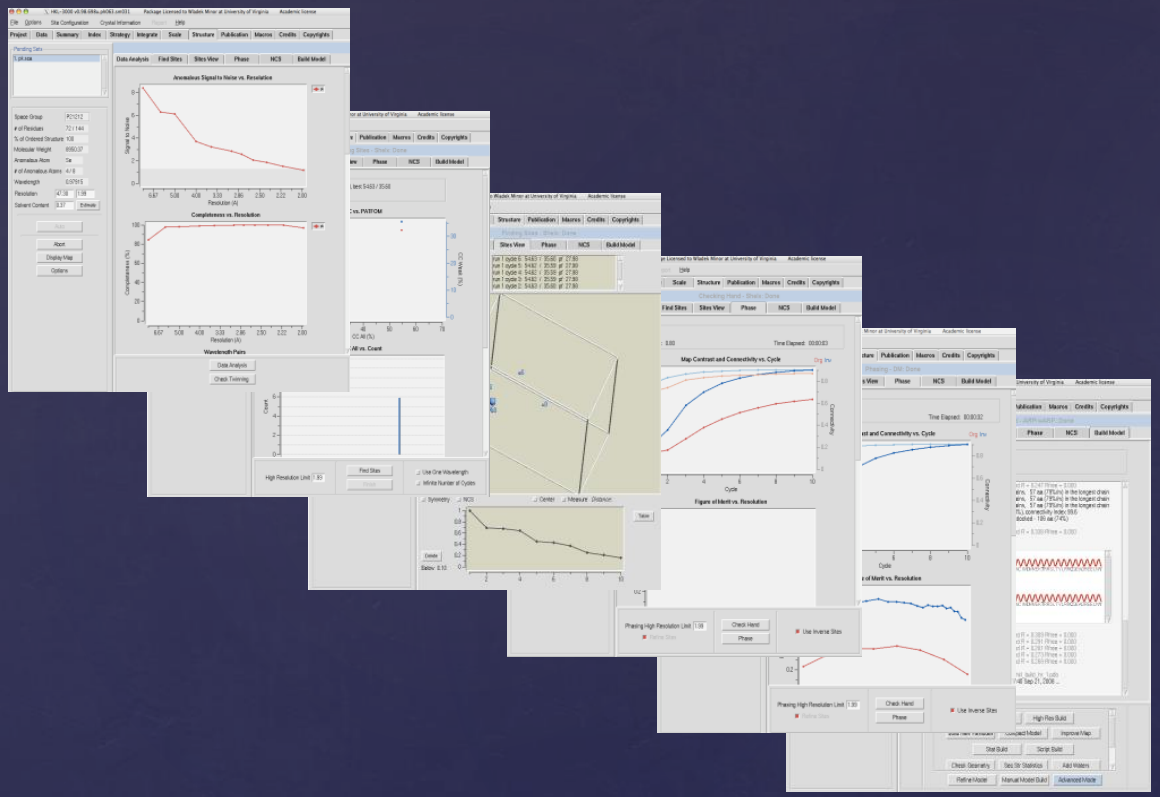


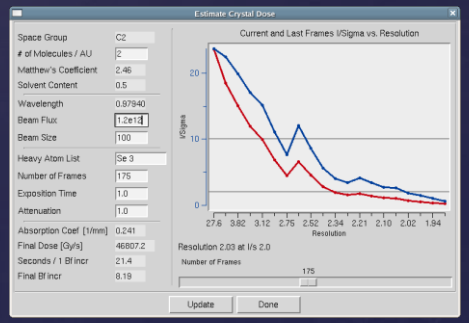
Space Group Determination



$P2_12_12$

Where are the real bottlenecks ?

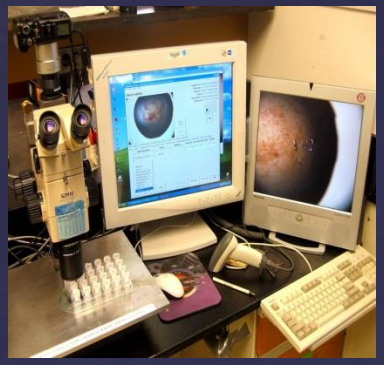




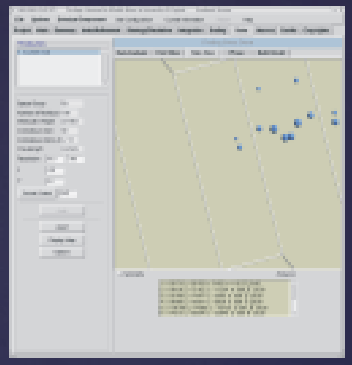
Database-controlled pipeline

lab e-book

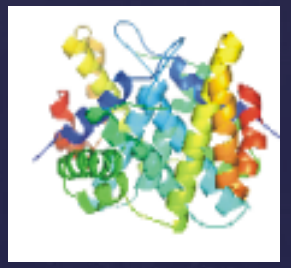
FPLC



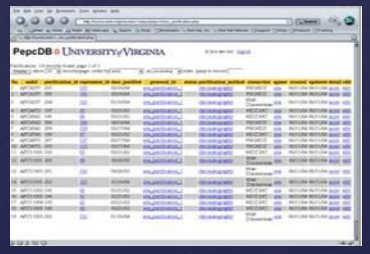
EQUIPMENT

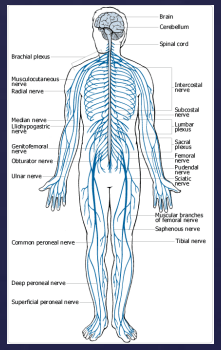


HKL-3000



DATABASES





Big brother?

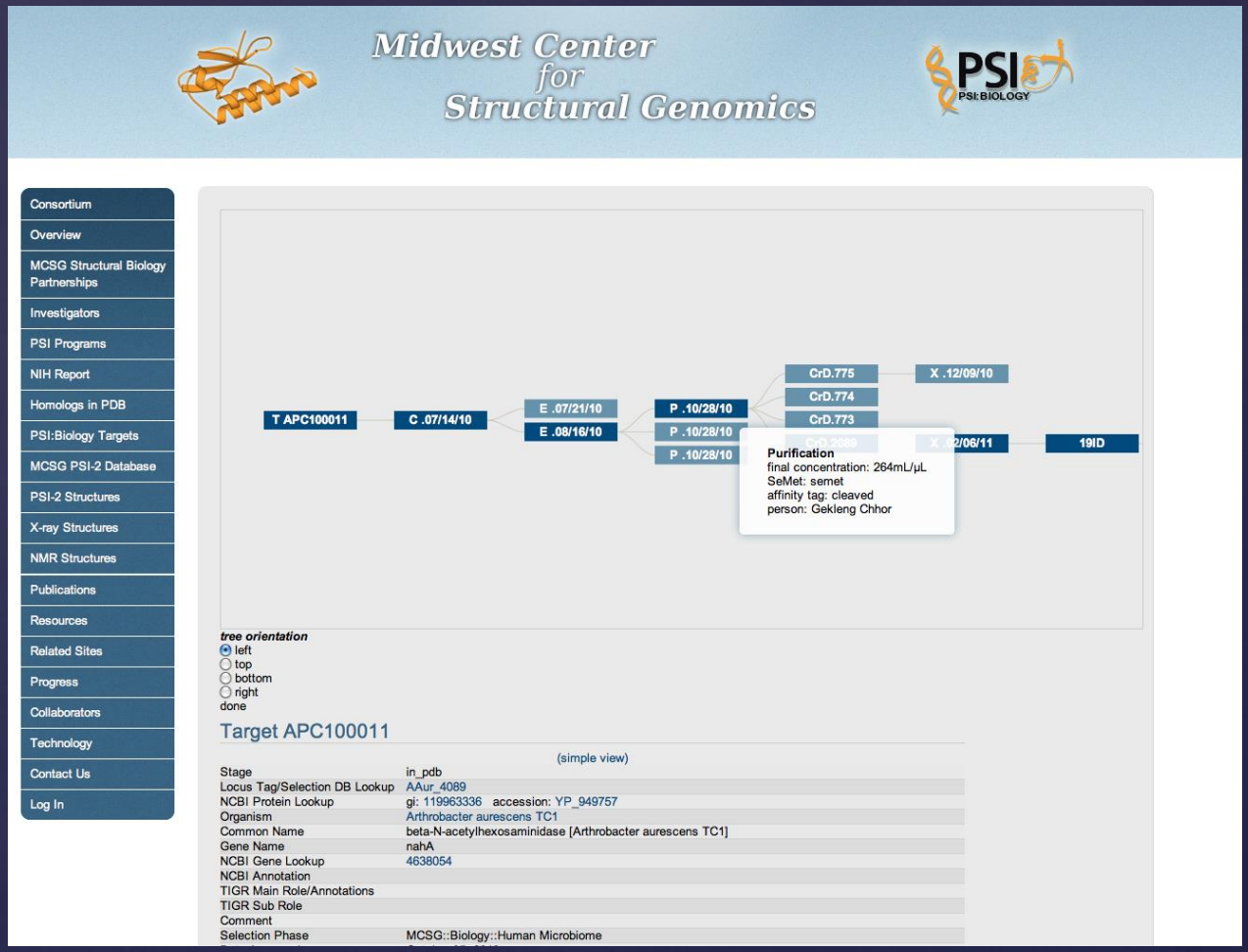
Statistics / Progress in Minor Lab LIMS by researcher

Last week (17 Apr 2015 - 24 Apr 2015)

Person	Clones	Exprs	Purifs	Macro preps	Plates	Drops	Crystals	Datasets processed	Structure refs	Kinetic assays	Thermal shift assays
Cooper, David	0	0	0	0	0	0	23	18	0	0	0
Handing, Katarzyna	0	0	0	0	0	0	51	53	13	0	0
Hou, Jing	0	0	0	0	1	30	0	1	1	0	0
Kowiel, Marcin	0	0	0	0	0	0	1	8	3	0	0
Shabalin, Ivan	0	0	0	0	0	0	125	14	9	0	0
Shumilin, Igor	0	0	0	0	0	0	0	3	2	0	0
Szlachta, Karol	0	0	0	0	0	0	34	20	3	0	0

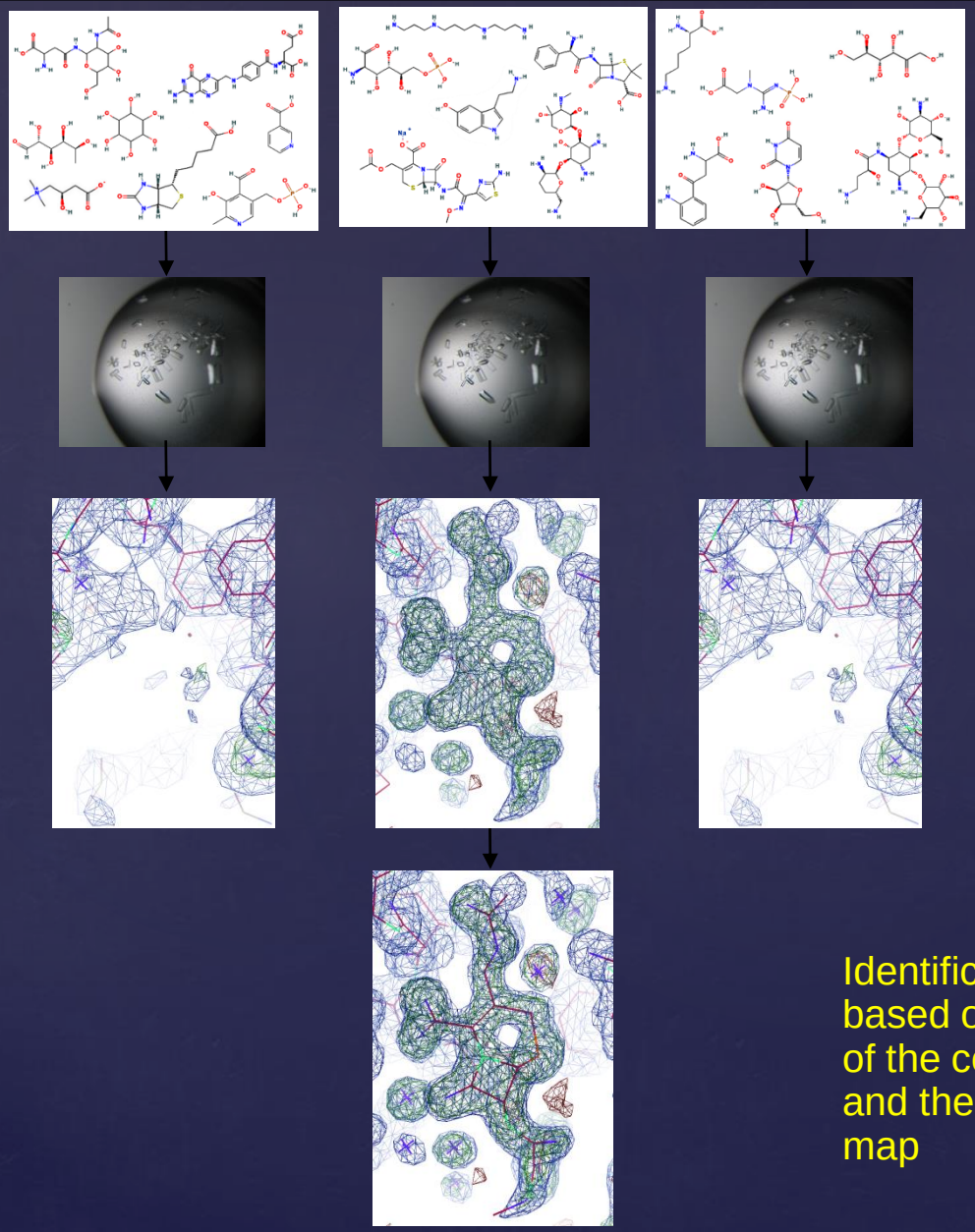
Last month (25 Mar 2015 - 24 Apr 2015)

Target status and path to success



HKL-3000 at SBC

Crystallographic ligand screening



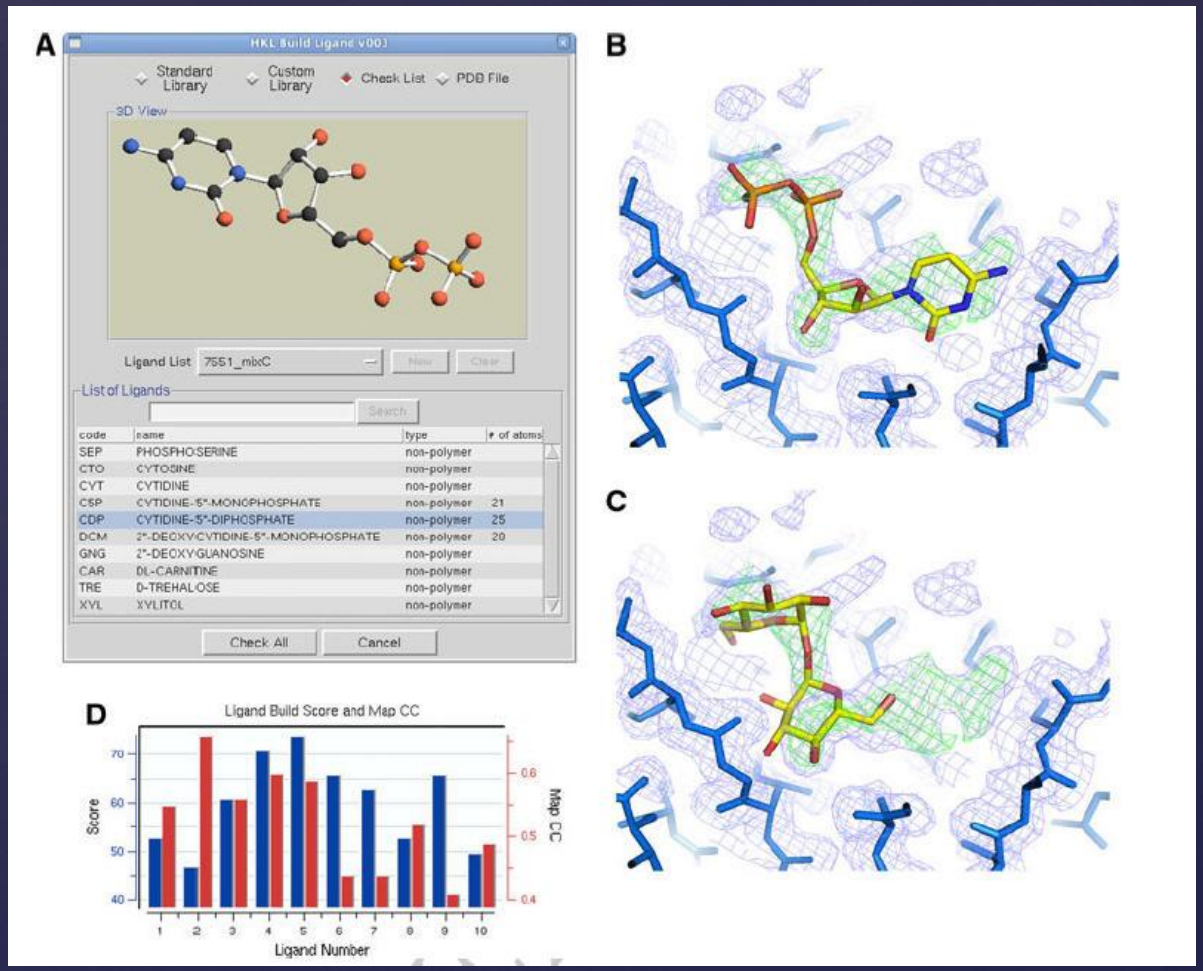
Cocktail solutions,
each composed of 5-10
potential ligands

Soaking protein crystals
with cocktail solutions

Data collection, structure
solution and identification of
additional electron density

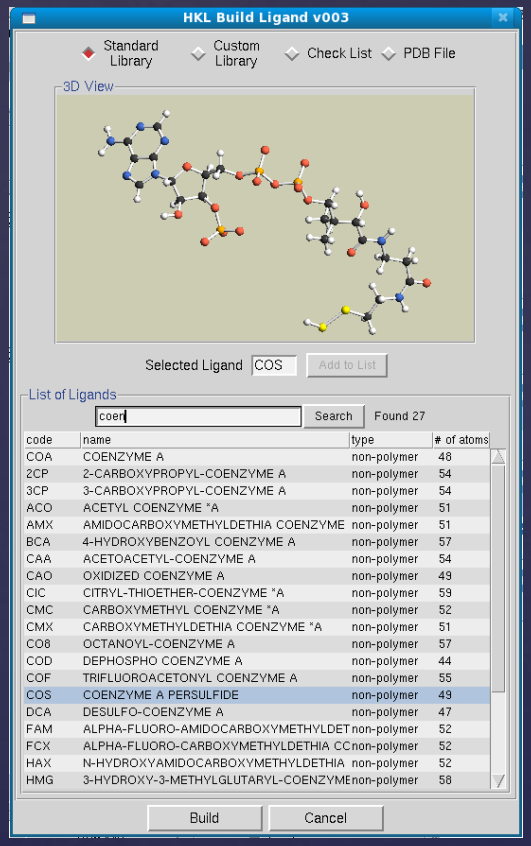
Identification of bound compound
based on the structures
of the cocktail components
and the difference electron density
map

Ligand fitting

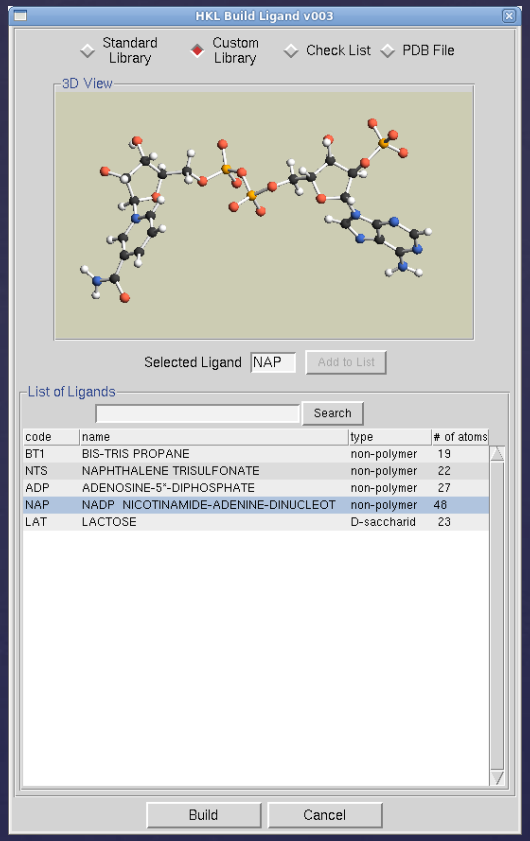


Automatic fitting of single ligand from particular library

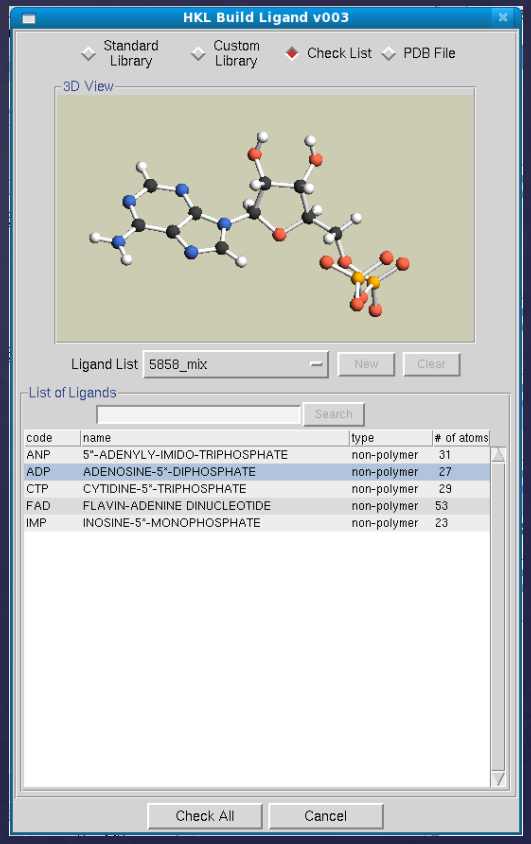
(a)



(b)

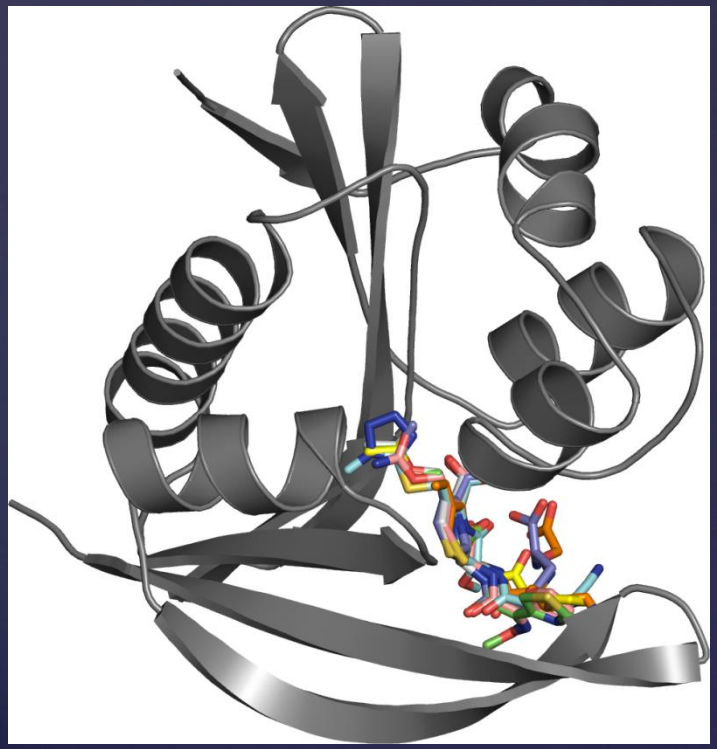


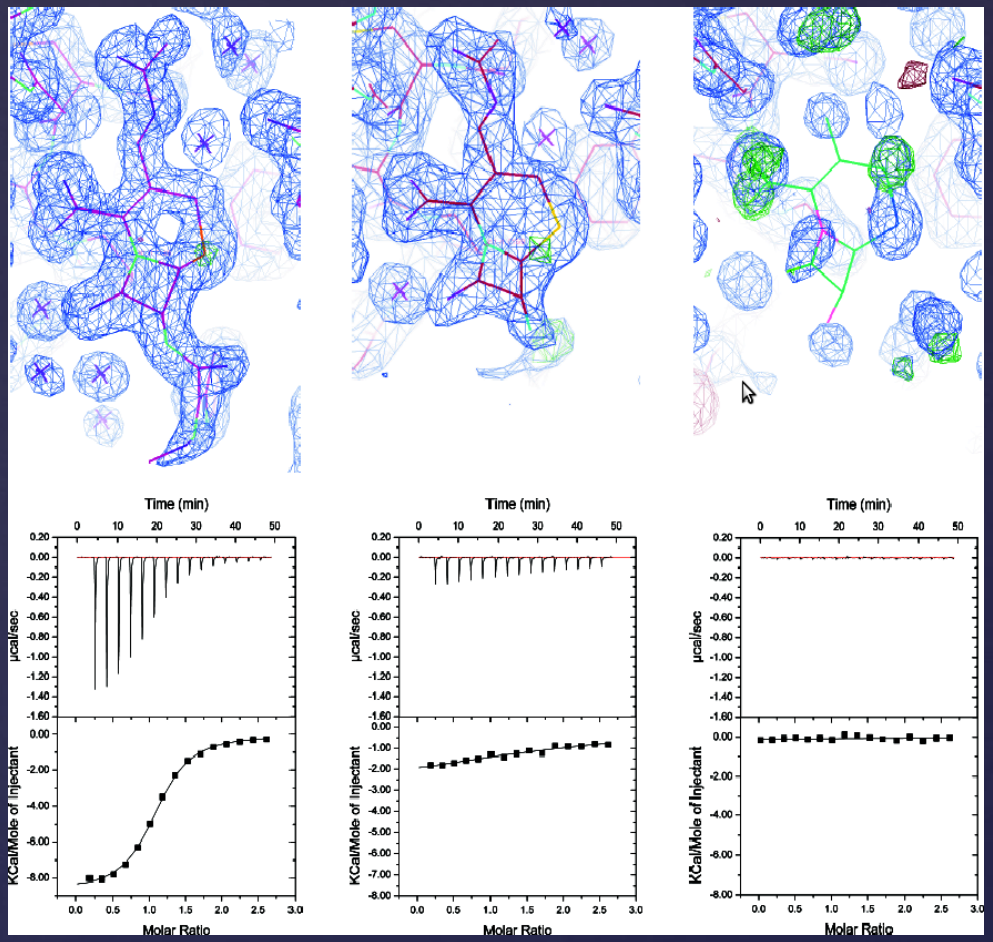
(c)



Ligands evaluation using crystallography and ITC

The search is extended to compounds showing similarity to the first hits.





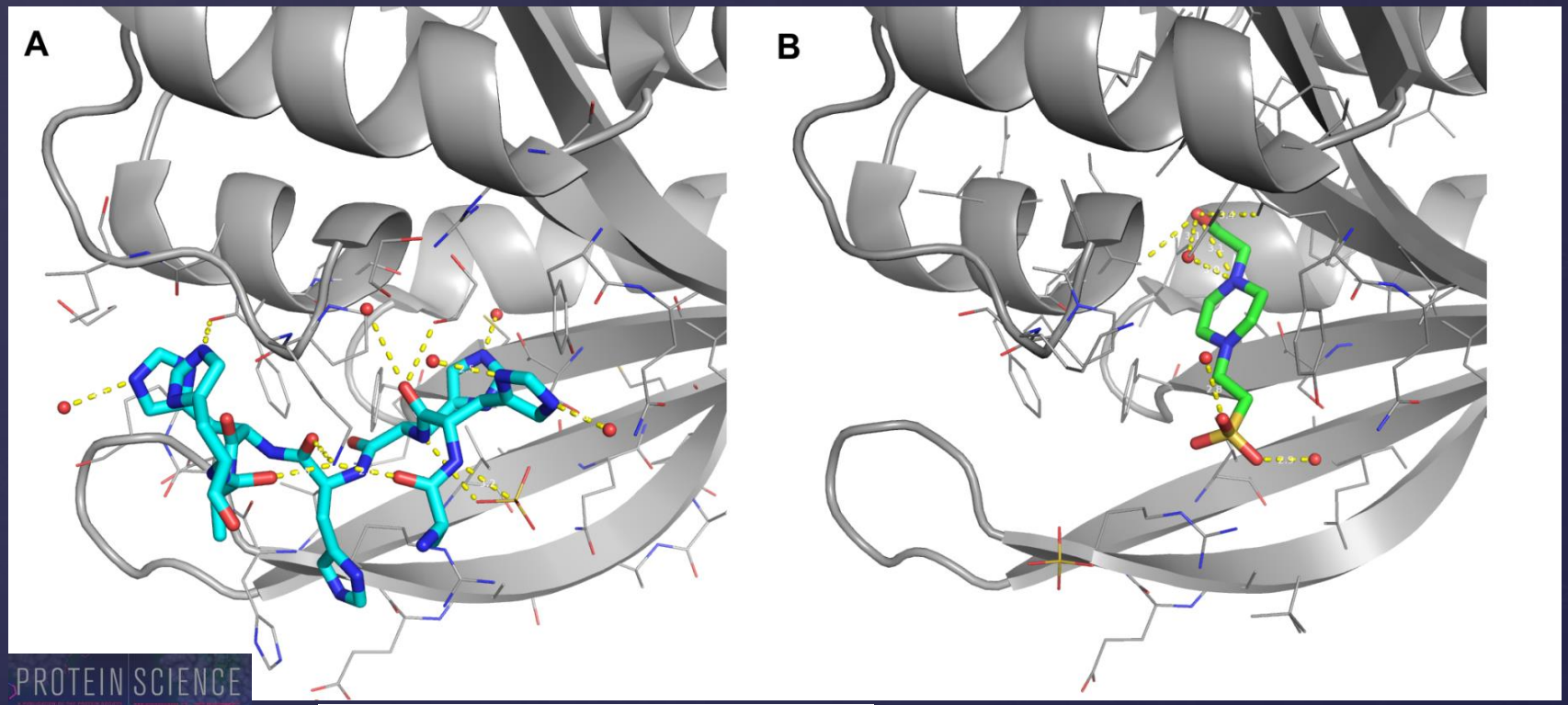
Cefotaxime
Kd = 3.5 ± 0.2 µM

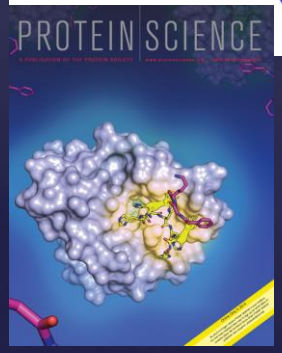
7-aminocephalosporanic acid

7-aminodesacetoxy-cephalosporanic acid

Evaluation of identified ligands using ITC allows to determine the binding affinity of the ligands and to identify which compounds bind with the highest affinity.

Structures of PA4794 with His-tag (A) and HEPES (B) bound to substrate binding site

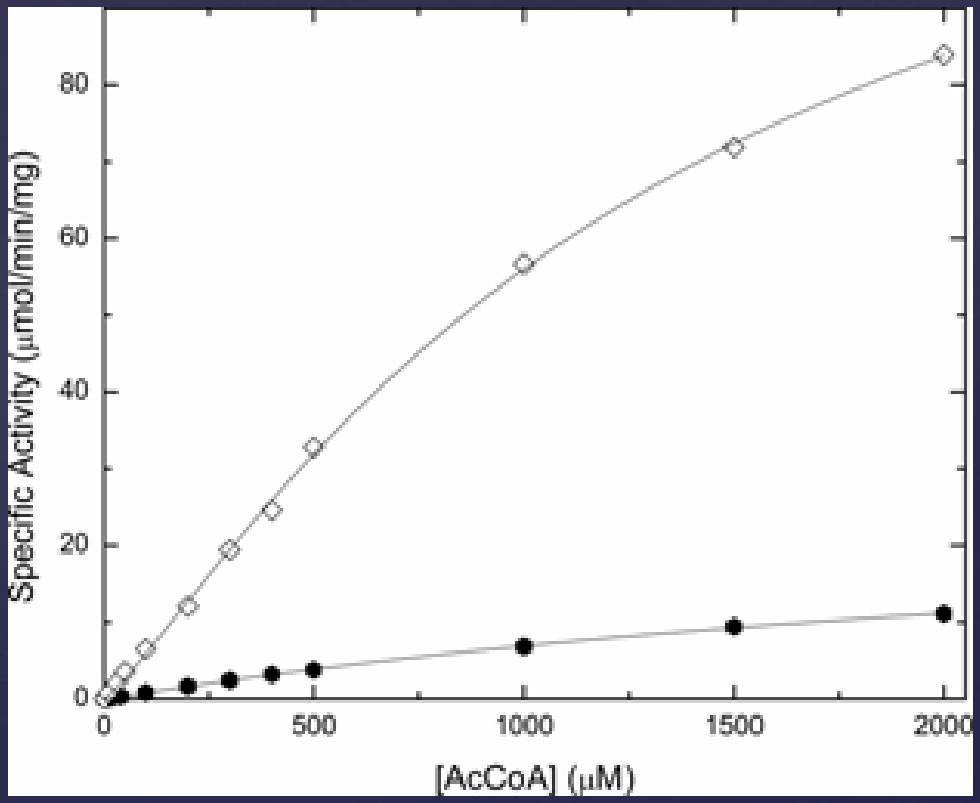




Double trouble—Buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments

Karolina A. Majorek,[1,2,3,4] Misty L. Kuhn,[4,5] Maksymilian Chruszcz,[1,3,4,6] Wayne F. Anderson,[3,4,5] and Wladek Minor[1,3,4]*

Effect of His-tag on SpeG



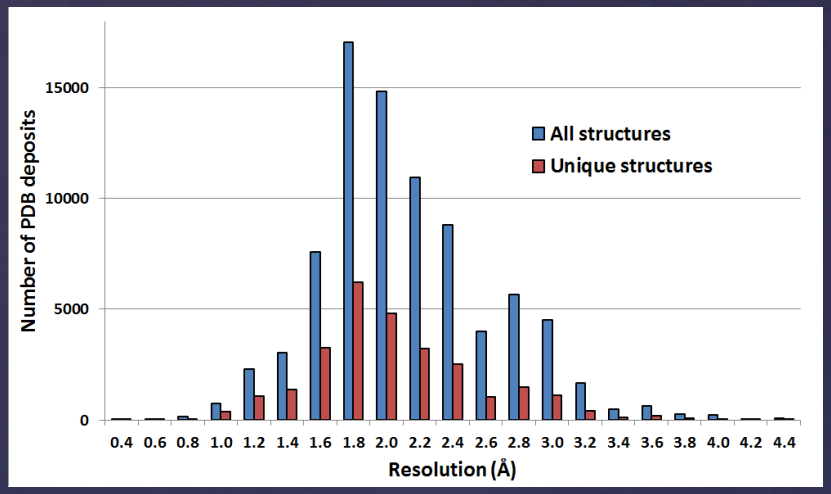
Protein purification and crystallization artifacts: The tale usually not told

Table I. *Known Structures of Proteins that Have Been Identified as Common Purification and Crystallization Artifacts*

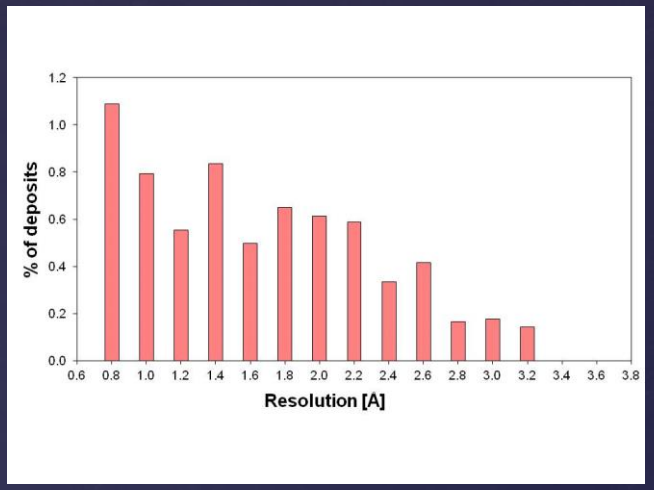
	Name of the protein	Molecular weight (kDa)	PDB ID
Affinity, solubility, anti-aggregation tags	Maltose-binding protein (MBP)	43	1LLS, 1MPB, 3PUW, 3SEU, 4KYC
	Glutathione-S-transferase (GST)	24	4ECB
	Thioredoxin (Trx)	11	1F6M, 2AJQ, 2H73, 4HU9, 4X43
	N-Utilization substance (NusA)	55	1U9L,[a] 1WCN,[b] 2KWP,[c] 4MTN[d]
	Small ubiquitin related modifier 1 (SUMO1)	12	2UYZ, 1Z5S, 4WJQ, 2IO2
	Haloalkane dehalogenase	33	4E46
***E. coli* native proteins**	Metal-binding lipocalin (YodA)	25	1OEJ, 4TNN
	Carbonic anhydrase (YadF)	25	2ESF
	Ferric uptake regulator (Fur)	16	2FU4
	cAMP-regulatory protein (CRP)	24	1CGP, 2CGP, 2GZW, 3FWE, 3HIF, 3N4M, 3QOP, 4FT8, 4HZF, 4I0A, 4I0B, 4N9H, 4N9I
	Glucosamine-6-phosphate synthase (GlmS)	67	4AMV, 1JXA, 3OOJ, 2J6H
	Glycogen synthase (GlgA)	53	2QZS
	Component 1 of the 2-oxoglutarate dehydrogenase complex (ODO1)	105	2JGD
	Component E2 of dihydrolipoamide succinyltransferase (ODO2)	44	1C4T
	Formyl transferase (YfbG, ArnA)	46	1U9J, 1YRW, 1Z7E, 2BLN, 4WKG
	Cu/Zn-superoxide dismutase (Cu/Zn-SODM)	16	1ESO
	Chloramphenicol-O-acetyl transferase (CAT)	26	1Q23
	Host factor-I protein (Hfq)	11	3VU3
Proteases	Tobacco etch virus (TEV)	28	1LVM
	Rhinovirus 3C protease	48	1CQQ
	SUMO protease C-terminal domain	26	2HL9
	Enterokinase	26	1EKB
	Trypsin	26	3UY9
	Chymotrypsin	26	1GGD
	Thrombin (active form)	36	3SQE, 1MH0, 4H6T
	Thermolysin	60	4D9W
	Proteinase K	40	3DVS
	Pepsin	41	5PEP
	Neutrophil elastase	29	5ABW
	LysN Peptidyl-Lys metalloendopeptidase	44	1GE7
	Lysyl endopeptidase	28	4NSY
	Factor Xa	55	1KIG
Exogenous proteins	Lysozyme	16	4TWS, 4PRQ, 1AKI
	DNase protein	31	2A40

Niedzialkowska *et al*, *Protein Sci. 2016 Mar;25(3):720-33*

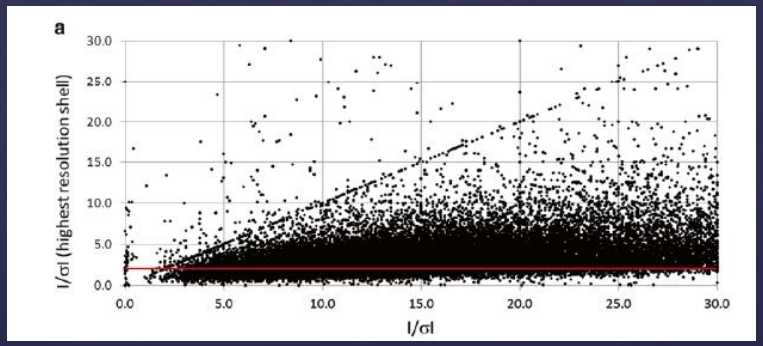
Resolution and ligands in PDB



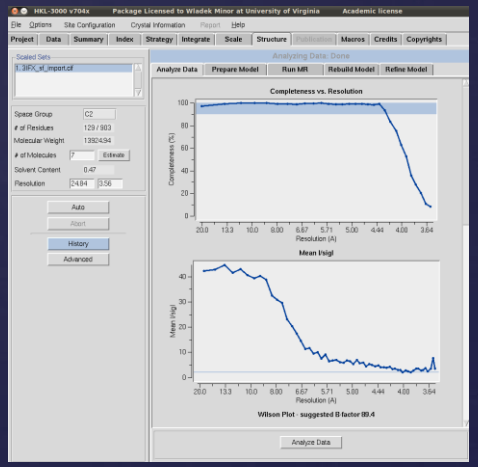
Cooper *et al, Expert Opinion on Drug Discovery (2011) 6: 771-782*



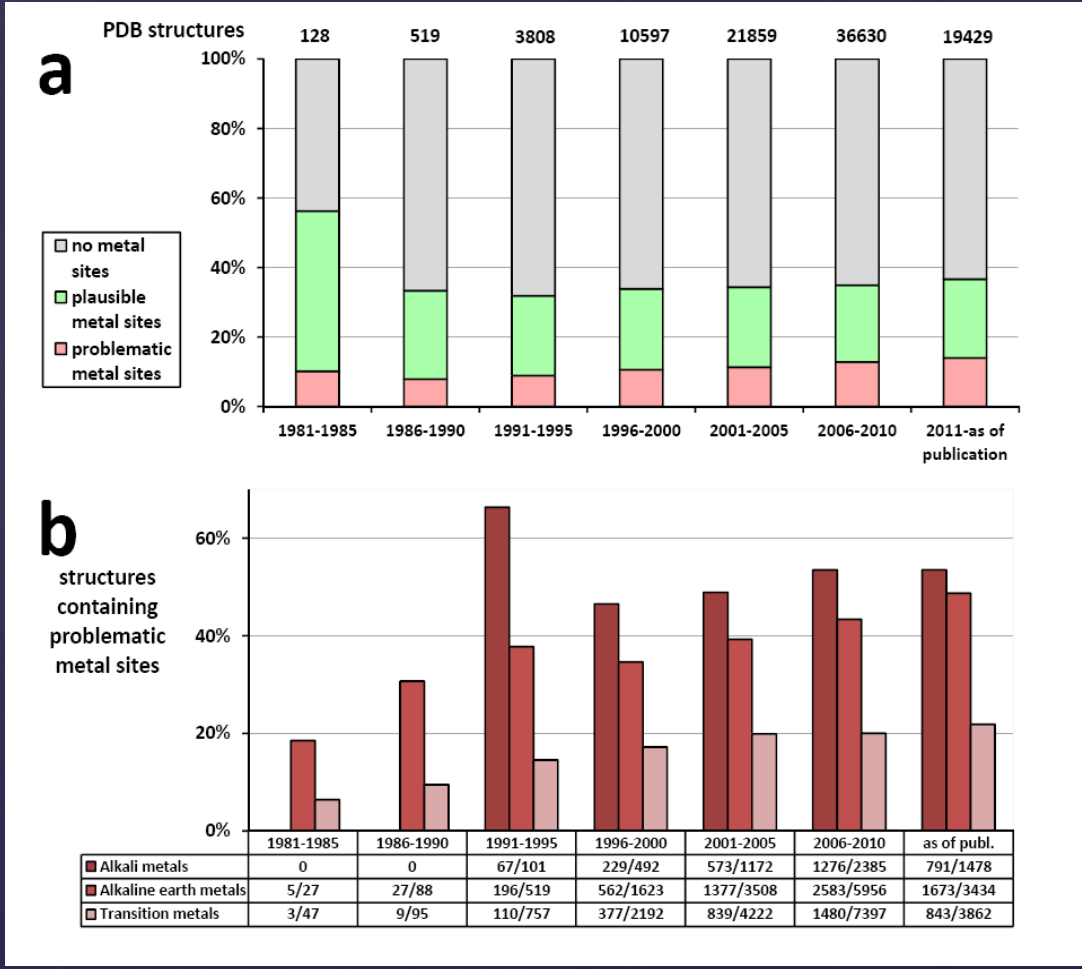
Cooper *et al, Expert Opinion on Drug Discovery (2011) 6: 771-782*



Domagalski *et al. Methods in Molecular Biology (2014) 1091: 297-314*



Metals in PDB

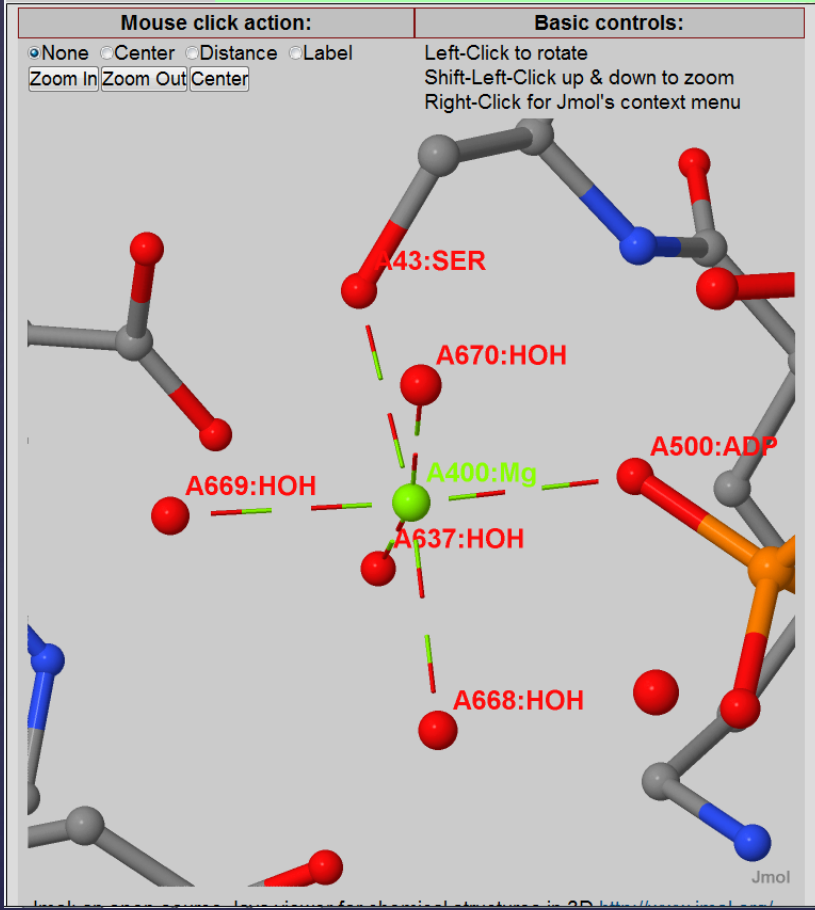


	1981-1985	1986-1990	1991-1995	1996-2000	2001-2005	2006-2010	as of publ.
Alkali metals	0	0	67/101	229/492	573/1172	1276/2385	791/1478
Alkaline earth metals	5/27	27/88	196/519	562/1623	1377/3508	2583/5956	1673/3434
Transition metals	3/47	9/95	110/757	377/2192	839/4222	1480/7397	843/3862

Zheng *et al Nature protocols* (2014)9: 156-70

Metal binding site validation: *CheckMyMetal* server

ID	Res	Atom	Valence	BV symmetry	Geometry	RMSD geometry angles	Missing vertices	Bidentate	CBVS	Alternative metal
400:A	_MG	MG--	2.1	0.111	Octahedral	5.36	0	0	4.41	
400:B	_MG	MG--	2.06	0.079	Octahedral	4.13	0	0	4.32	



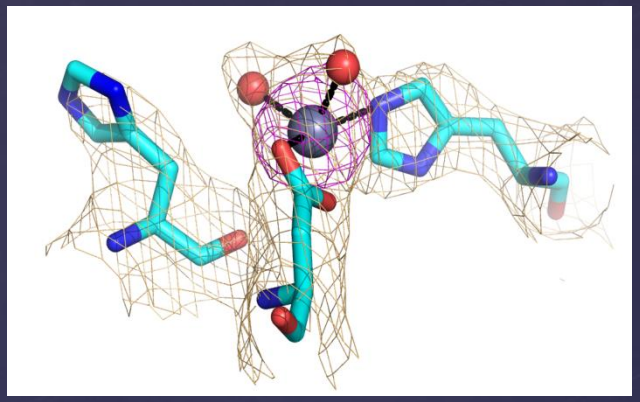
Use the buttons below to control the view

Residue Name:	On	Off
Metal Distances:	On	Off
Protein Cartoon:	On	Off
Spin:	On	Off
Antialiasing:	On	Off

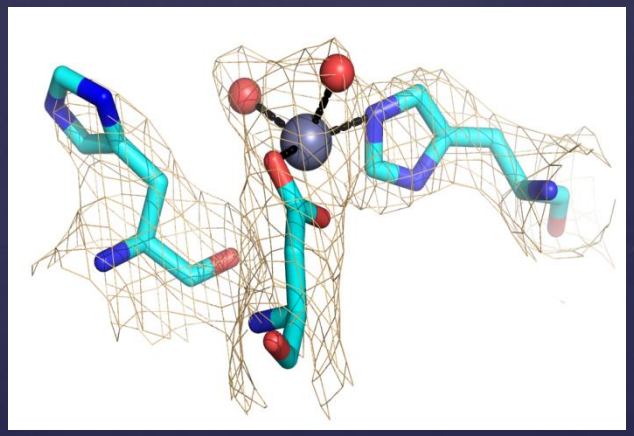
Legend	Explanation
Valence	Summation of bond valence values for an ion binding site
BV symmety	Summation of bond valence vectors, weighted by bond valence values. Increase when the coordination sphere is not symmetrical due to incompleteness.
Geometry	Ion binding site geometry, as calculated by the NEIGHBORHOOD algorithm
RMSD geometry angles	R.M.S. Deviation of observed geometry angles (L-M-L angles) compared to ideal geometry, in degrees
Missing vertices	Number of sites that is not occupied in the coordination sphere for the assigned geometry
Bidentate	Number of residues that form a bidentate interaction instead of being considered as multiple ligands
CBVS	Calcium Bond Valence Sum, used for alternative metal(s) prediction [Muller,P. et al. (2003) Is the bond-valence method able to identify metal atoms in protein structures? Acta Crystallogr. D Biol. Crystallogr., 59, 32-37.]
Alternative metal	A list of alternative metal(s) is proposed in descending order of confidency, assuming metal environment is accurately determined

Zheng *et al Nature protocols* (2014) 9: 156-70

The way to prove presence of metal



Metal binding site A
9668eV



Metal binding site A
9618eV

Data collected below and above zinc absorption edge - APS 19BM

Dissonance in Science

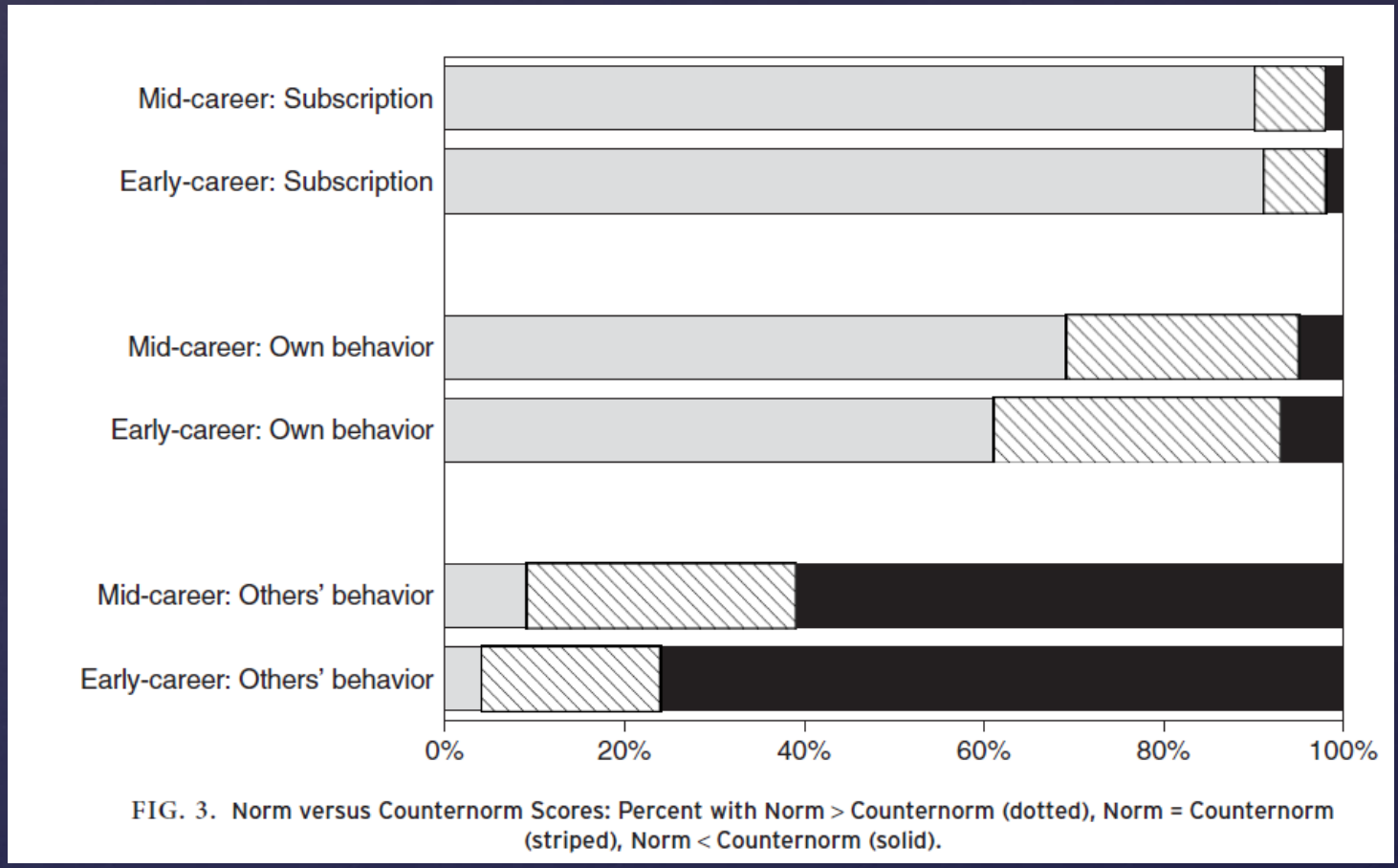


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

Anderson *et al J. Empirical Res* (2007) (4)6-70

Acknowledgments

Wladek Minor
- Marcin Cymborowski
- Maks Chruszcz
- Matt Zimmerman
- Heping Zheng
- Ivan Shabalin
- Ewa Niedzialkowska
- Karolina Majorek
- Katherine Handing

Zbyszek Otwinowski
- Dominika Borek

Andrzej Joachimiak
MCSG and SBC staff

Wayne Anderson
CSGID staff

Steve Almo
NYSGRC Staff

Zbyszek Dauter
Mariusz Jaskolski
Alexander Wlodawer

- Garib Murshudow
- Paul Emsley
- Tom Terwilliger
- George Sheldrick

Grants:
NIH GM53163, GM62414, GM74942
GM117080, HG008424, GM117325
HHSN272200700058C,HHSN272201200026C
HKL Research. Inc.