Workshop on Advanced Data Collection with Multi-Axis Goniometry
MADaC 2015

Synchrotron SOLEIL - L'orme des Merisiers
Saint-Aubin, France
12th-14th of November 2015

The use of multi-axis goniometers for phasing of glycoprotein crystals

Thomas Krey

Institut Pasteur, Paris, France, and Hannover Medical School, Germany

Acknowledgements

Institut Pasteur, Paris, France

Unite Virologie Structurale

Felix Rey

Jimena Perez-Vargas

Ieva Vasiliauskaite

Juliette Fedry

Marie-Christine Vaney

Stephane Duquerroy

Pablo Guardado-Calvo

Platform of Crystallization

Ahmed Haouz

Patrick Weber

SOLEIL, Paris, France

Pierre Legrand

Andrew Thompson

William Sheperd

SLS, Villigen, Switzerland

Vincent Olieric

Global Phasing, Cambridge, UK

Gerard Bricogne

Clemens Vonrhein

What kind of proteins do we usually work with ?

- soluble ectodomains of glycoproteins of viral or cellular origin involved in membrane fusion
- stabilized by a number of disulfide bridges !
- proteins are produced in insect cells (production of SeMet protein difficult - incorporation rate ~50-70% + lower yields !)
- X-tals diffract to intermediate/low resolution
- X-tals often difficult to reproduce (batch-to-batch variation !)

What kind of proteins do we usually work with ?

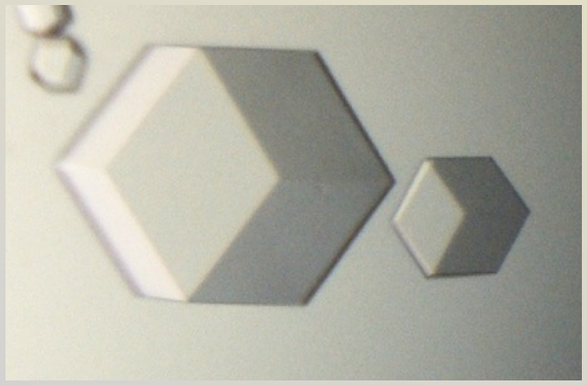
- soluble ectodomains of glycoproteins of viral or cellular origin involved in membrane fusion
- stabilized by a number of disulfide bridges !
- proteins are produced in insect cells (production of SeMet protein difficult - incorporation rate ~50-70% + lower yields !)
- X-tals diffract to intermediate/low resolution
- X-tals often difficult to reproduce (batch-to-batch variation !)

↳ experimental phasing is usually challenging and requires specific considerations !

Case I

EFF-1 - a C.elegans cell-cell fusion protein

(Perez-Vargas et al., 2014)



- Cubic space group ($I2_13$)
- best native diffraction to ~ 3 Å
- most likely one molecule per AU - i.e., no NCS

Two heavy atom derivatives

XDS/XSCALE

1. Au-derivative (Diffraction to ~ 3.9 Å, anomalous signal to ~ 4.6 Å)

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

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
17.45	734	147	170	86.5%	2.7%	3.4%	729	44.60	3.0%	99.9*	98*	6.934	59
12.34	1700	314	315	99.7%	2.8%	3.5%	1698	44.70	3.1%	99.9*	97*	6.167	138
10.07	2198	397	397	100.0%	3.0%	3.5%	2197	44.53	3.3%	99.9*	95*	4.809	177
8.72	2561	478	478	100.0%	3.4%	3.6%	2558	40.78	3.8%	99.9*	92*	4.280	219
7.80	2949	540	542	99.6%	4.2%	4.1%	2947	34.69	4.7%	99.8*	88*	3.714	248
7.12	3342	595	595	100.0%	4.7%	4.9%	3341	29.45	5.2%	99.8*	85*	2.938	276
6.59	3573	632	632	100.0%	6.2%	6.3%	3572	23.37	6.8%	99.7*	81*	2.655	296
6.17	3984	697	697	100.0%	7.3%	7.4%	3984	20.11	8.0%	99.6*	71*	2.011	331
5.82	4280	744	744	100.0%	9.5%	9.7%	4280	16.34	10.5%	99.3*	62*	1.711	354
5.52	4435	771	771	100.0%	9.8%	10.0%	4435	15.70	10.8%	99.3*	50*	1.420	364
5.26	4829	840	840	100.0%	11.2%	11.5%	4829	13.97	12.4%	99.2*	36*	1.152	402
5.04	4922	854	854	100.0%	13.8%	13.8%	4922	11.82	15.2%	98.7*	23*	1.052	405
4.84	5152	894	895	99.9%	14.4%	14.4%	5151	11.28	15.8%	98.8*	22*	1.001	429
4.66	5189	905	906	99.9%	18.2%	18.4%	5187	9.26	20.0%	98.0*	17*	0.930	433
4.50	5548	961	962	99.9%	24.7%	25.0%	5548	6.91	27.2%	96.6*	18*	0.906	460
4.23	5970	1034	1034	100.0%	45.0%	46.5%	5970	3.88	49.6%	88.4*	6	0.787	498
4.11	6015	1050	1051	99.9%	53.8%	53.7%	6013	3.30	59.3%	86.7*	2	0.751	504
4.00	6365	1100	1100	100.0%	81.9%	82.4%	6365	2.21	90.2%	73.1*	-2	0.762	529
3.90	6249	1092	1094	99.8%	111.3%	112.6%	6245	1.58	122.7%	61.6*	8	0.776	525
total	85877	15061	15093	99.8%	8.3%	8.7%	85853	14.00	9.2%	99.9*	61*	1.607	7137

Data collected on a single crystal using inverse beam strategy with wedges of 10 degrees.

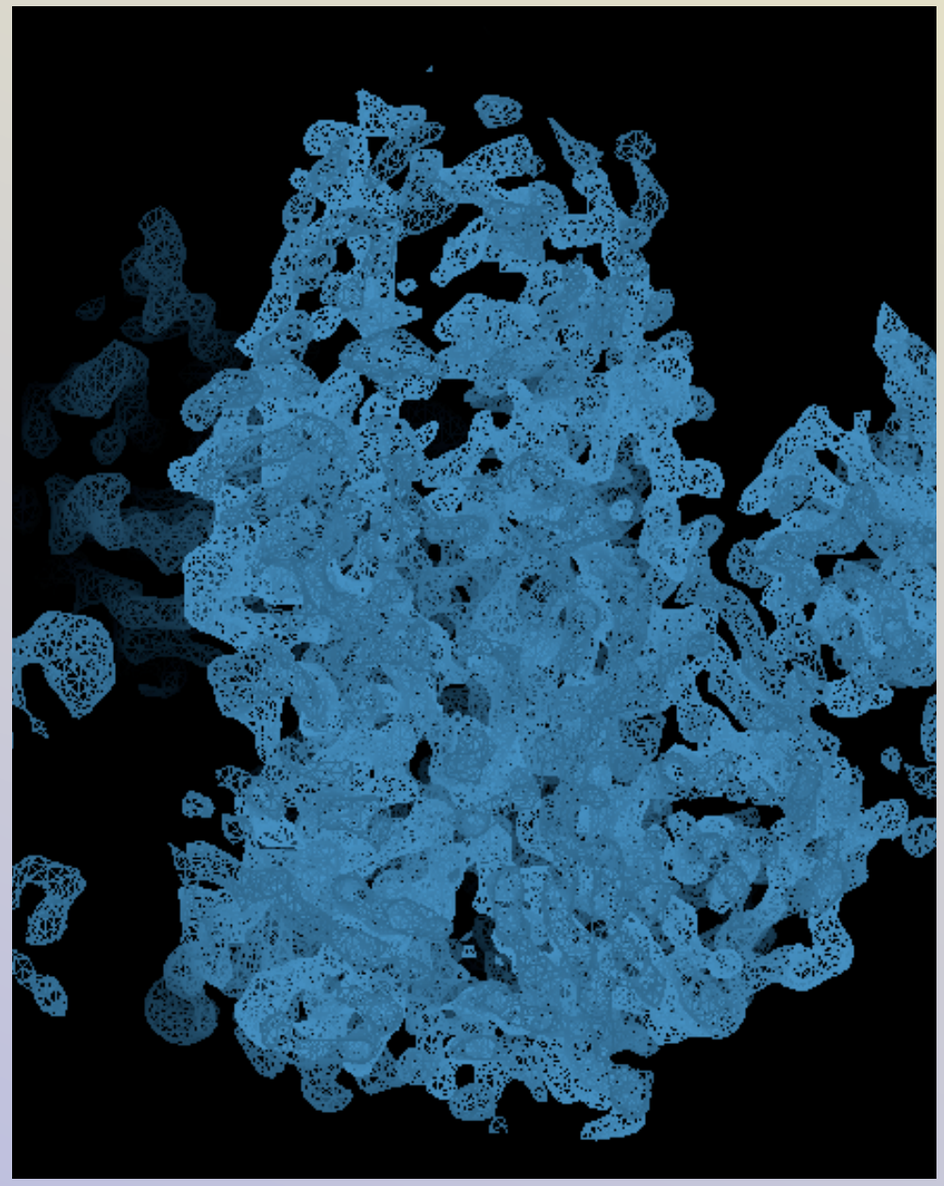
1. Yb-derivative (Diffraction to ~ 4.6 Å, anomalous signal to ~ 6 Å)

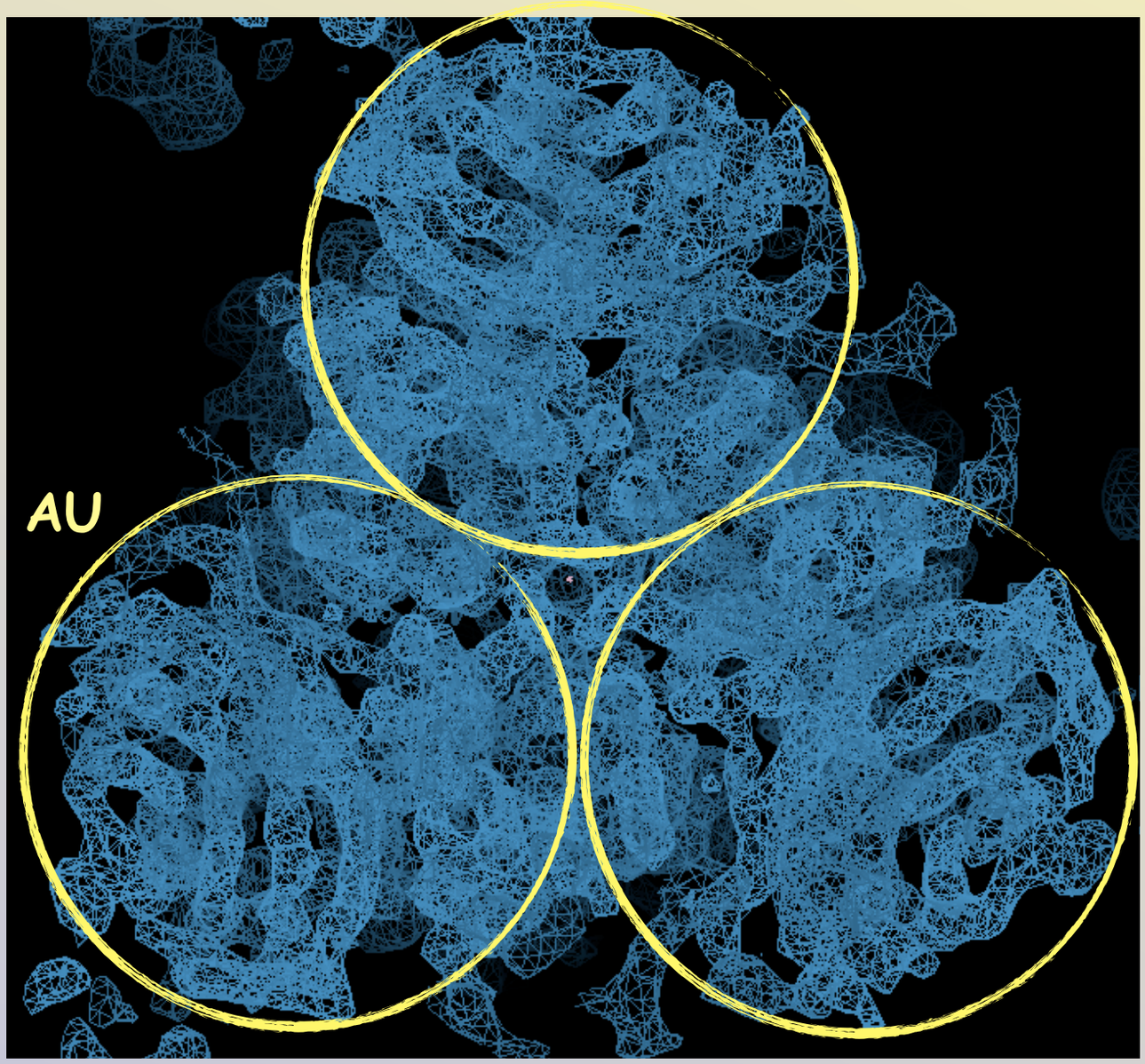
SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
20.60	524	99	109	90.8%	4.3%	5.8%	524	27.60	4.8%	99.8*	97*	6.355	39
14.56	1092	190	190	100.0%	4.1%	6.0%	1092	27.71	4.5%	99.9*	95*	5.159	81
11.89	1441	248	248	100.0%	4.5%	5.9%	1441	27.92	5.0%	99.6*	92*	3.772	110
10.30	1591	286	288	99.3%	4.7%	5.9%	1587	26.63	5.2%	99.7*	90*	3.398	126
9.21	1775	326	327	99.7%	6.2%	6.1%	1775	24.88	6.9%	99.4*	79*	2.886	148
8.41	2048	353	354	99.7%	6.6%	6.6%	2045	22.38	7.2%	99.6*	79*	2.654	159
7.79	2311	400	400	100.0%	7.6%	7.7%	2310	18.42	8.3%	99.3*	72*	2.090	185
7.28	2514	421	421	100.0%	9.1%	9.3%	2512	15.60	10.0%	99.4*	63*	1.808	196
6.87	2706	453	454	99.8%	12.7%	12.8%	2699	11.57	13.9%	99.0*	55*	1.472	205
6.51	2732	459	460	99.8%	16.6%	17.1%	2726	9.14	18.2%	98.0*	48*	1.263	210
6.21	2973	499	499	100.0%	19.3%	19.9%	2966	8.05	21.1%	97.8*	30*	1.041	230
5.71	3380	558	558	100.0%	32.7%	34.5%	3374	5.00	35.9%	93.8*	18	0.876	258
5.50	3388	564	565	99.8%	39.3%	39.3%	3384	4.33	43.1%	93.2*	16	0.890	263
5.32	3611	598	601	99.5%	51.5%	52.3%	3604	3.33	56.5%	88.6*	-6	0.680	279
5.15	3466	578	583	99.1%	63.7%	65.5%	3462	2.78	69.9%	85.2*	12	0.736	270
5.00	3828	638	642	99.4%	76.9%	80.7%	3821	2.19	84.4%	85.8*	6	0.683	296
4.85	3735	621	627	99.0%	70.5%	71.8%	3731	2.42	77.3%	83.4*	-8	0.652	289
4.73	3987	669	683	98.0%	100.2%	101.3%	3979	1.78	110.0%	80.4*	4	0.671	311
4.61	4062	671	683	98.2%	148.4%	150.3%	4057	1.22	162.6%	67.3*	0	0.654	311
total	54166	9136	9198	99.3%	9.8%	10.8%	54083	9.06	10.8%	99.8*	59*	1.376	4197

EFF-1 phasing

Initial set of experimental phases obtained
from AutoSharp using MIR followed by
automatic solvent flattening





EFF-1 phasing

Buccaneer and Arp/Warp failed !!

initial model building (by hand using a skeletonized map in Coot), but sequence motifs required to help unambiguously assigning the aa sequence !!

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

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
13.82	13081	348	348	100.0%	3.4%	3.9%	13081	113.52	3.4%	99.9*	92*	4.061	144
9.77	25511	650	650	100.0%	3.4%	4.0%	25511	110.35	3.4%	100.0*	89*	3.026	293
7.98	29966	803	803	100.0%	4.3%	4.4%	29966	90.79	4.3%	100.0*	81*	2.758	370
6.91	39386	970	970	100.0%	5.6%	5.4%	39386	76.43	5.7%	100.0*	66*	2.240	455
6.18	46222	1084	1084	100.0%	7.0%	6.8%	46222	64.34	7.1%	99.9*	56*	1.813	512
5.64	47967	1217	1217	100.0%	8.2%	7.7%	47967	54.42	8.3%	99.9*	36*	1.347	577
5.22	54488	1307	1307	100.0%	9.0%	8.7%	54488	50.33	9.1%	99.9*	25*	1.159	621
4.61	61928	1484	1484	100.0%	11.1%	10.8%	61928	40.23	11.2%	99.9*	6	0.918	711
4.37	63386	1599	1599	100.0%	14.8%	14.3%	63386	31.63	15.0%	99.8*	1	0.876	769
4.17	68594	1650	1650	100.0%	20.0%	19.8%	68594	24.55	20.2%	99.8*	4	0.867	795
3.99	74451	1750	1750	100.0%	28.5%	29.0%	74451	17.59	28.8%	99.6*	-4	0.778	843
3.83	76970	1816	1816	100.0%	44.7%	45.7%	76970	11.67	45.2%	98.9*	-3	0.771	879
3.69	76200	1933	1937	99.8%	65.3%	65.9%	76200	7.99	66.1%	97.8*	-1	0.752	934
3.57	71753	1813	1950	93.0%	108.4%	100.0%	71753	5.31	109.8%	94.3*	5	0.814	869
3.46	83844	2010	2010	100.0%	117.9%	123.8%	83844	4.23	119.4%	92.9*	2	0.758	974
3.35	89067	2111	2111	100.0%	166.0%	175.5%	89067	2.98	168.0%	88.1*	1	0.725	1025
3.26	82753	2166	2166	100.0%	224.5%	240.7%	82753	2.00	227.5%	76.0*	-4	0.662	1052
3.17	64447	2194	2194	100.0%	319.4%	342.7%	64447	1.24	325.0%	48.8*	-2	0.668	1066
3.09	52128	2266	2266	100.0%	480.1%	525.5%	52128	0.68	490.9%	27.9*	-1	0.597	1102
total	1182859	30600	30741	99.5%	13.2%	13.6%	1182859	25.48	13.4%	100.0*	12*	1.017	14676

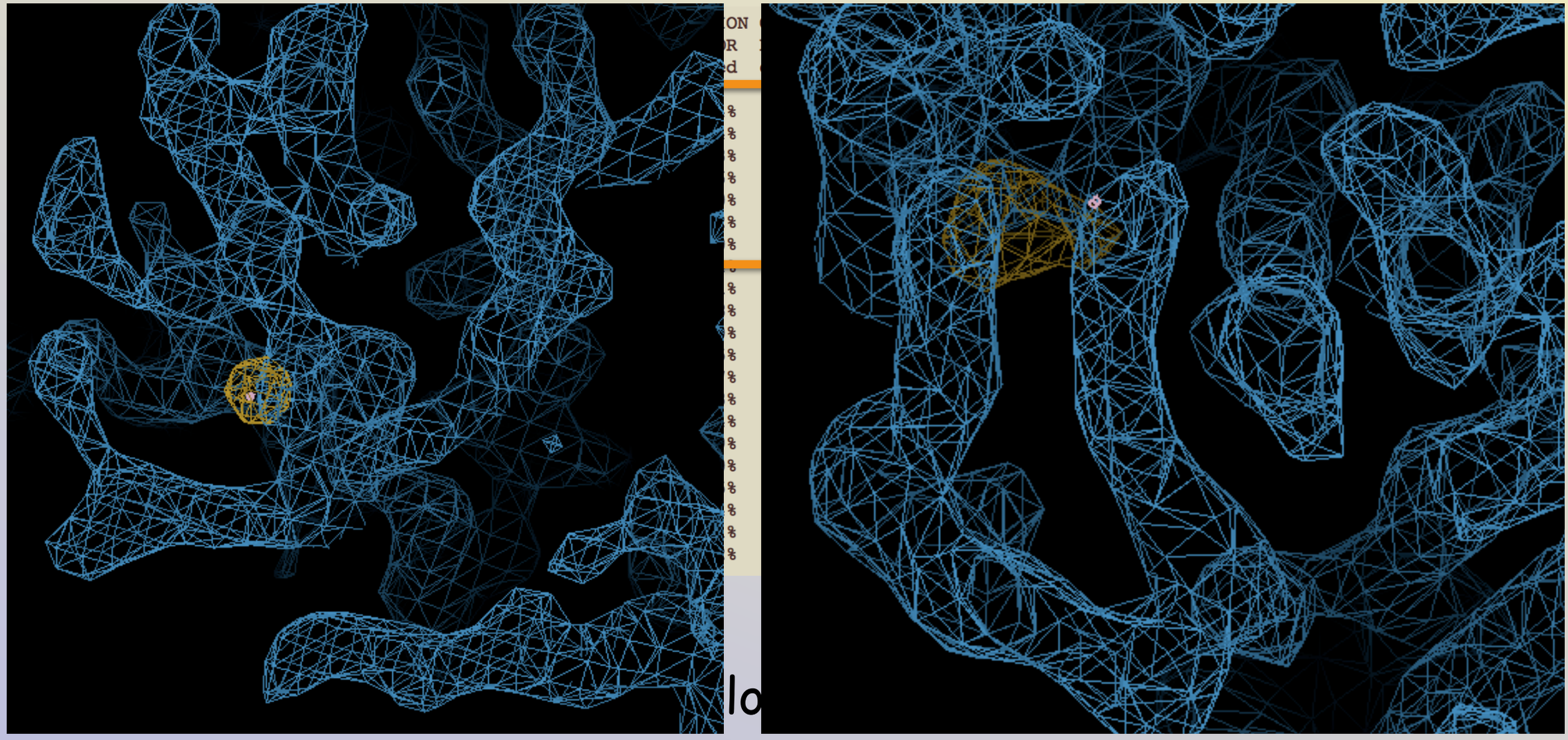
XDS/XSCALE

Sulfur-SAD, but no low-dose data collection !

EFF-1 phasing

Buccaneer and Arp/Warp failed !!

initial model building (by hand using a skeletonized map in Coot), but sequence motifs required to help unambiguously assigning the aa sequence !!



lo

EFF-1 - structure determination

- Using these anchoring points (Disulfide bridges, NGS) an initial model (~ 55% of the protein backbone) was built

- Based on this model a truncated protein version was designed that crystallized in a different space group with three molecules per AU

- Multi-crystal averaging then allowed structure determination

EFF-1 - conclusions

- Identification of two derivatives was not sufficient to build a full model

- The additional anchoring points provided by the anomalous sulfur map were essential to build the initial model

- Every bit of additional phase information can make the difference … even if it is not able to solve the structure on its own !

Summary

Advantages	Disadvantages
Less time-consuming than inverse-beam collection strategies	Requires more computing power (storage and calculation)
Higher true completeness / less radiation damage	?
Simple and easy to use	?
Can be adapted to both native and derivative crystals	?
Possibility to align along a symmetry axis	?

It does not cost much … but it can help a lot !!!