

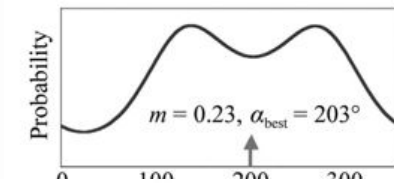
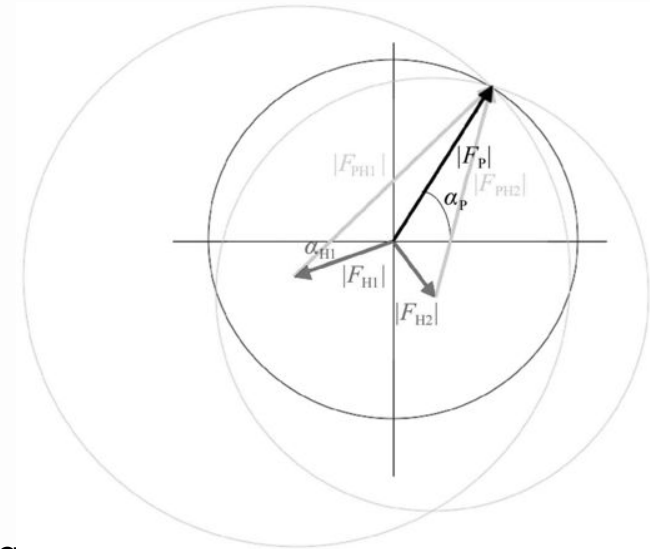
Biology IV: Experiments / Refinement

Thomas R. Schneider, EMBL Hamburg
2/7/2015

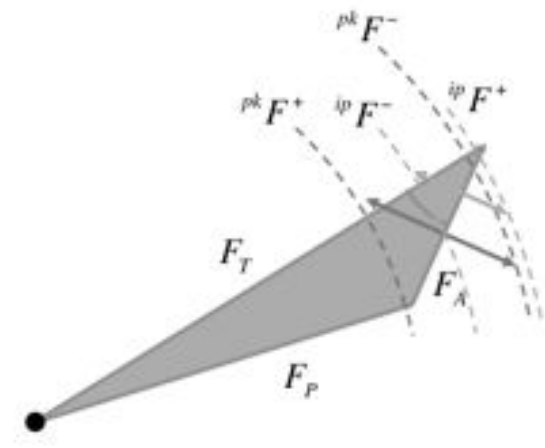
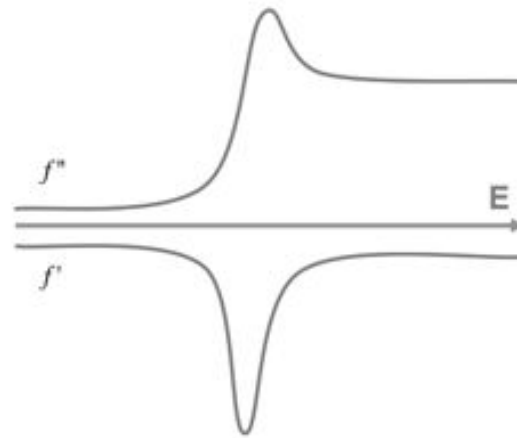
thomas.schneider@embl-hamburg.de

Recap

- Phasing methods
 - Ab initio
 - Molecular Replacement – 25%. Model bias
 - Multiple Isomorphous Replacement
 - Multiple Anomalous Diffraction
 - Single Anomalous Diffraction – 1%. phase ambiguity
- Importance of accurate measures of data quality



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



Today

- Experiments
 - Beams
 - Diffractometry
 - Low-energy data collection
 - Automation
- Refinement
- XFEL

Diffraction angles and resolution

- Smaller lattice spacings d give rise to higher diffraction angles Θ :

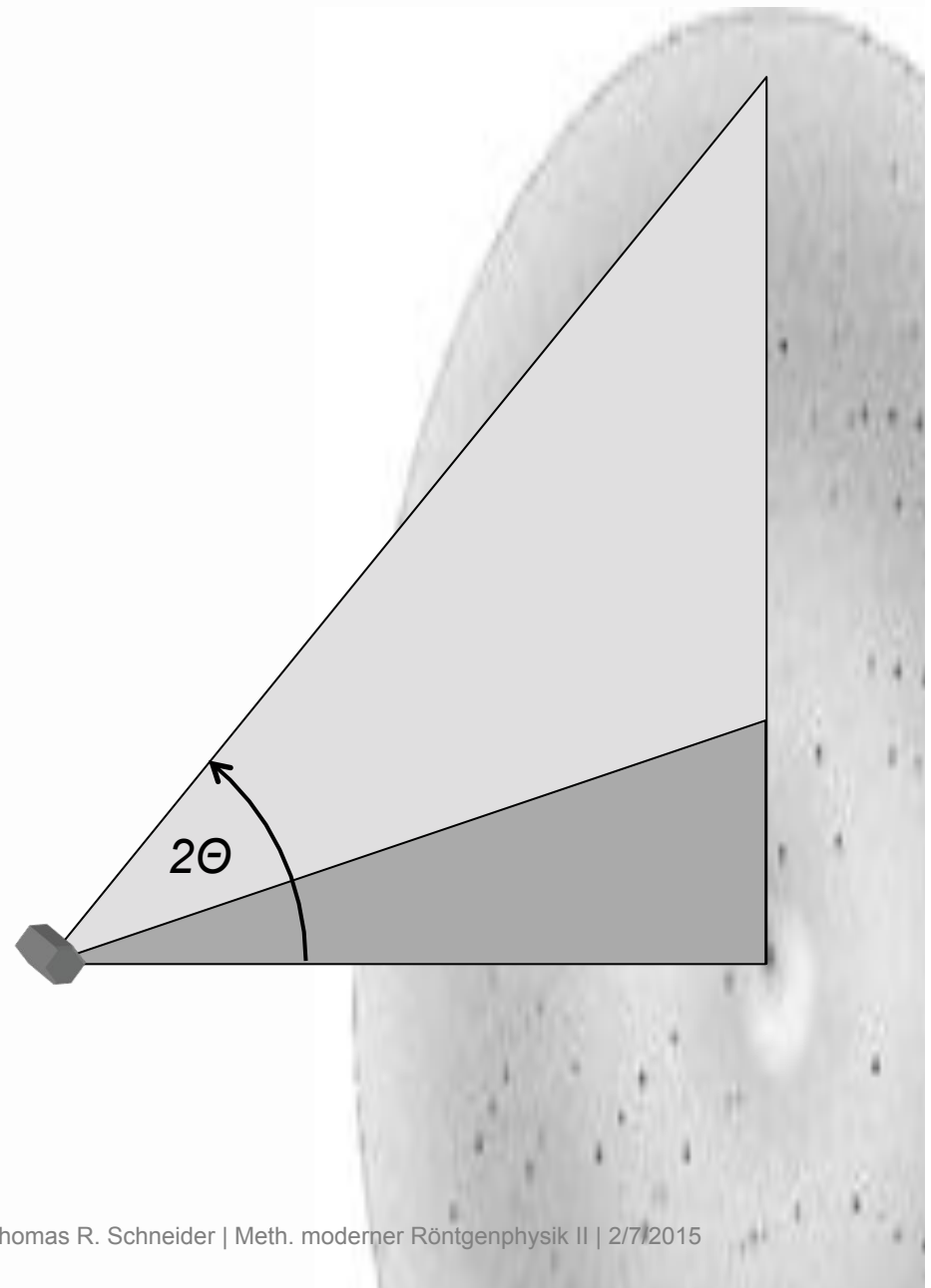
$$2 d \sin\Theta = n \lambda$$

$$\Theta = \arcsin(n \lambda / 2d)$$

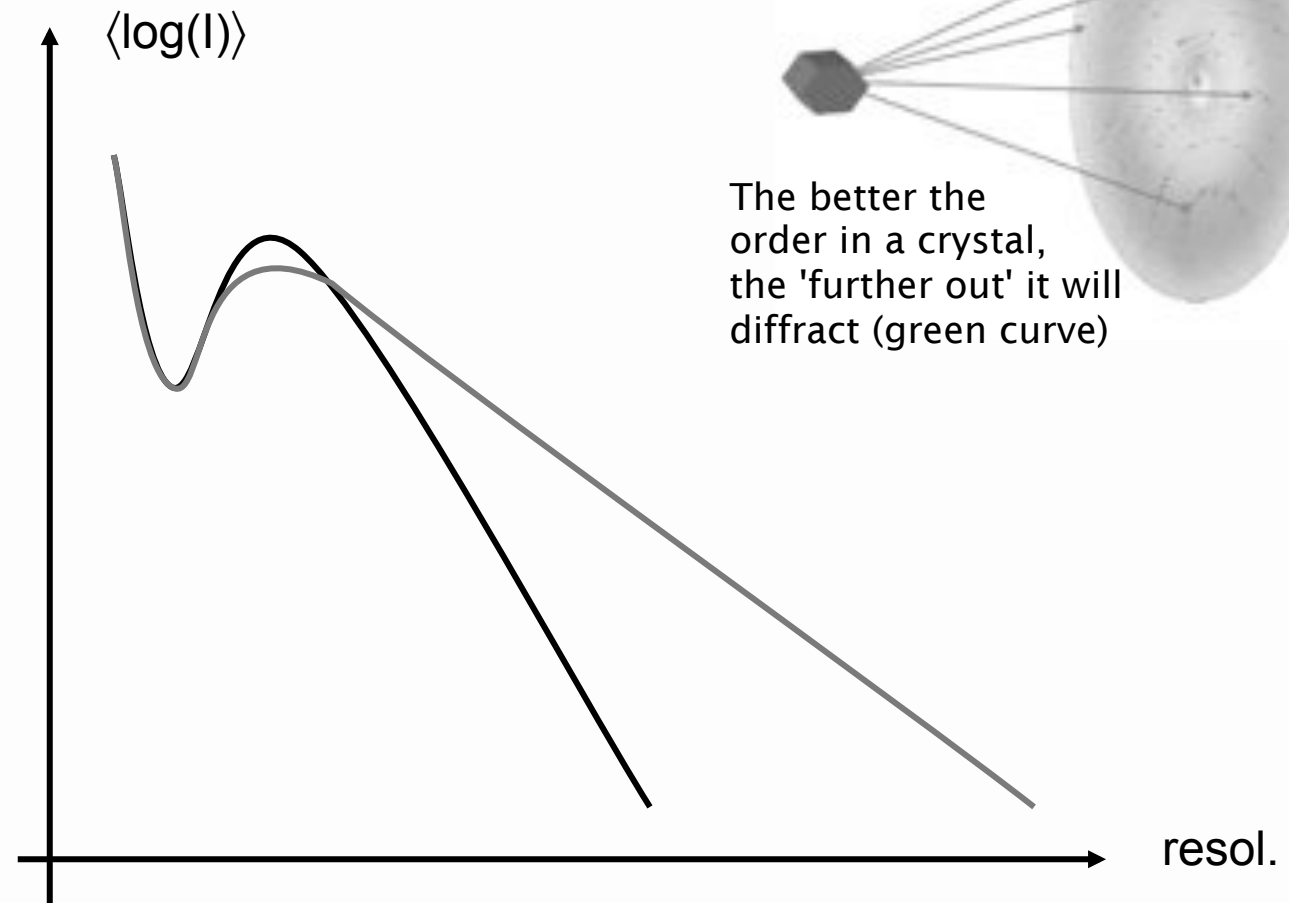
$$\arcsin(1 \text{ \AA} / 2 * 3 \text{ \AA}) = 9.6^\circ$$

$$\arcsin(1 \text{ \AA} / 2 * 1 \text{ \AA}) = 30.0^\circ$$

- The lattice spacing roughly corresponds to the optical resolution.
- The higher the diffraction angle, the weaker is the average diffracted intensity

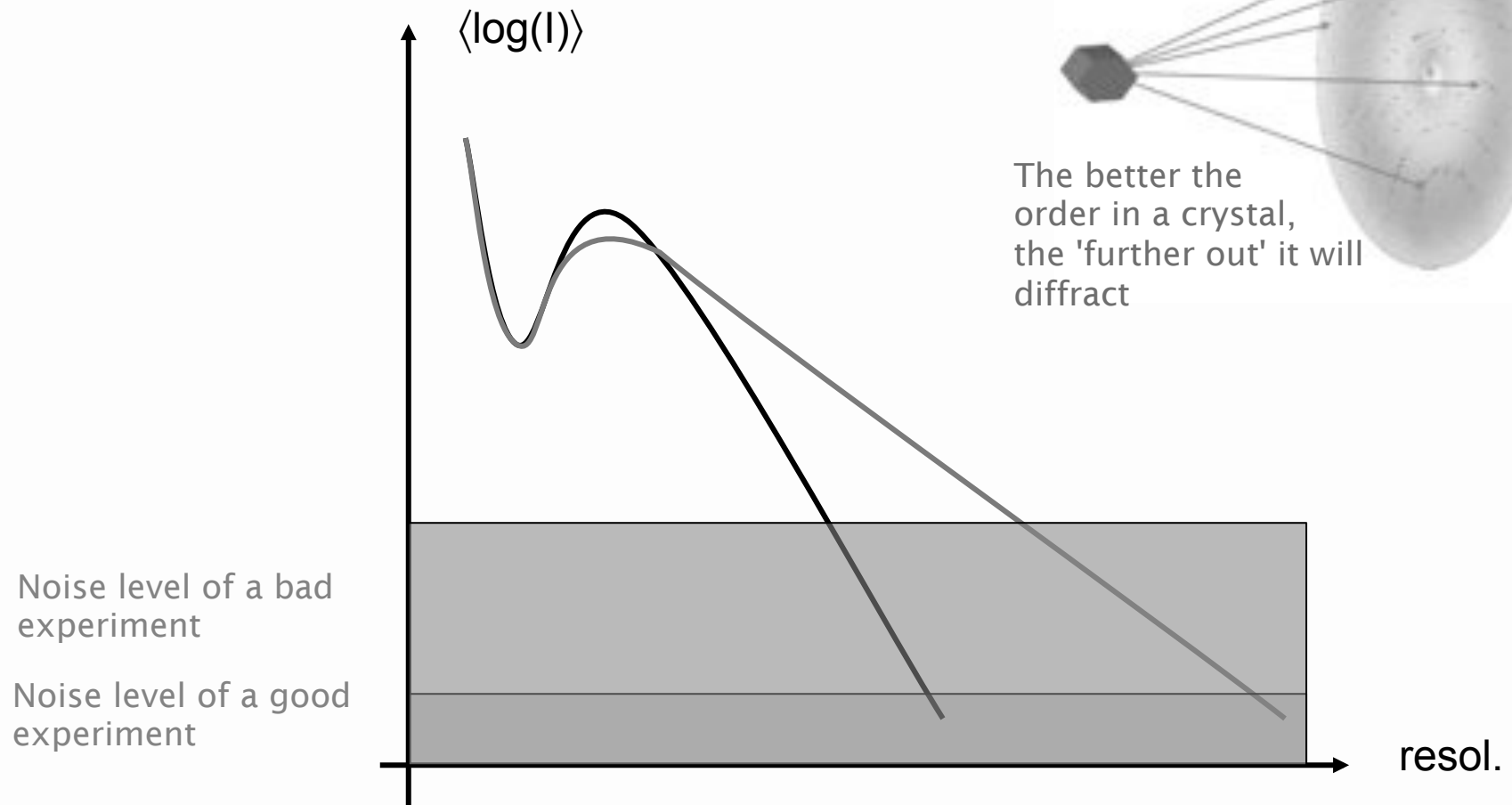


Limited resolution in crystallography



- Diffraction data have an enormous dynamic range
- The higher the resolution / the larger the unit cell, the smaller the diffracted intensities

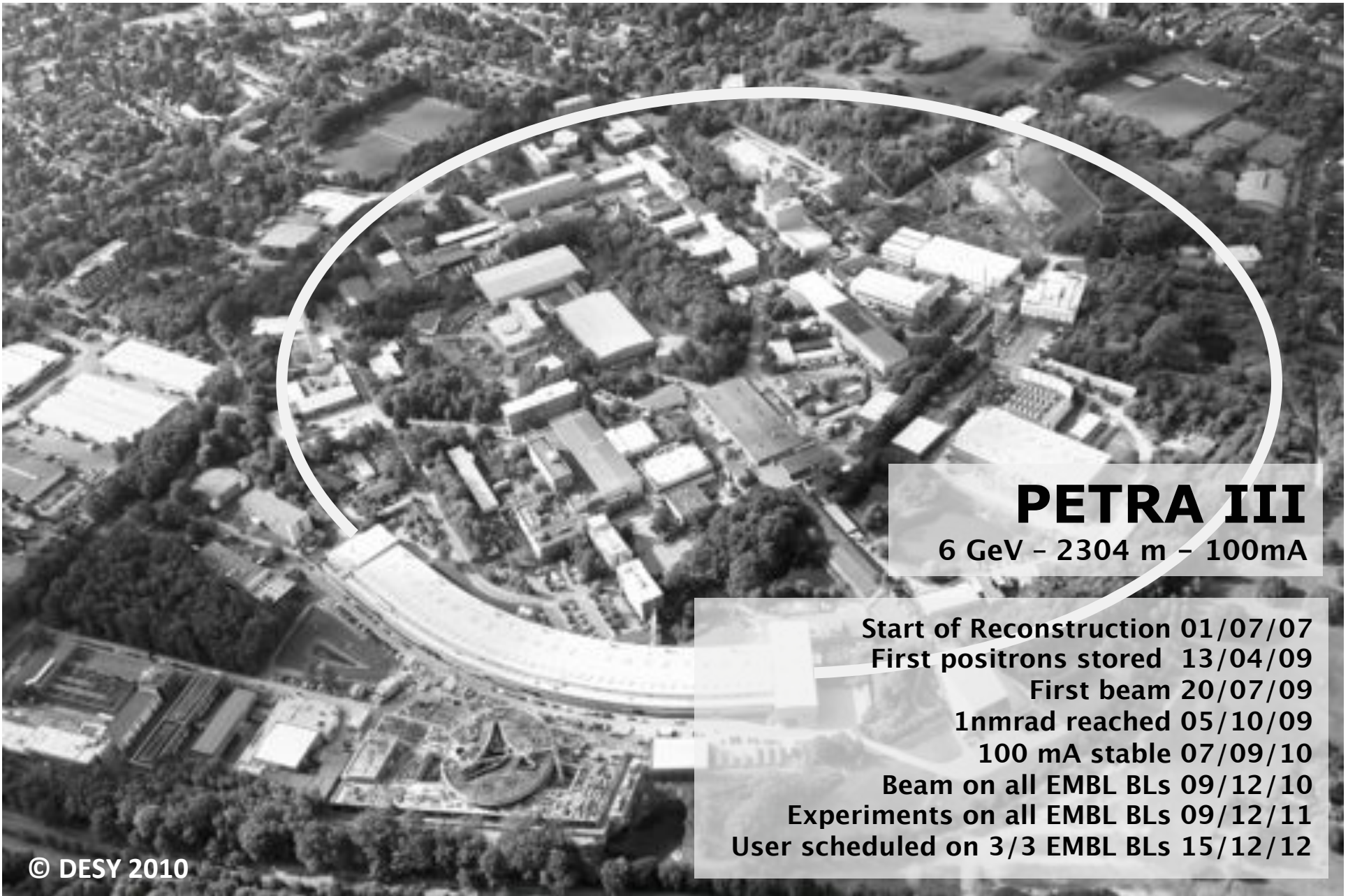
Limited resolution in crystallography



- Diffraction data have an enormous dynamic range
- The higher the resolution / the larger the unit cell, the smaller the diffracted intensities

Why does anomalous phasing not work all the time?

- Molecule of interest does not contain anomalous scatterers
 - Seleno-Met incorporation
 - Derivatization
 - Co-crystallization
- The signal is too weak (small differences between large numbers)
 - Increase signal by choosing the right wavelength (synchrotron ...)
 - Enhance signal e.g. by oxidation of Se-atoms
 - Decrease background by optimizing the hardware, crystal mount, experiment
 - Add extra data -> SIRAS, MIRAS
 - More / more accurate data (inverse beam, kappa geometry)
- Crystals suffer from radiation damage
 - Keep crystal at cryogenic temperatures
 - Play tricks prolonging crystal lifetime (scavengers)
 - Careful use of tolerable dose (-> BEST)
 - Data from several positions on the same crystal
 - Data from several crystals



PETRA III

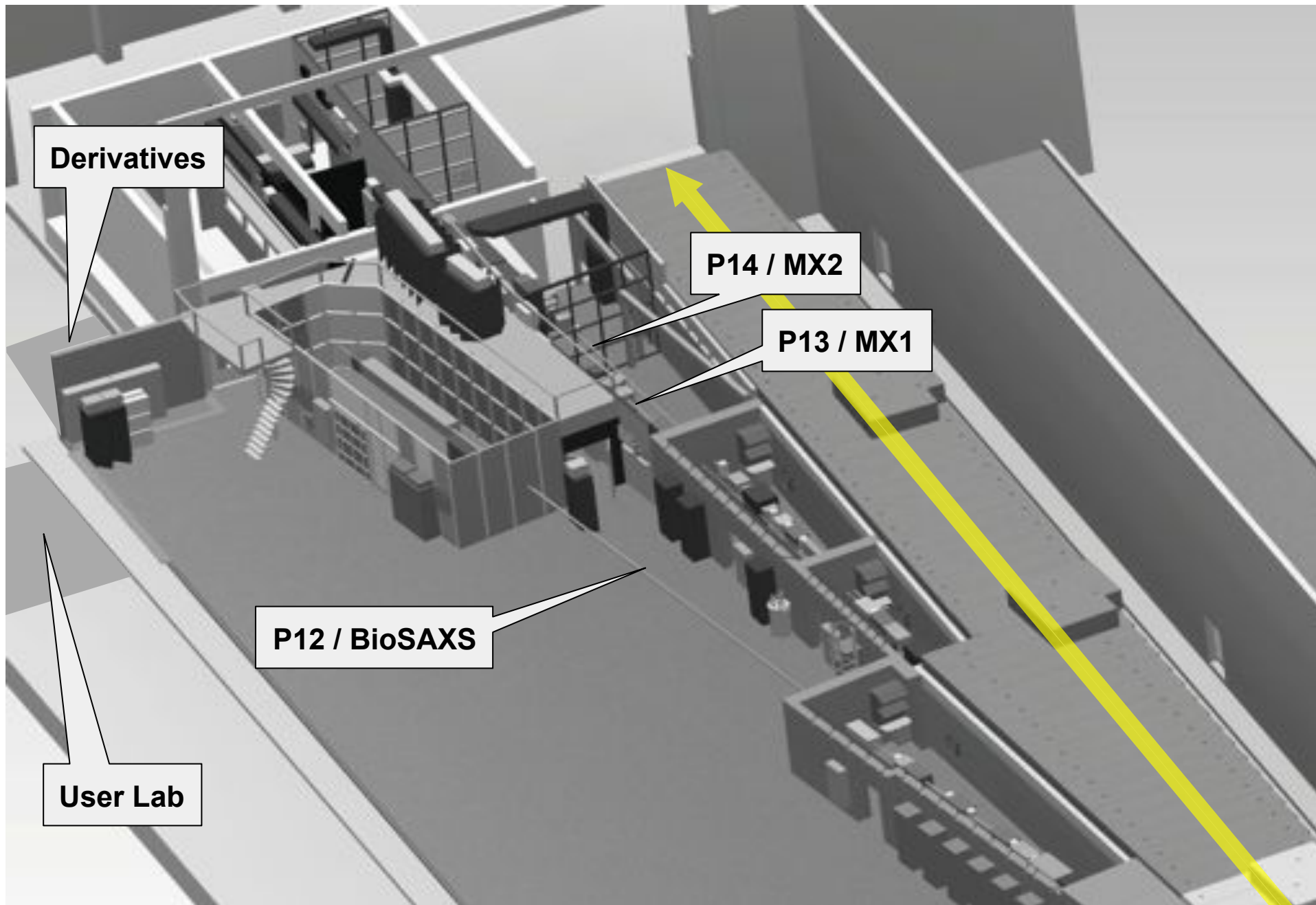
6 GeV - 2304 m - 100mA

Start of Reconstruction 01/07/07
First positrons stored 13/04/09
First beam 20/07/09
1nmrad reached 05/10/09
100 mA stable 07/09/10
Beam on all EMBL BLs 09/12/10
Experiments on all EMBL BLs 09/12/11
User scheduled on 3/3 EMBL BLs 15/12/12

© DESY 2010

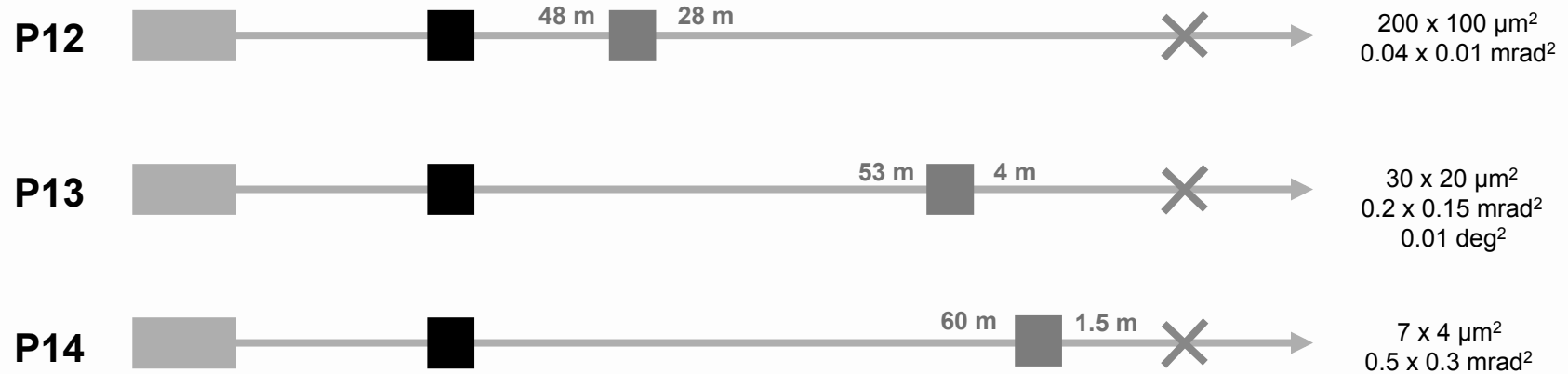


© DESY 2010



EMBL@PETRA3 - Optics

ADAPTIVE OPTICS



	P12:BioSAXS	P13:MX1	P14:MX2
Energy [keV]	4-20	5(4)-17	7-35
Monochromators	Si(111)	Si(111)	Si(111)
Beam size H x V [μm^2]	200 x 60	28 x 13	4 x 1
Divergence H x V [mrad]	0.04 x 0.01	0.2 x 0.15	<0.5 x <0.3
Demagnification H / V	1:1.4 / 1: 1.2	1:12 / 1:15	1:60 / 1:40
Intensity [ph/sec]	10^{13}	10^{13}	10^{12}

P13 / MX1

4-16keV, 30 μm focus

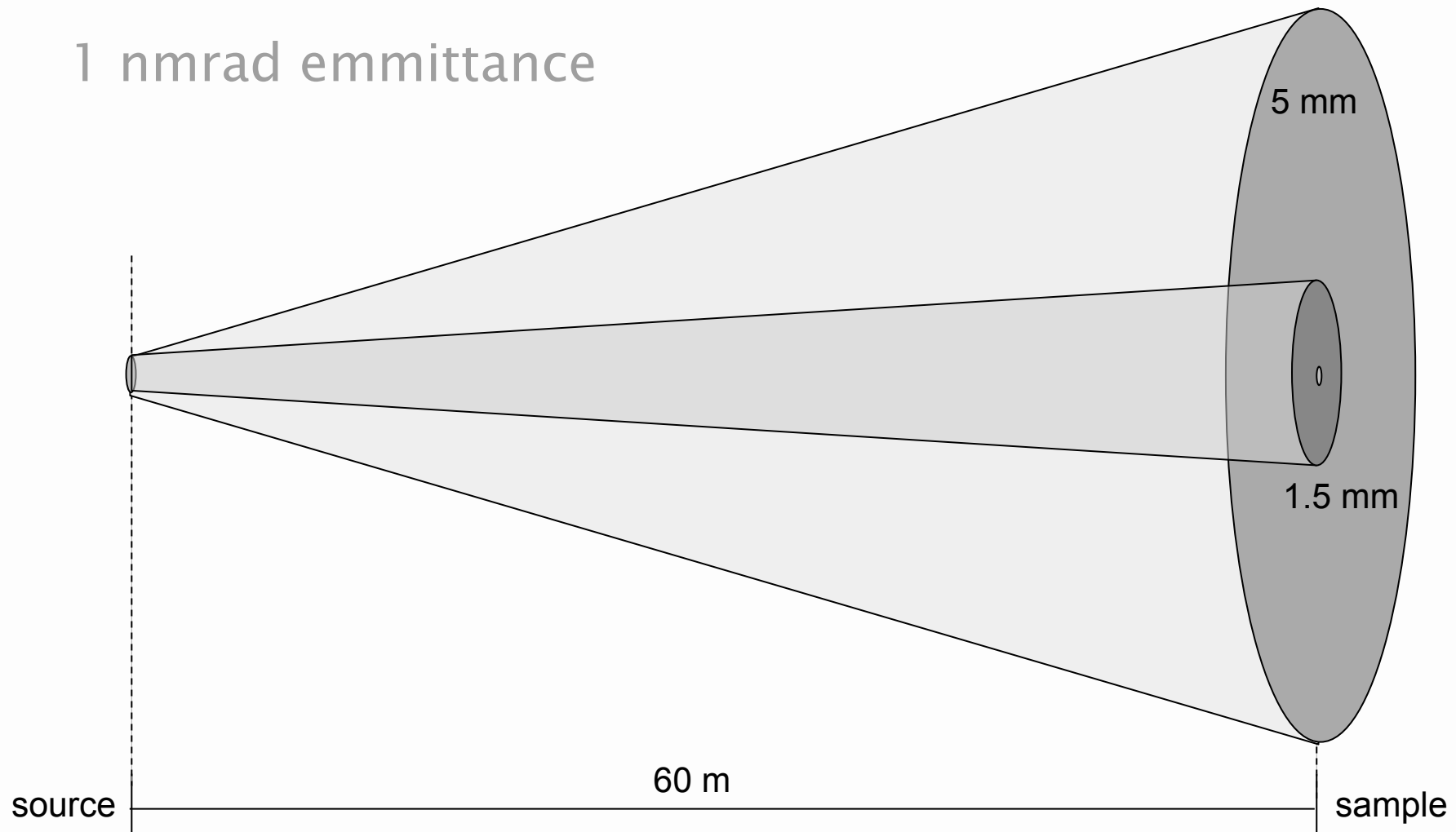


P14 / MX2

7-15 [35] keV, 5 x 7 [1x5] μm focus

DECTRIS
6M
PILATUS

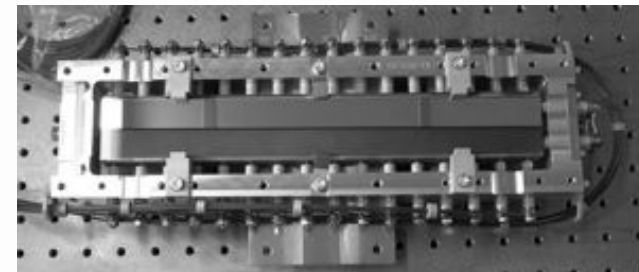
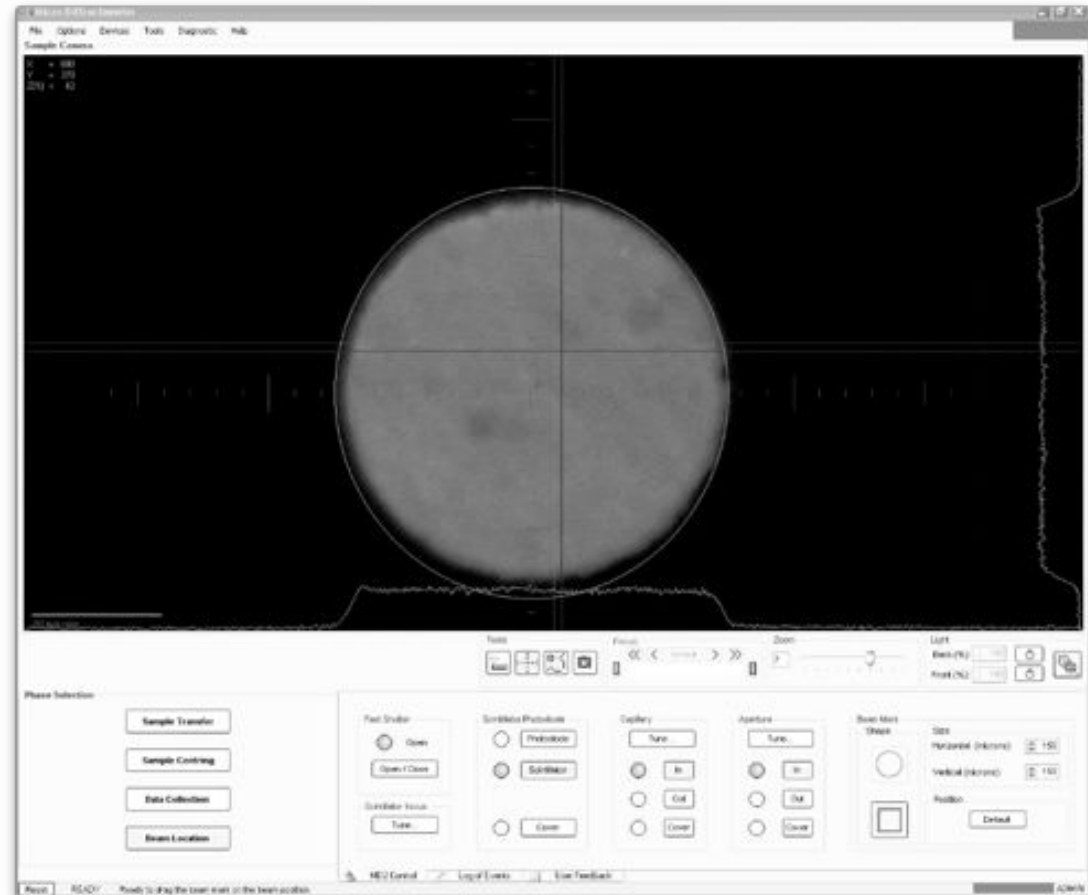
1 nmrad emittance



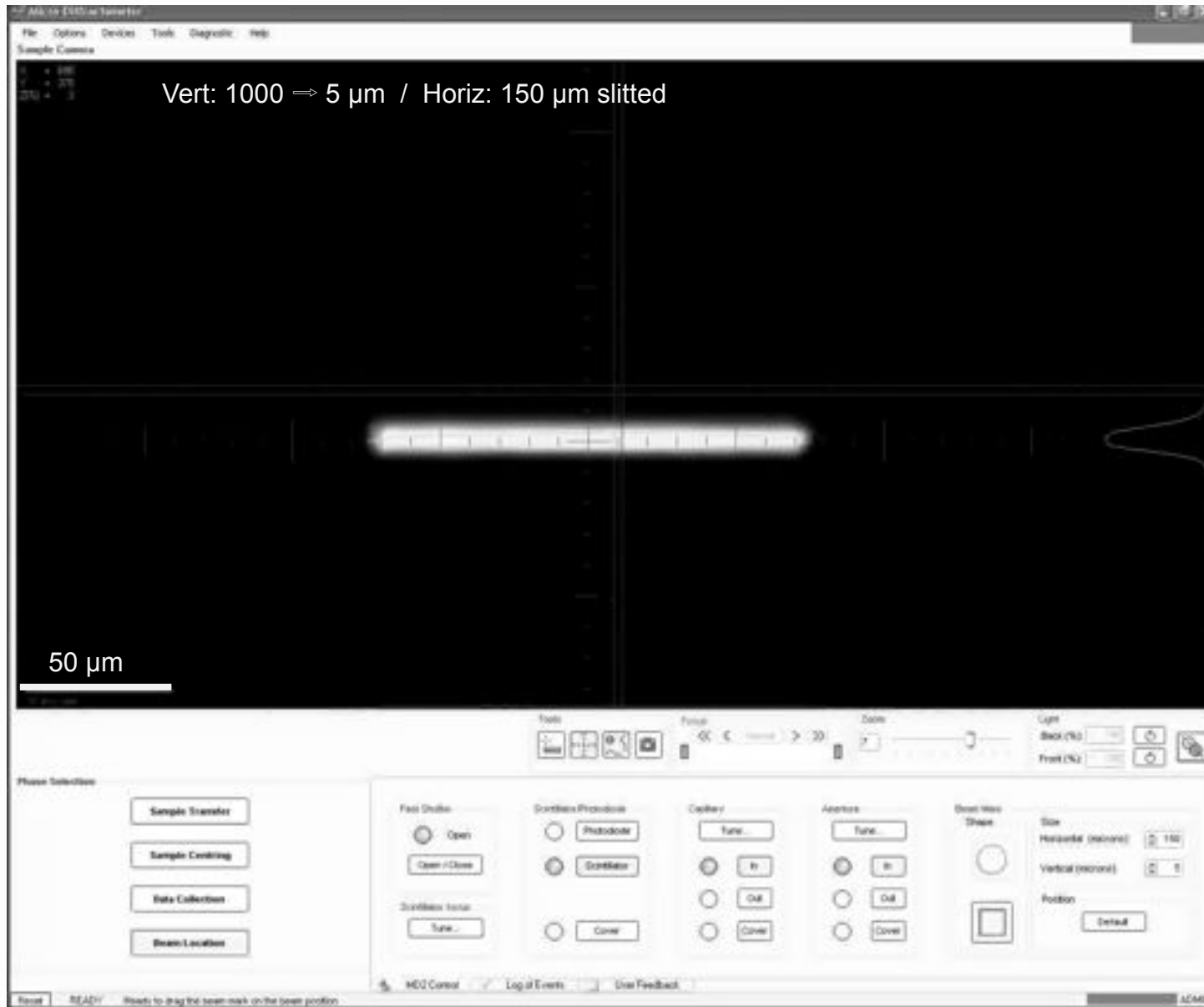
- 10^{13} monochromatic ph/sec to start with
- Flux through a $150 \mu\text{m}$ aperture is ca. 5×10^{11} ph/sec \Rightarrow 20 min. xtal-lifetime
- $20 \mu\text{rad} = 20 \mu\text{m}/\text{m}$. Compare to $0.1^\circ \text{mos.} = 1700 \mu\text{rad}$

P14: unfocused beam

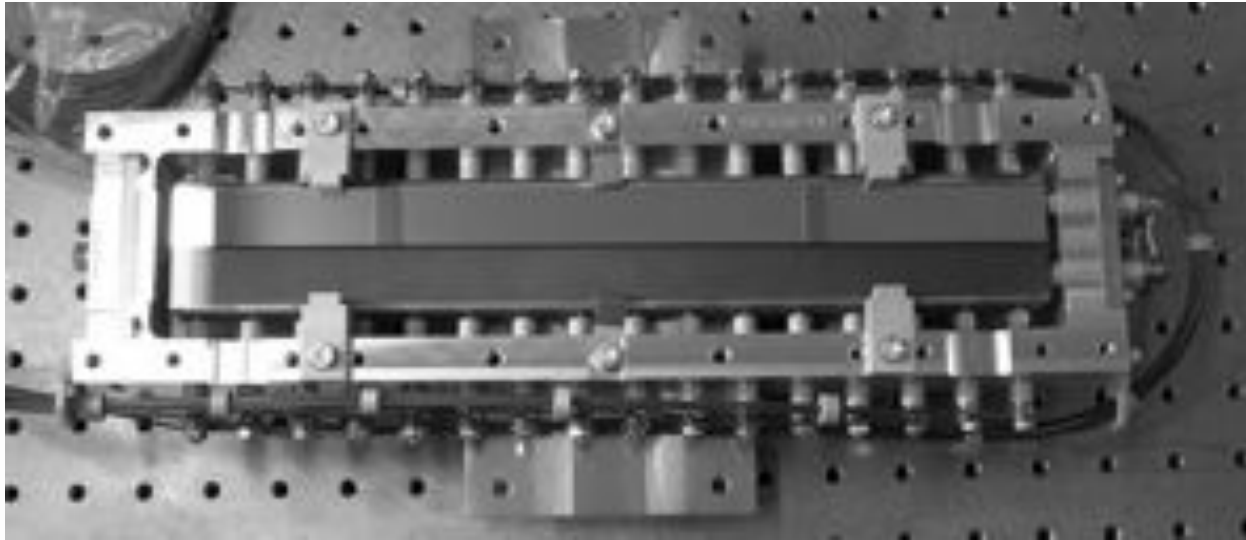
- Central beam profile is essentially top-hat
- flux: 5×10^{11} ph/sec in $150 \mu\text{m} \varnothing$
- Beam position is very stable:
 - rmsd $1.5 \mu\text{m}$ @ <10 Hz
 - rmsd $7 \mu\text{m}$ @ > 30 Hz
 - N.B. 60 m from the source, 16m from the DCM
- Introduce vertically focusing mirror to reduce beam size from $1000 \mu\text{m}$ to $5 \mu\text{m}$



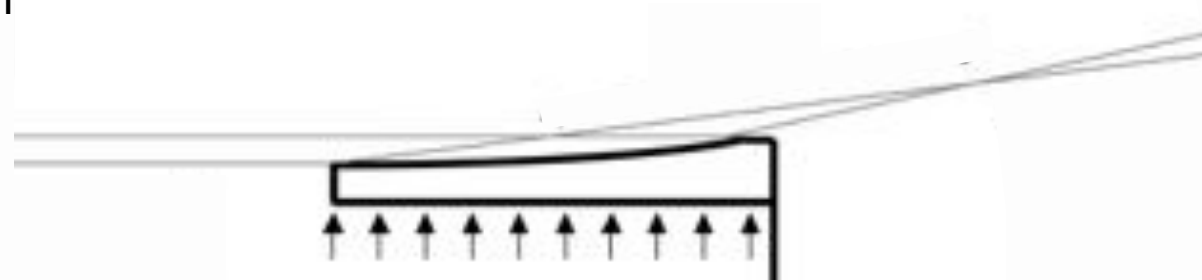
Vertical (Un)focusing ...



Adaptive X-ray focusing Mirrors

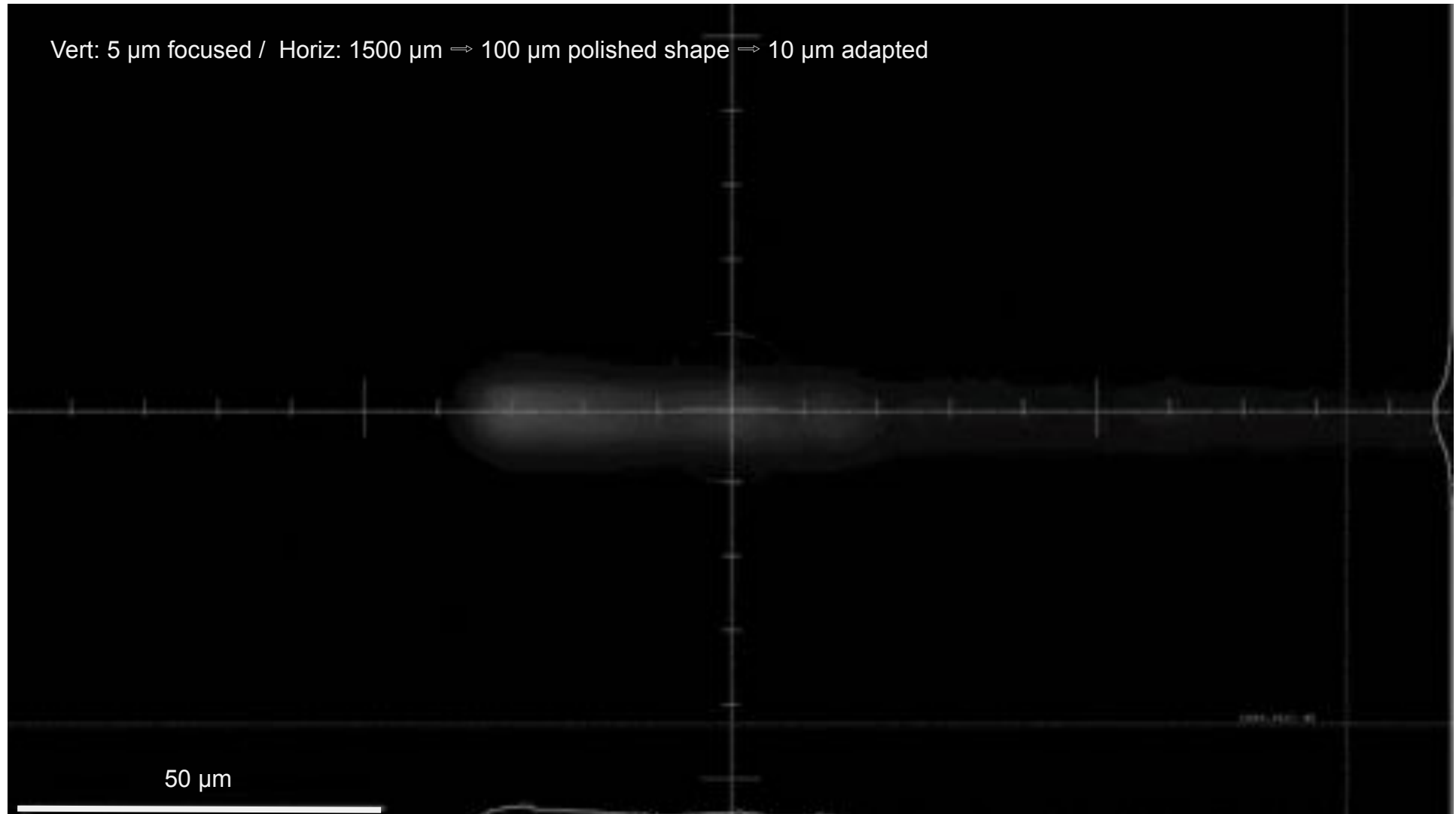


SESO bimorph mirror



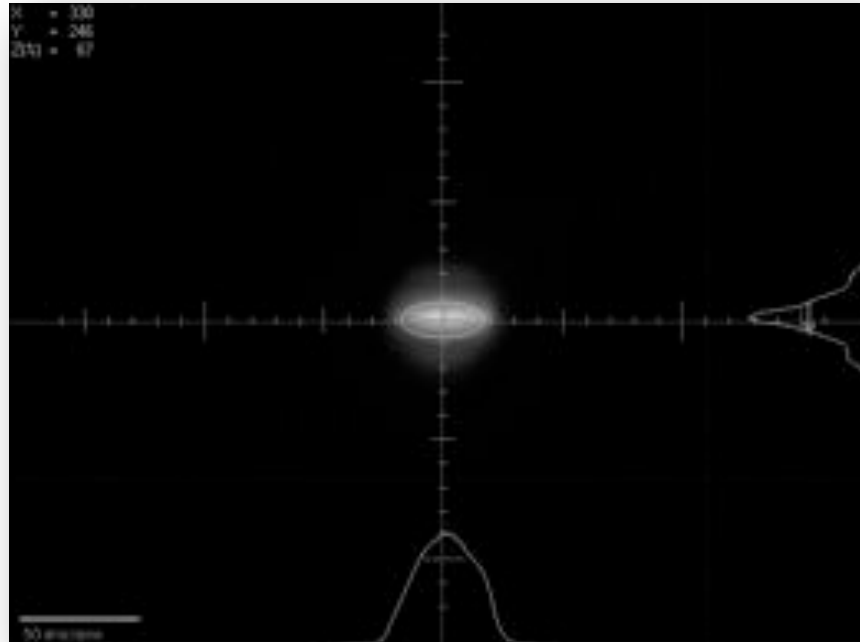
P14: horizontal focusing (11/12/12)

Vert: 5 μm focused / Horiz: 1500 μm \Rightarrow 100 μm polished shape \Rightarrow 10 μm adapted

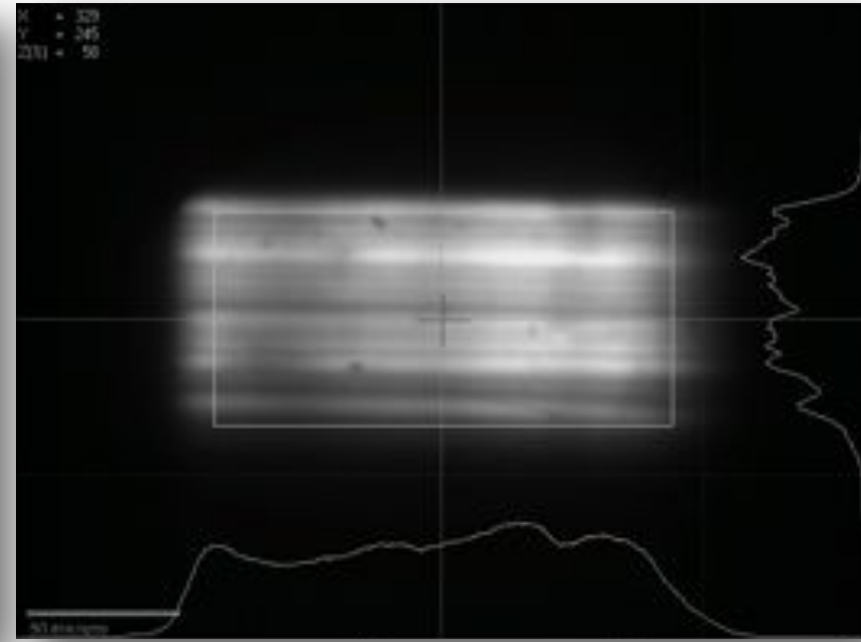


- 10^{12} ph/sec into $7 \times 4 \mu\text{m}^2$

P13: Focused beam I and II



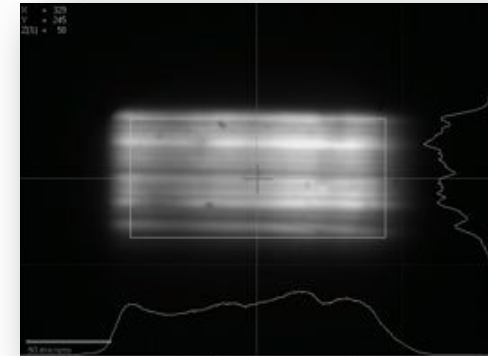
- $H \times V$ (FWHM) = $35 \times 15 \mu\text{m}^2$



- Focused, but detuned
- Shaped with Attocube-slits
- $H \times V$ (FVHM) = $150 \times 70 \mu\text{m}^2$

P13: Focused large beam, ConA [5/2/13]

- Concanavalin A crystal $\sim (50 \mu\text{m})^3$
- 1222, 61.8 Å, 85.1 Å, 89.0 Å, 90°, 90°, 90°
- E = 15 keV, $\lambda = 0.827 \text{ Å}$, $3 \times 10^{12} \text{ ph/sec}$ (no att.)
- 1600 frames @ 0.05 sec / 0.1° = 160° in 80 sec



SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE ≥ -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIB	COMPL OF DATA	R-FACTOR observed	R-FACTOR expected	I/SIGMA	R-meas	CC(1/2)	Anomal	SigAno Corr	Nano
5.08	5743	1858	1876	99.0%	1.8%	1.8%	52.99	2.2%	99.9*	24*	0.974	718
3.60	10368	3354	3373	99.4%	1.9%	2.0%	47.48	2.3%	99.9*	11*	0.900	1432
2.94	13992	4371	4394	99.5%	2.8%	2.7%	35.16	3.3%	99.9*	7	0.853	1965
2.54	15771	5159	5186	99.5%	4.5%	4.6%	21.05	5.5%	99.7*	-2	0.793	2262
2.28	18306	5829	5874	99.2%	7.0%	7.0%	14.39	8.4%	99.4*	12*	0.849	2633
2.08	20072	6463	6488	99.6%	9.2%	9.6%	10.54	11.1%	99.0*	1	0.782	2886
1.92	21413	7019	7080	99.1%	13.7%	13.7%	7.27	16.6%	98.4*	2	0.808	3077
1.80	23807	7521	7582	99.2%	25.5%	25.1%	4.26	30.7%	93.9*	-1	0.795	3416
1.70	23621	7833	8113	96.5%	46.2%	44.9%	2.39	56.0%	78.7*	2	0.784	3386
total	153093	49407	49966	98.9%	5.0%	5.0%	15.66	6.1%	99.9*	3	0.818	21775

Diffraction

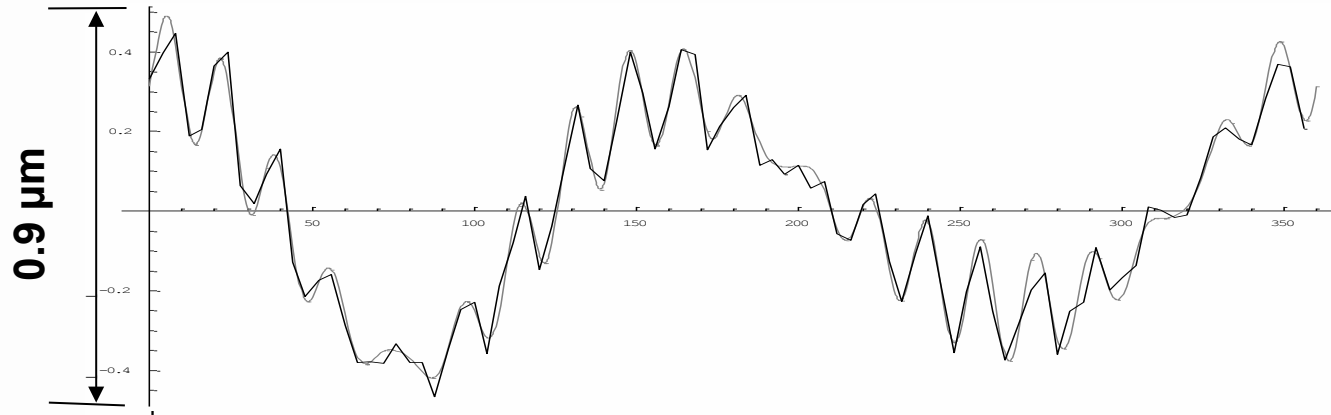


MD2 – measured sample run out

Sample vs spindle position (video microscope + image processing), 1 Turn w/o Kappa

Black line – average of multiple measurements (varying conditions)

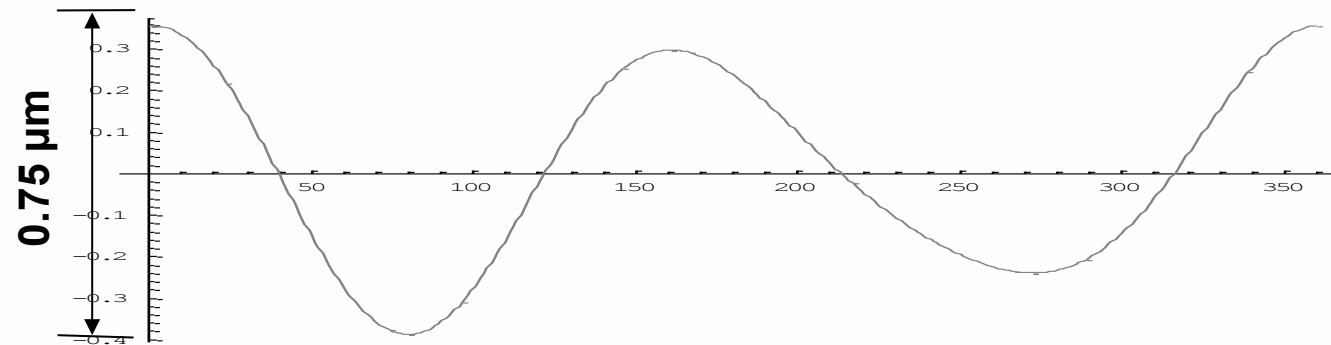
Red lines – synchronous model



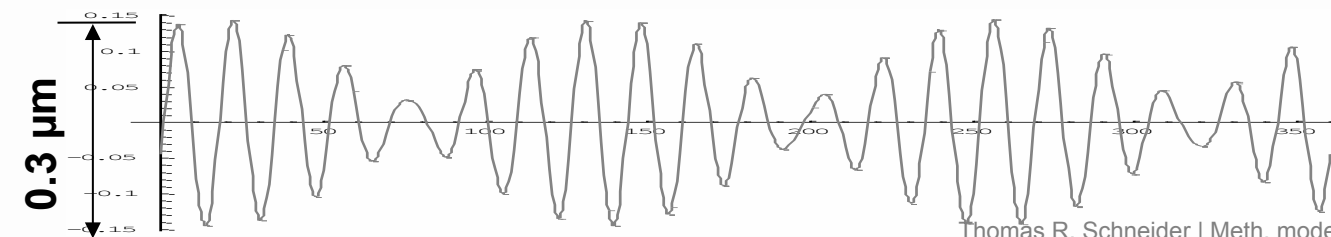
r.m.s.d. = 0.25 μm

peak-to-peak diameter = 0.9 μm

2nd (+3rd) harmonic
Gravity effect on the centering table

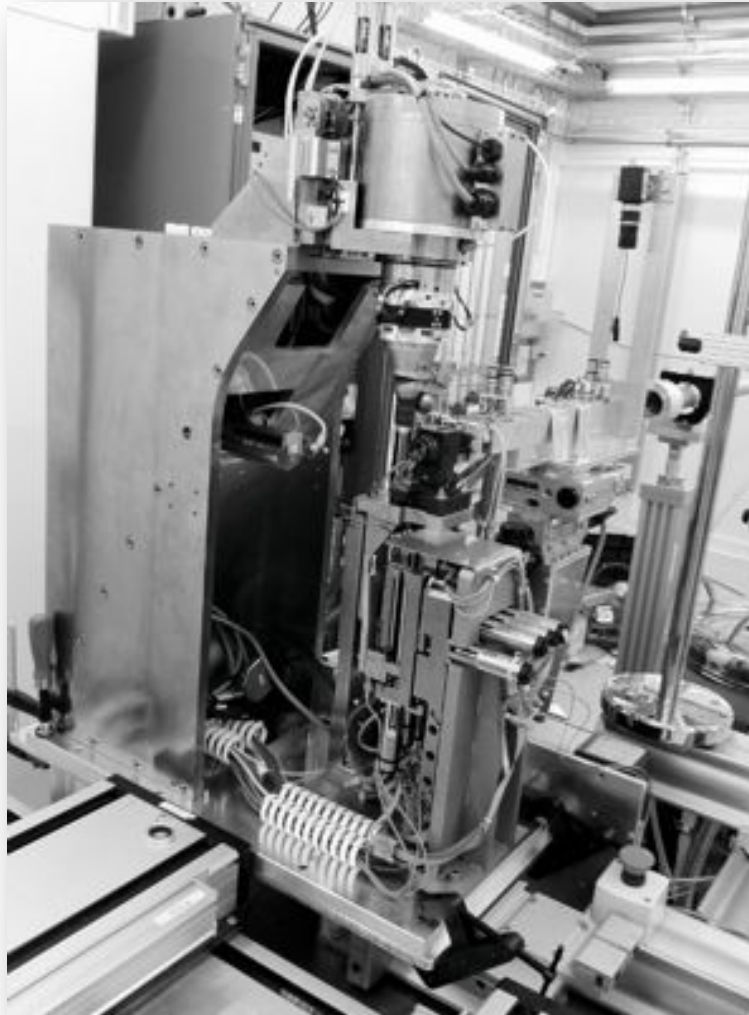


2.5 μm with Kappa

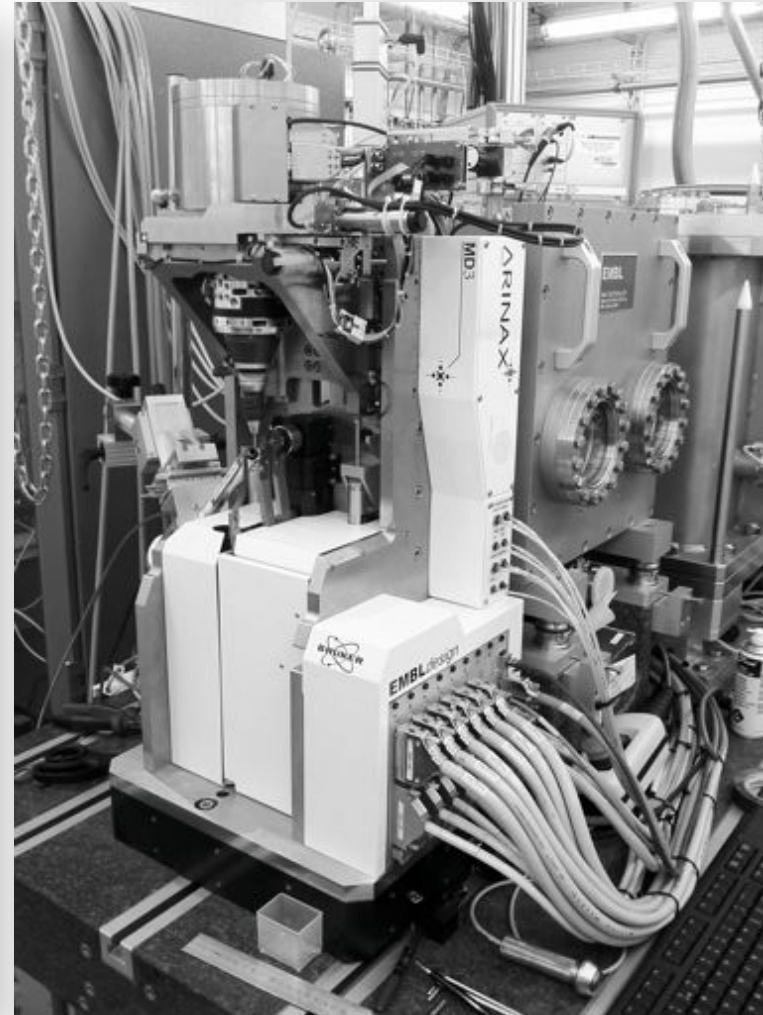


20th (+23rd) harmonic
Poles of the OMEGA motor

Vertical Diffractometry



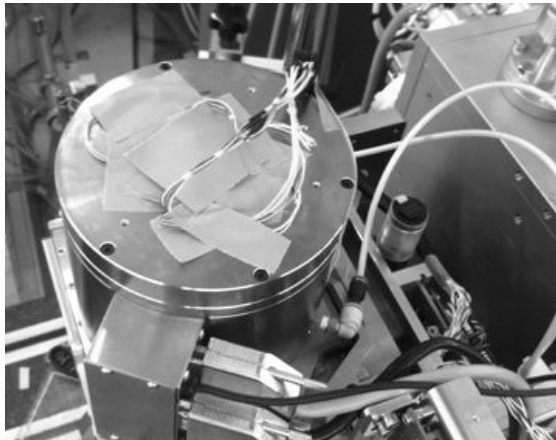
HPGonioV – summer 2011



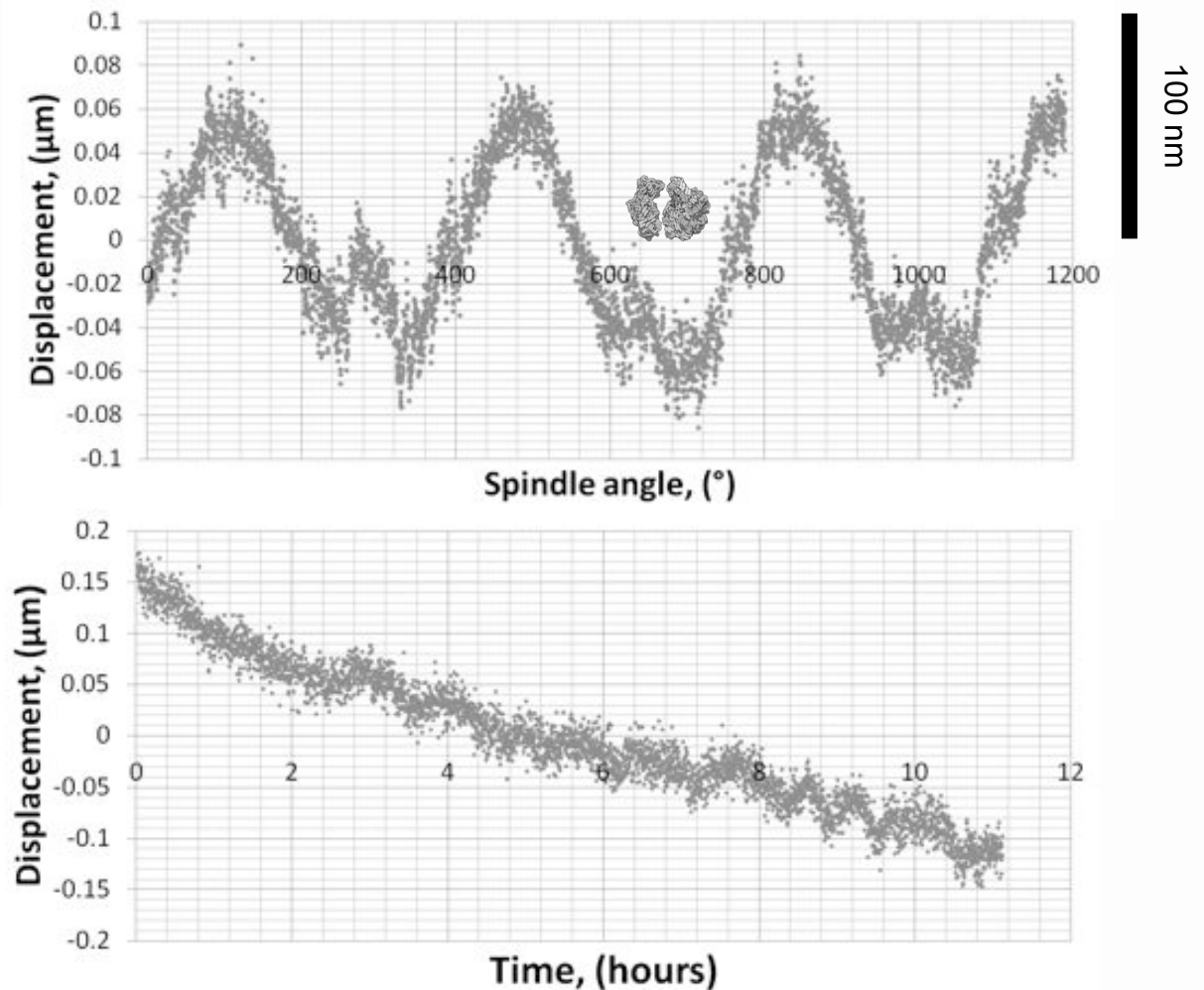
MD3 – 30/10/2012

MD3 [as of 19/01/2013]

- off-center component: 50 nm
- rmsd from ideal: 16 nm



- drift <math>< 0.5 \mu\text{m}</math> / 12 hrs
(was >



First data from MD3

Zn-free Insulin, 10 keV unfocussed, 5/11/12

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE ≥ -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION	NUMBER OF REFLECTIONS			COMPLETENESS	R-FACTOR	R-FACTOR COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal	SigAno	Nano	
LIMIT	OBSERVED	UNIQUE	POSSIBLE	OF DATA	observed	expected				Corr			
4.80	14704	747	747	100.0%	1.9%	2.1%	14704	154.15	1.9%	100.0*	94*	3.780	321
3.40	27439	1350	1350	100.0%	1.8%	2.1%	27439	153.49	1.9%	100.0*	82*	2.267	624
2.78	36345	1753	1753	100.0%	2.1%	2.3%	36345	135.22	2.1%	100.0*	69*	2.103	827
2.41	41700	2036	2036	100.0%	2.5%	2.7%	41700	110.24	2.6%	100.0*	66*	1.922	966
2.16	47670	2363	2363	100.0%	3.2%	3.4%	47670	85.60	3.3%	100.0*	54*	1.578	1130
1.97	51825	2565	2565	100.0%	4.4%	4.6%	51825	63.89	4.5%	100.0*	46*	1.394	1233
1.82	56586	2789	2789	100.0%	7.0%	7.2%	56586	43.06	7.1%	99.9*	26*	1.122	1342
1.70	60983	3051	3051	100.0%	11.0%	11.6%	60983	28.58	11.3%	99.8*	14*	0.970	1473
1.61	44887	3061	3224	94.9%	17.6%	18.4%	44839	15.69	18.2%	99.2*	6	0.835	1453
total	382139	19715	19878	99.2%	2.5%	2.7%	382091	71.28	2.6%	100.0*	36*	1.481	9369

Yb-loaded Thaumatin, 10 keV unfocussed, 6/11/12

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE ≥ -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION	NUMBER OF REFLECTIONS			COMPLETENESS	R-FACTOR	R-FACTOR COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal	SigAno	Nano	
LIMIT	OBSERVED	UNIQUE	POSSIBLE	OF DATA	observed	expected				Corr			
5.13	25296	1931	1931	100.0%	2.6%	2.7%	25296	89.08	2.7%	100.0*	98*	8.469	735
3.64	47111	3491	3492	100.0%	2.5%	2.7%	47110	91.40	2.6%	100.0*	95*	5.418	1510
2.97	61442	4521	4521	100.0%	2.8%	3.0%	61442	80.26	2.9%	100.0*	94*	4.770	2021
2.58	71550	5335	5335	100.0%	3.7%	3.8%	71550	61.80	3.8%	99.9*	91*	4.172	2428
2.30	81067	6034	6034	100.0%	4.7%	4.8%	81067	49.94	4.9%	99.9*	87*	3.453	2780
2.10	90969	6729	6729	100.0%	5.8%	6.1%	90969	40.72	6.1%	99.9*	81*	2.696	3122
1.95	98538	7278	7278	100.0%	7.9%	8.3%	98538	31.50	8.2%	99.9*	73*	2.138	3403
1.82	105584	7822	7822	100.0%	12.0%	13.2%	105584	21.52	12.5%	99.7*	64*	1.712	3669
1.72	74909	7658	8320	92.0%	16.5%	18.3%	74728	13.37	17.4%	99.0*	48*	1.285	3499
total	656466	50799	51462	98.7%	4.1%	4.3%	656284	44.47	4.2%	100.0*	82*	3.032	23167

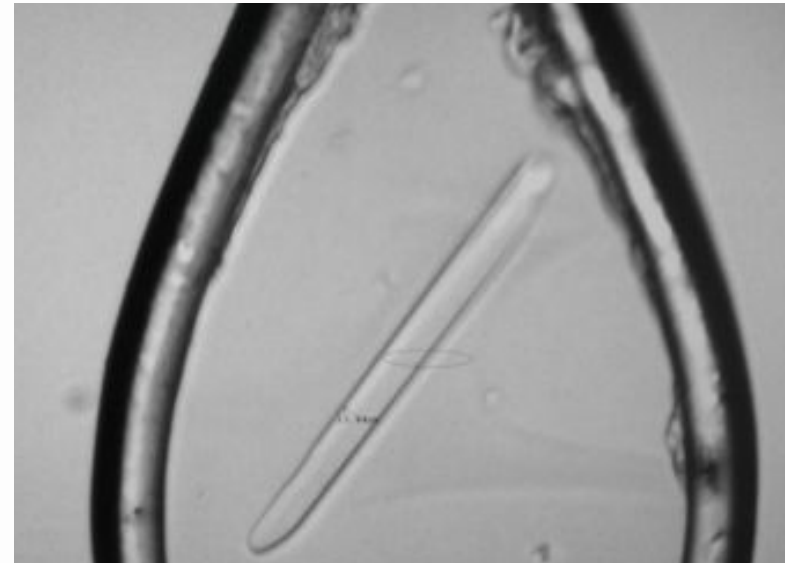
'4D-Scan' (Florent Cipriani, Alexandre Gobbo, EMBL-GR)



- Synchronous movement of omega, centring table and alignment table using direct PMAC programming to collect data along a long needle in a 'continuous helical scan'
- Here: carbon fibre 10 micron thick, 650 micron long.
- Real movement is smooth. Jumps are due to recording.

P14: 1D vs. 4D scan

- 'Monoclinic lysozyme' P2₁
31.5 53.0 72.0 90 99.2 90.0
70% solvent for 129 residues.
crystal 150 x 15 x ?? μm
- 10 keV = 1.24 Å
- half-focussed beam: 10 x 30 micron
- 1D: 0.2 deg / 0.04 sec, total = 36 secs.
4D: 0.1 deg / 0.5 sec, total = 15 min.



SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

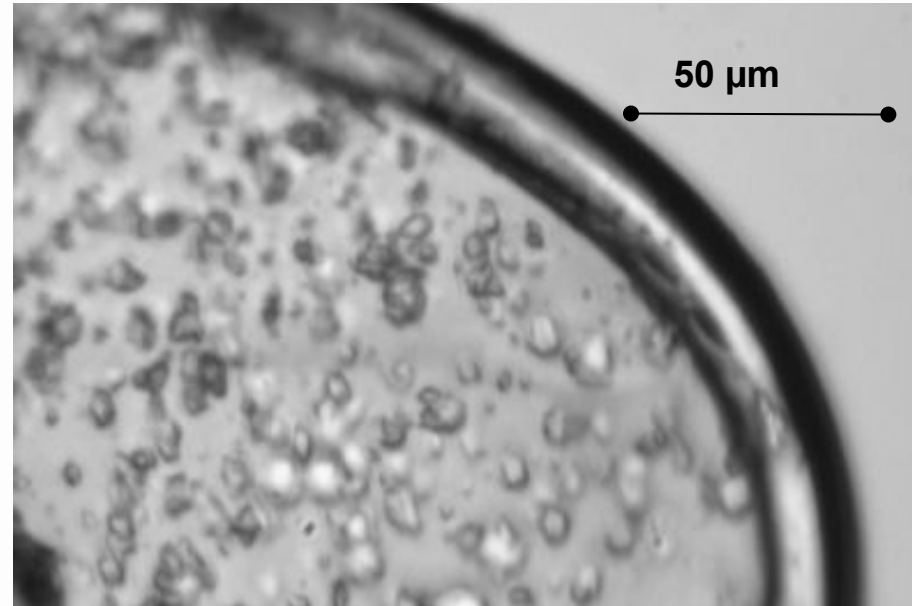
RESOLUTION LIMIT	NUMBER OF REFLECTIONS			COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano	
6.83	1422	427	432	98.8%	5.3%	5.3%	418	18.34	6.4%	99.4*	-11	0.760	238
4.85	2375	723	747	96.8%	7.5%	7.8%	2367	12.69	9.0%	99.1*	-6	0.769	352
3.97	3216	922	938	98.3%	6.6%	6.7%	3207	15.37	7.9%	99.3*	4	0.830	488
3.44	3764	1098	1122	97.9%	8.0%	8.1%	3752	12.68	9.5%	99.1*	-9	0.709	540
3.08	4341	1245	1264	98.5%	12.1%	12.0%	4328	8.93	14.3%	98.1*	-5	0.804	597
2.81	4774	1371	1379	99.4%	19.4%	19.3%	4767	5.93	23.0%	94.8*	-9	0.761	635
2.60	5122	1484	1503	98.7%	25.1%	25.3%	5106	4.44	29.8%	92.7*	-1	0.777	674
2.43	5442	1583	1611	98.3%	32.9%	32.0%	5430	3.59	39.1%	87.8*	-2	0.796	656
2.29	5008	1559	1707	91.3%	43.5%	41.1%	4942	2.61	52.2%	81.9*	-8	0.760	535
total	55464	10412	10703	97.3%	13.2%	13.1%	35317	7.75	15.7%	98.7*	-5	0.775	4715

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS			COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano	
4.77	3927	1208	1239	97.5%	4.0%	4.3%	3907	23.99	4.9%	99.5*	-17	0.701	576
3.38	7081	2107	2165	97.3%	4.3%	4.4%	7054	23.84	5.1%	99.6*	-16	0.744	964
2.76	9386	2737	2770	98.8%	5.7%	5.5%	9360	18.22	6.8%	99.4*	-15	0.763	1197
2.39	10840	3204	3272	97.9%	7.9%	7.8%	10799	13.02	9.4%	99.1*	-9	0.787	1285
2.14	12134	3623	3694	98.1%	11.6%	11.4%	12084	9.37	13.9%	98.1*	-4	0.784	1329
1.96	13387	3969	4078	97.3%	18.4%	19.1%	13328	6.03	21.9%	96.2*	-7	0.732	1404
1.81	14515	4346	4428	98.1%	35.3%	37.3%	14445	3.27	42.1%	88.9*	-9	0.682	1426
1.69	15419	4635	4725	98.1%	61.6%	66.7%	15338	1.89	73.5%	73.4*	-8	0.679	1430
1.60	10819	3945	5056	78.0%	82.1%	92.7%	10552	1.14	101.7%	52.5*	-9	0.645	791
total	97508	29774	31427	94.7%	8.3%	8.5%	96867	8.60	9.9%	99.6*	-9	0.728	10402

Microcrystal data collection

Ultralente Insulin in H3
a=b=80 Å c=37 Å
crystal: (3-5) x (3-5) x (5-10) μm³
beam: 4x5 μm ('half'-focused)
exp: 1 s per 1° frame
~40 frames/crystal
Pilatus6MF

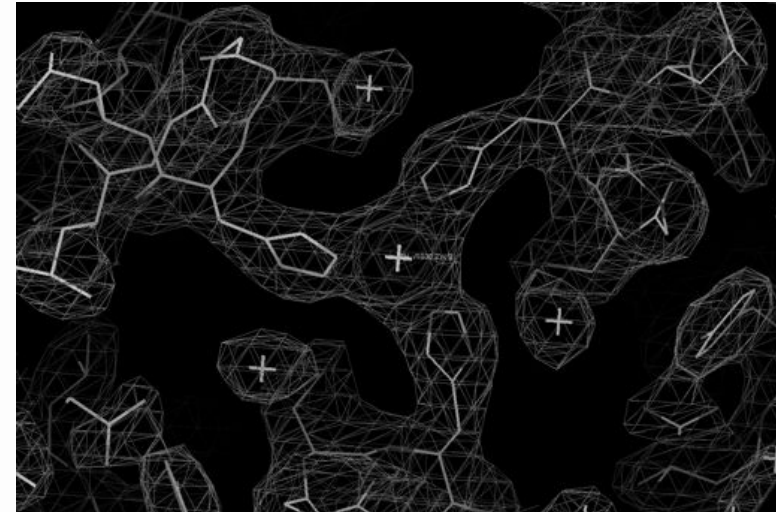
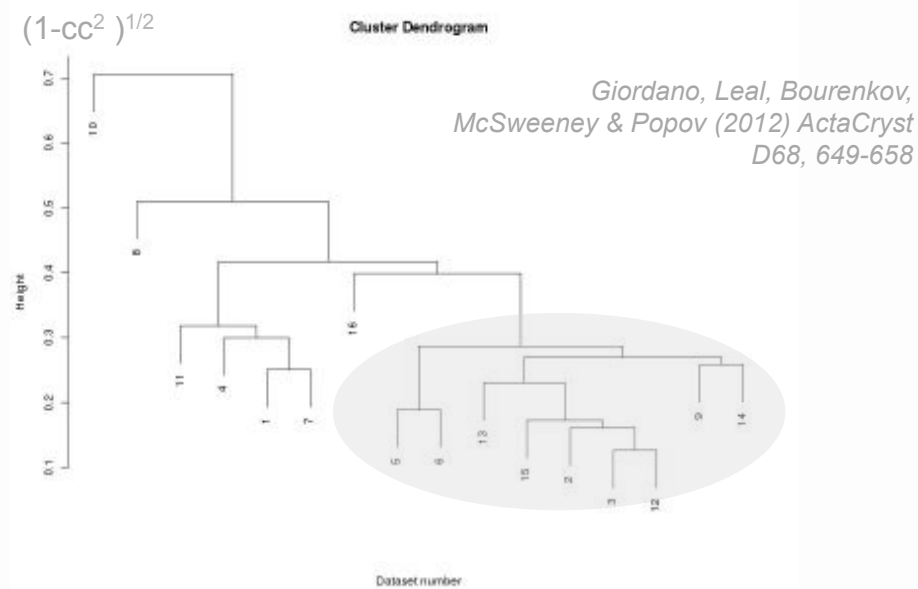


Partial data set statistics

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE \geq -3.0 AS FUNCTION OF RESOLUTION												
RESOLUTION	NUMBER OF REFLECTIONS			COMPLETENESS	R-FACTOR	R-FACTOR COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal		
LIMIT	OBSERVED	UNIQUE	POSSIBLE	OF DATA	observed	expected				Corr		
8.34	86	78	168	46.4%	3.3%	3.3%	16	15.89	4.7%	99.6*	0	
5.28	272	253	478	52.9%	4.5%	3.8%	38	13.53	6.4%	98.7*	0	
3.93	540	496	960	51.7%	3.2%	3.6%	88	13.06	4.5%	99.4*	0	
3.27	662	599	1154	51.9%	5.2%	5.1%	126	9.47	7.3%	99.1*	0	
2.95	587	543	1012	53.7%	9.5%	8.7%	88	6.10	13.4%	97.8*	0	
2.65	591	531	1420	37.4%	12.8%	12.7%	120	4.40	18.1%	97.0*	0	
total	2738	2500	5192	48.2%	4.5%	4.5%	476	8.99	6.3%	99.5*	56	

Ultralente insulin microcrystal structure solution

Isomorphous subset via cluster analysis



Electron density map, model phases
REFMAC R/Rfree=0.18/0.24

Combined data set statistics

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE ≥ -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS			COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano	
	OBSERVED	UNIQUE	POSSIBLE										
8.34	1989	168	168	100.0%	8.2%	9.3%	1987	25.91	8.6%	99.7*	61*	1.800	83
5.28	6054	486	486	100.0%	11.2%	11.5%	6053	21.13	11.7%	99.4*	52*	1.548	242
3.93	11427	964	964	100.0%	13.5%	12.5%	11424	19.50	14.1%	99.1*	38*	1.227	479
3.27	14393	1182	1182	100.0%	22.6%	21.8%	14393	12.74	23.6%	99.2*	16*	0.967	591
2.95	12132	987	988	99.9%	43.8%	48.5%	12132	7.79	45.7%	97.5*	14	0.840	493
2.65	12892	1341	1442	93.0%	68.5%	82.9%	12822	4.28	72.0%	89.8*	-2	0.708	617
total	58887	5128	5230	98.0%	17.4%	17.7%	58811	12.07	18.1%	99.5*	23*	1.012	2505

Refinement

A 'good' model should ...

- Agree with the data

$$R = \frac{\sum_{hkl} |F_{hkl}^{obs} - F_{hkl}^{calc}|}{\sum_{hkl} |F_{hkl}^{obs}|}$$

- Agree with what we know

Progress of a Refinement

#par is number of parameter in model

#obs is number of reflections used in refinement

Rw is Rwork in percent

Rf is Rfree in percent

Rd is the difference between Rw and Rf in percent

CPU is approximate CPU time [secs] normalized to a Pentium III running at 500 MHz

step	data used	#par	#obs	Rw	Rf	Rd	CPU
molecular replacement	15.0-4.0	3	775	44.5	-	-	30
rigid body	10.0-2.5	9	2997	46.4	47.4	1.0	140
first round	10.0-1.5	3887	13818	30.2	34.3	4.1	500
after first rebuild 10cgl	"	3735	"	26.2	31.4	5.2	360
SHELXWAT	"	4091	"	20.2	24.6	4.4	2000
bld + SHELXWAT	"	4231	"	18.7	22.8	4.1	2000
include all data 20cgl	10.0-1.1	4203	33993	19.1	21.7	2.6	1450
ANIS 20cgl	"	9453	"	16.0	19.2	3.2	2850
Rebuild 10 SHELXWAT	"	9557	"	13.4	15.8	1.8	1480
rebuild 10 cgl	"	10481	"	12.3	15.3	3.0	1610
rebuild 10 cgl	"	10819	"	12.1	15.1	3.0	1650
rebuild 10 cgl	"	10838	"	11.6	14.6	3.0	1800
rebuild 10 cgl	"	11494	"	11.3	14.4	3.1	1800
rebuild 10 cgl	"	11576	"	11.0	14.0	3.0	1800
rebuild 10 cgl	"	11774	"	10.7	13.8	3.1	1800
put hydrogens	10.0-1.1	11765	"	9.7	12.8	3.1	2500
include test set	10.0-1.1	11693	35786	9.8	-	-	2600

Target Function

- Implements our idea of agreement between the data and the measurements. One possible form is:

$$E_{X-ray} = \sum_{hkl} \left| F_{hkl}^{obs} - k F_{hkl}^{calc} \right|^2$$

- One could replace F_{obs} by I_{obs} .
- The hkl -terms can be weighted by the uncertainties of the measurements: $1/\sigma(F_{obs})$.
- Other functions, such as the $CC(F_{obs}, F_{calc})$, can be used.

Observables and Parameters

res	#r/a	parameters	d:p
3.0	2	x y z	0.6:1
2.5	4	x y z B?	1:1
2.0	8	x y z B	2:1
1.5	20	x y z B	5:1
1.1	50	x y z U_{11} U_{22} U_{33} U_{12} U_{13} U_{23}	5:1
0.9	90	x y z U_{11} U_{22} U_{33} U_{12} U_{13} U_{23}	10:1

Numbers are for a protein crystal with 40% solvent.
 Lysozyme (129 amino acids 1001 atoms) will produce
 ca. 32000 reflections to 2 Å resolution.

Stereochemical Restraints

Acta Cryst. (1991). A47, 392-400

Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement

BY RICHARD A. ENGH AND ROBERT HUBER

Max Planck Institut für Biochemie, D8033 Martinsried bei München, Germany

Table 1. CSD parametrization atom types

Atom type	Description
C	Carbonyl C atom of the peptide backbone
C5W*	Tryptophan C ^γ
CW*	Tryptophan C ^{α2} , C ^{γ2}
CF*	Phenylalanine C ^γ
CY*	Tyrosine C ^γ
CY2*	Tyrosine C ^γ
C5*	Histidine C ^γ
CN*	Neutral carboxylic acid group C atom
CH1E	Tetrahedral C atom with one H atom
CH2E	Tetrahedral C atom with two H atoms (except CH2P, CH2G)
CH2P*	Proline C ^γ , C ^δ
CH2G*	Glycine C ^α
CH3E	Tetrahedral C atom with three H atoms
CR1E	Aromatic ring C atom with one H atom (except CR1W, CRH, CRHH, CR1H)
CR1W*	Tryptophan C ^{α2} , C ^{γ2}
CRH*	Neutral histidine C ^{α1}
CRHH*	Charged histidine C ^{α1}
CR1H*	Charged histidine C ^{α2}
N	Peptide N atom of proline
NR	Unprotonated N atom in histidine
NP	Pyrrrole N atom
NH1	Singly protonated N atom (His, Trp, peptide)
NH2	Doubly protonated N atom
NH3	Triply protonated N atom
NC2	Arginine N ^{η1} , N ^{η2}
O	Carbonyl O atom
OC	Carboxyl O atom
OH1	Hydroxyl O atom
S	S atom
SM*	Methionine S atom
SH1E	Singly protonated S atom

* Atom types marked with an asterisk are new (non-XPLOR) types.

Table 2. Bond parameters

Bond type	σ	Bond length (Å)
C5W-CW	0-018	1-433
CW-CW	0-017	1-409
C-CH1E	0-021	1-525
C5-CH2E	0-014	1-497
C5W-CH2E	0-031	1-498
CF-CH2E	0-023	1-502
CY-CH2E	0-022	1-512
C-CH2E		
CN-CH2E		
C-CH2G		
C5W-CR1E		
CW-CR1E		
CW-CR1W		
CF-CR1E		
CY-CR1E		
CY2-CR1E		
C5-CR1H		
C5-CR1E		
C-N		
C-NC2		
C5-NH1		
CW-NH1		

Table 3. Angle parameters

Angle type	σ	Angle (°)	Angle type
C5W-CW-CW	1-2	107-2	CH3E-C
CW-C5W-CH2E	1-4	126-8	CH3E-C
C5W-CW-CR1E	1-0	133-9	CH3E-C
CW-CW-CR1E	1-0	118-8	C-CH2E
CW-CW-CR1W	1-0	122-4	C5-CH2
CW-C5W-CR1E	1-6	106-3	CF-CH2
CW-CW-NH1	1-3	107-4	C5W-CH
CH1E-C-N	1-5	116-9	CY-CH2
CH1E-C-NH1	2-0	116-2	C-CH2E
CH1E-C-O	1-7	120-8	C-CH2C
CH1E-C-OC	2-5	117-0	C-CH2C
CH2E-C5-CR1E	1-3	129-1	CH1E-C
CH2E-C5-CR1H	1-3	131-2	CH1E-C
CH2E-CF-CR1E	1-7	120-7	CH1E-C
CH2E-C5W-CR1E	1-5	126-9	CH1E-C
CH2E-CY-CR1E	1-5	120-8	CH1E-C
CH2E-C-N	2-1	118-3	CH1E-C

Bond lengths and angles should be consistent with these values ('Engh&Huber Dictionary') derived from small molecule structures. ITC-F contains an updated version (§18.3).

Enforcing stereochemistry

$$E_{bond} = \sum_{bonds} \frac{|d^{obs} - d^{EH}|^2}{\sigma_{d^{EH}}}$$

- E_{bond} is zero for a perfect agreement between expected bond lengths and bond lengths measured in the model. The value of E_{bond} increases for increasing deviations.
- The weight for each bond-type depends on the spread observed in small molecule structures

Enforcing stereochemistry

- Target functions can be defined for various geometrical properties:

$$E_a = \sum_a \frac{|a^{obs} - a^{ex}|^2}{\sigma_{a^{ex}}}$$

- And combined into an overall function:

$$E_{chem} = E_{bond} + E_{angle} + E_{chir} + \dots$$

Combined Target Function

$$E_{tot} = E_{chem} + w_{X-ray} E_{X-ray}$$

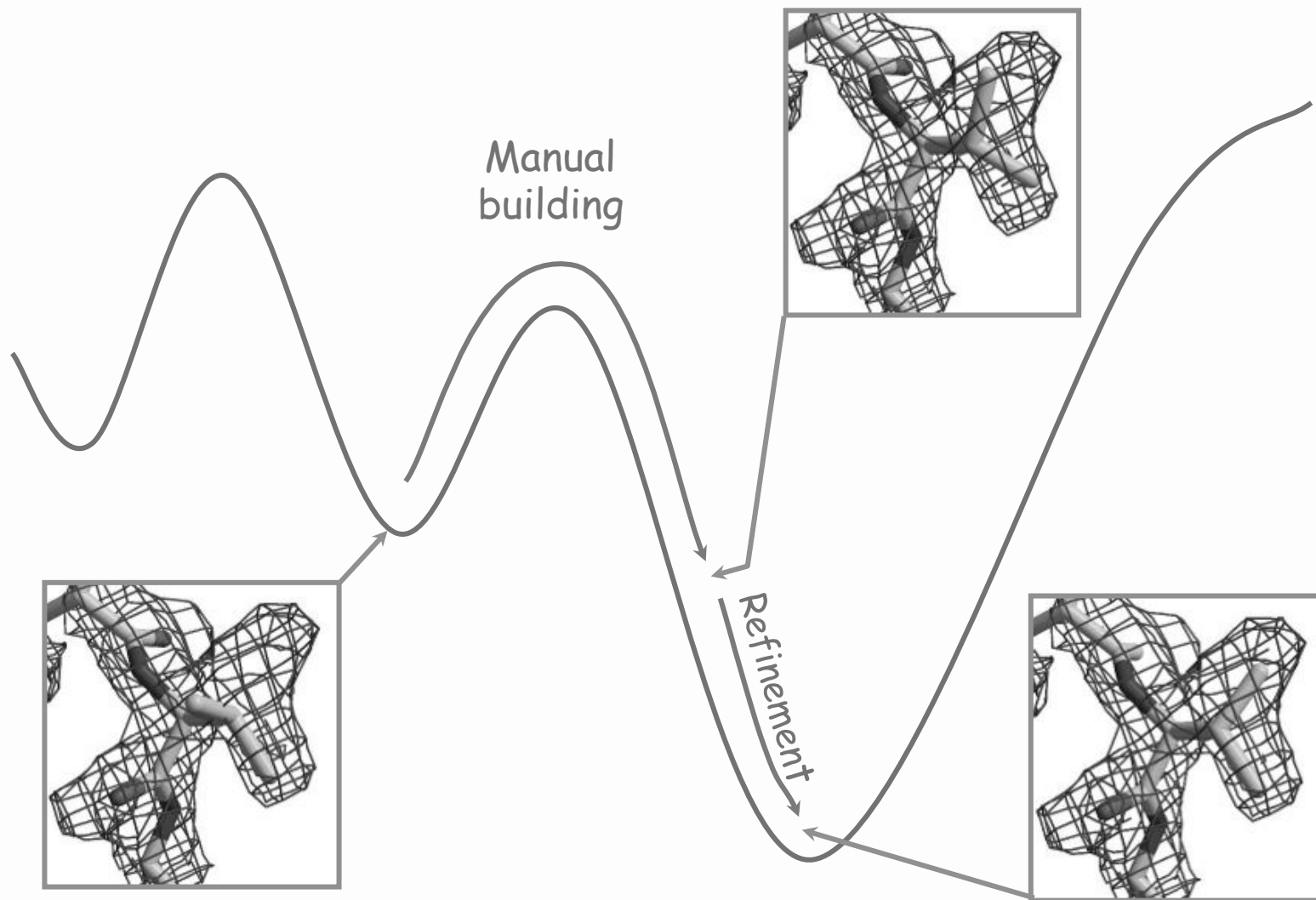
- E_{chem} contains information about bond-lengths, -angles, planarity, chirality, non-bonded repulsion, electrostatic interactions etc. (the *a priori* knowledge).
- w_{X-ray} is the relative weight between chemical information and experimental data. It depends not only the amount of data but also on the quality of the data.

Optimization of the Target Function

$$E_{tot} = E_{chem} + w_{X-ray} E_{X-ray}$$

- Minimization of E_{tot} should deliver a model that (1) is consistent with the data and (2) has reasonable stereochemistry.
- At 2 Å, 5000 atoms correspond to ca. 20000 parameters to be refined against 50000 reflections.

Getting out of Local Minima



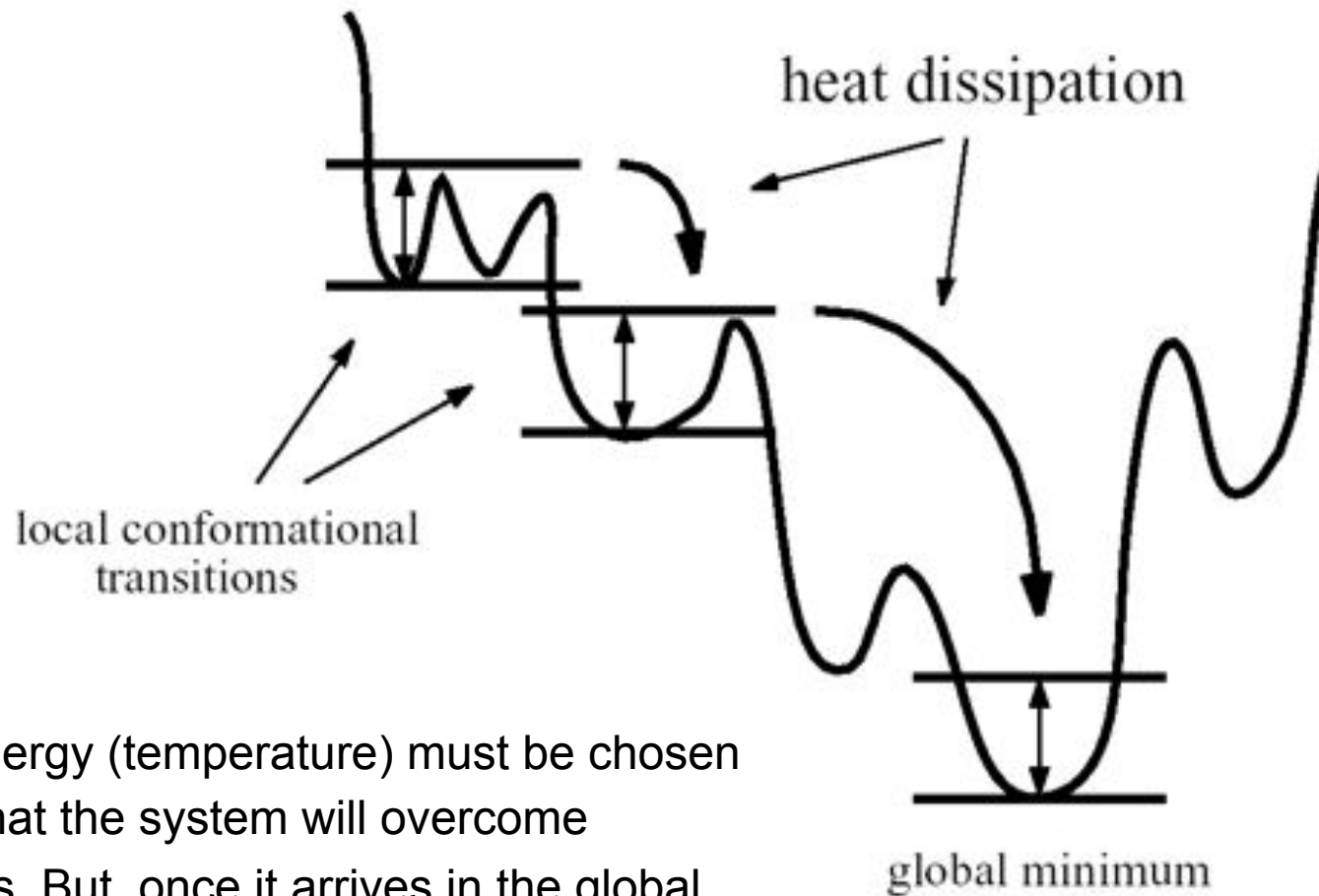
Simulated Annealing



- A method to search conformational space
- *Annealing is a physical process wherein a solid is heated so that the particles assume random positions and then is cooled slowly so that all particles can arrange into their lowest energy state.*
- The target function E can be considered as the potential energy of the system and then a molecular dynamics simulation is run (in fact, historically, XPLOR (later CNS, later PHENIX) was an MD-program with an additional energy term $E_{X\text{-ray}}$).

Simulated Annealing

Figure from
Paul Adam's lecture notes



The energy (temperature) must be chosen such that the system will overcome barriers. But, once it arrives in the global minimum, it should not escape from there.

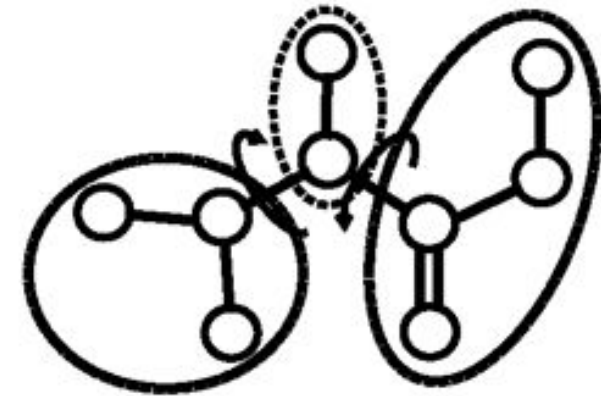
Observables and Parameters

res	#r/a	parameters	d:p
3.0	2	x y z	0.6:1
2.5	4	x y z B?	1:1
2.0	8	x y z B	2:1
1.5	20	x y z B	5:1
1.1	50	x y z U_{11} U_{22} U_{33} U_{12} U_{13} U_{23}	5:1
0.9	90	x y z U_{11} U_{22} U_{33} U_{12} U_{13} U_{23}	10:1

Constraints

- Constraints couple parameters to each other such that the overall number of parameters is reduced.
- Segmented rigid bodies
 - Entire residues
 - Rigid groups (guanodinio, carboxylate, aromatic rings)
 - Entire domains

Torsion Angle Refinement



- Deviations in bond lengths and angles are usually small.
- Torsion angles can be used as ‘normal coordinates of protein dynamics’
- This reduces the number of parameters dramatically (up to a factor of 10)
- Initially introduced by Diamond (1971); implemented in XPLOR by Rice & Brunger (1996).
- Large radius of convergence (e.g. for bad molecular replacement models).

Diamond (1971) *Acta Cryst.* A27:436-452
Sussmann (1977) *Acta Cryst.* A33:800
Rice & Brunger (1996) *Proteins* 19:277-290

B-factors

- A 'crystal structure' represents a space and time average (many molecules in the crystal for the time of the experiment).
- The coordinates of an atom in the pdb-file represent the mean coordinates. The spread around the mean coordinates is described by the B-factor.
- Some Examples:

$$\begin{array}{ll} B = 5 \text{ \AA}^2 & \langle u \rangle = 0.25 \text{ \AA} \\ B = 20 \text{ \AA}^2 & \langle u \rangle = 0.50 \text{ \AA} \\ B = 40 \text{ \AA}^2 & \langle u \rangle = 0.71 \text{ \AA} \\ B = 80 \text{ \AA}^2 & \langle u \rangle = 1.01 \text{ \AA} \end{array}$$

- A typical bondlength is 1.5 Å !

$$B = 8\pi^2\langle u^2 \rangle,$$

where $\langle u^2 \rangle$ is the mean square displacement

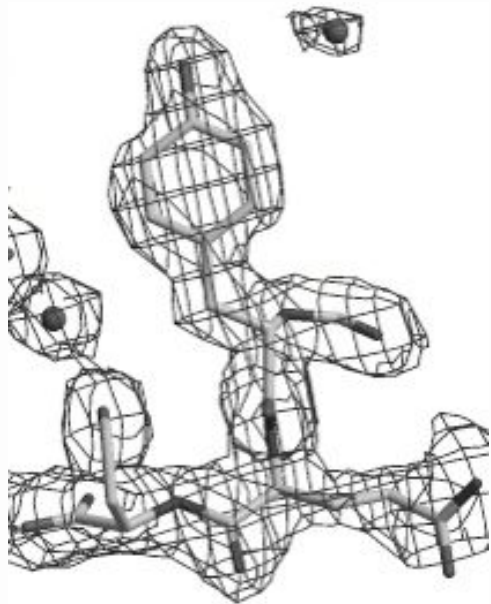
B-factors

- Low
 - In well ordered region
- High
 - In loops
 - On the surface
 - In terminal regions
- Here
 - Red is high
 - Blue is low



B-factors and electron density

All pictures were taken from the same structure refined at 2.0 Å resolution.
Countoured at the same level



Tyr48
 $B = 20.9 \text{ \AA}^2$
RSR = 0.12



Met285/Asn286
 $B = 69.9 \text{ \AA}^2 / 69.1 \text{ \AA}^2$
RSR = 0.25 / 0.18



Glu229/Asp230
 $B = 73.9 \text{ \AA}^2 / 73.4 \text{ \AA}^2$
RSR = 0.39 / 0.27

Summary

- Achievable resolution depends (1) on the quality of the crystal (2) on the signal/noise ratio enabled by the experimental setup.
- On PETRA III, high-flux & flexible beams with diameters in the μm range can be produced.
- X-ray Optics and mechanics are accurate in the sub- μm range.
- Refinement is the process of improving the fit of a model to the experimental data. Over-fitting can be a problem.
- Free electron lasers have opened new possibilities: serial crystallography on μm -sized crystals.
- The 'diffraction-before-destruction' paradigm has been verified experimentally.