Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/2/9/e1600292/DC1

Supplementary Materials for

Native phasing of x-ray free-electron laser data for a G protein–coupled receptor

Alexander Batyuk, Lorenzo Galli, Andrii Ishchenko, Gye Won Han, Cornelius Gati, Petr A. Popov, Ming-Yue Lee, Benjamin Stauch, Thomas A. White, Anton Barty, Andrew Aquila, Mark S. Hunter, Mengning Liang, Sébastien Boutet, Mengchen Pu, Zhi-jie Liu, Garrett Nelson, Daniel James, Chufeng Li, Yun Zhao, John C. H. Spence, Wei Liu, Petra Fromme, Vsevolod Katritch, Uwe Weierstall, Raymond C. Stevens, Vadim Cherezov

Published 23 September 2016, *Sci. Adv.* **2**, e1600292 (2016)
DOI: 10.1126/sciadv.1600292

This PDF file includes:

- table S3. Comparison of interactions involving charged residues between PDB: 4EIY and A_{2A}_SSAD_1.9 structures.
- table S4. Comparison of protein and data collection parameters for successful S-SAD phasing of XFEL data.

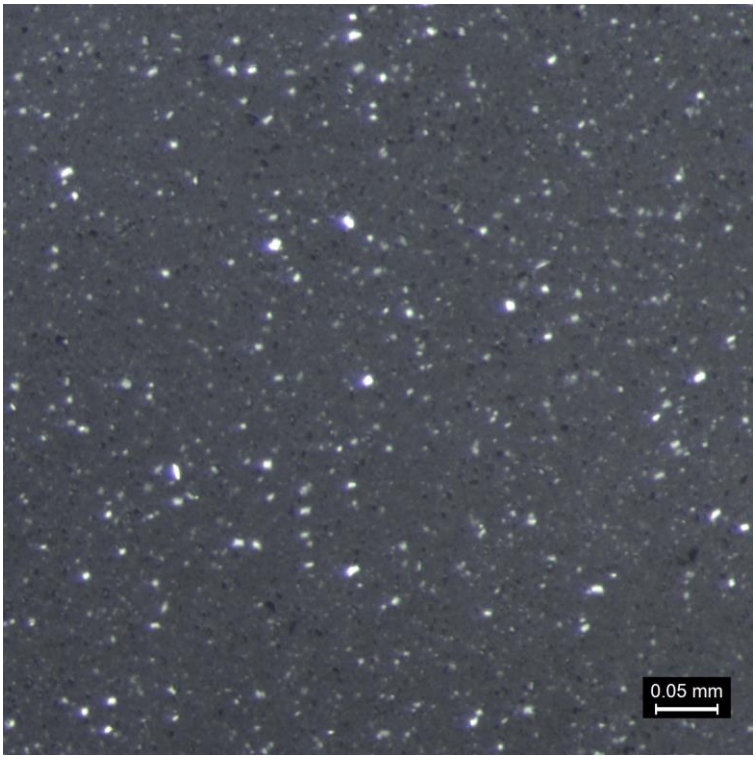

fig. S1. A_{2A}-BRIL/ZM241385 microcrystals used for data collection. The image was taken using a high-resolution zoom stereo microscope with cross-polarizers.

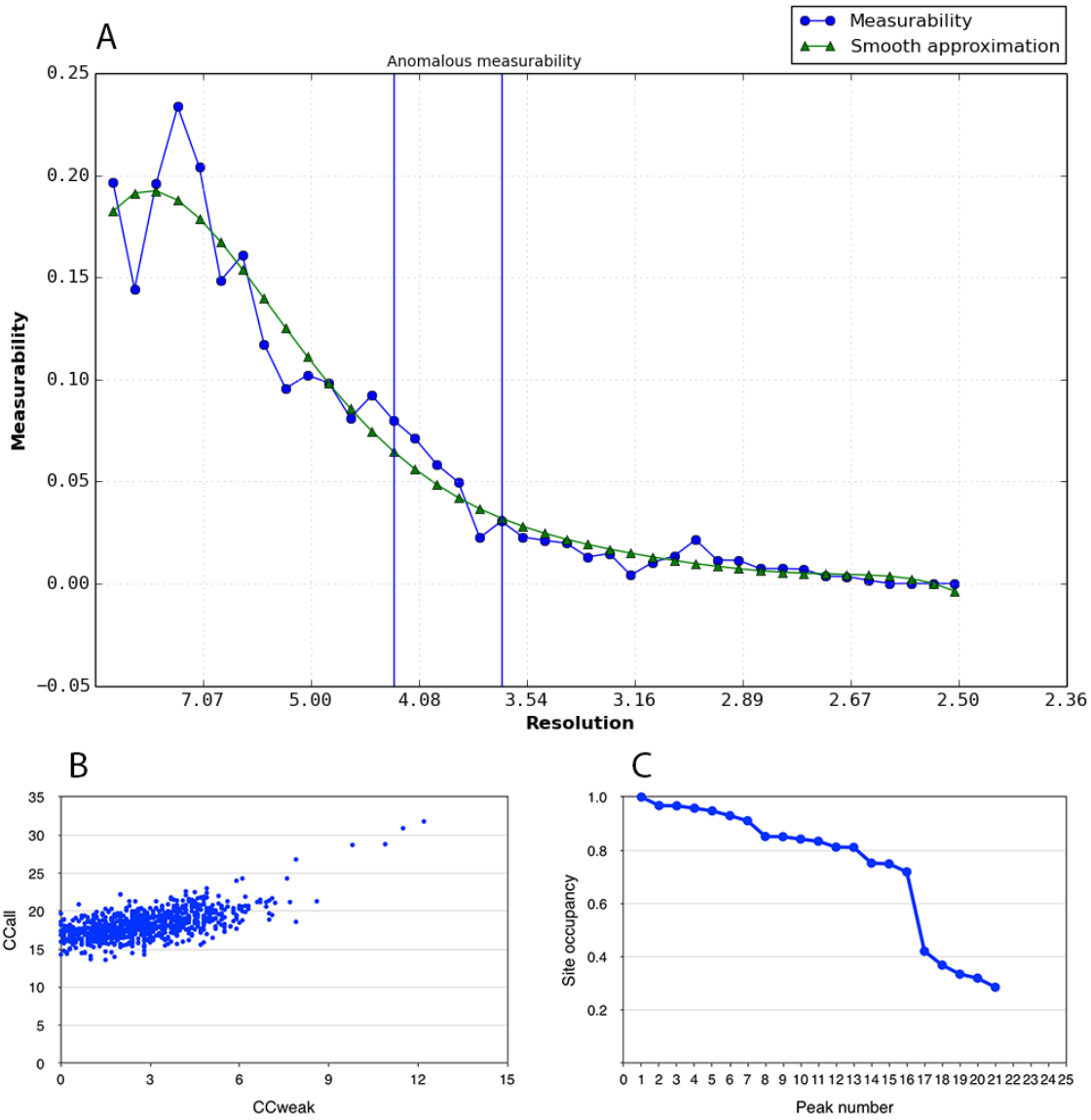

fig. S2. Strength of anomalous signal and sulfur atom search. (**A**) Dependence of the anomalous signal measurability on resolution calculated with phenix.xtriage (31). The anomalous signal measurability, is defined as the fraction of Bijvoet related intensity differences, for which $|I^+-I^-|/\sigma(I^+-I^-) > 3.0$, $\min[I^+/\sigma(I^+), I^-/\sigma(I^-)] > 3.0$ holds. (**B**) SHELXD correlation coefficients, CC, between observed and calculated Bijvoet differences. (**C**) Sulfur site occupancy vs. peak number plot.

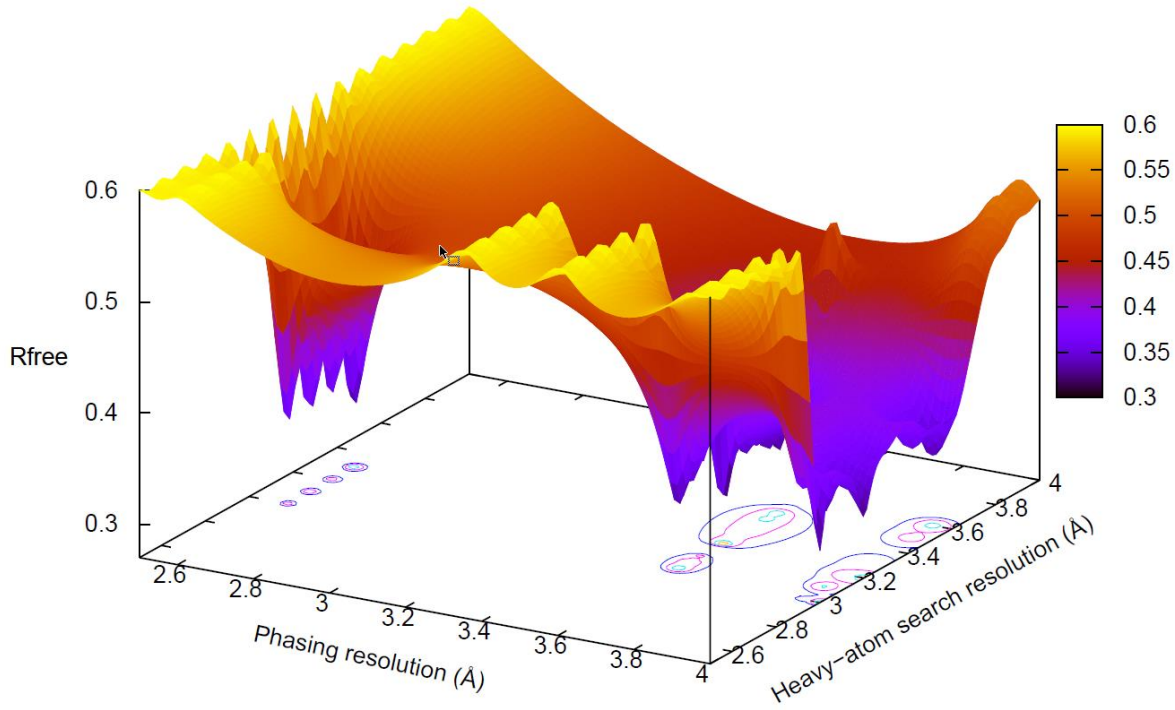

fig. S3. **Parameter-space screening results for S-SAD phasing using the X²DF pipeline.** A three-dimensional screening was performed varying the high-resolution limits for the sulfur site search and phasing, and the number of sulfur atoms to search. This figure represents an intersection at the number of sulfur atoms to search of 10. However, SHELXD yielded 18 sites, and 6 low occupancy sites were eliminated automatically before phasing. The results show that there are multiple combinations of high-resolution limits for sulfur site search and phasing which led to structure solutions. Colors represent the respective R_{free} value as indicated by the color scalebar.

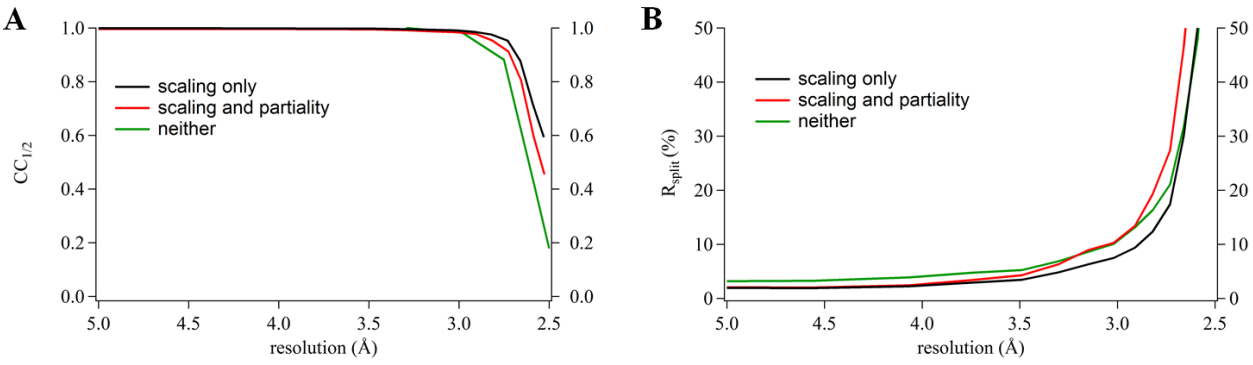

fig. S4. **Effect of different data processing methods on data merging metrics.** (**A**) Correlation coefficient $CC_{1/2}$ and (**B**) R_{split} factor.

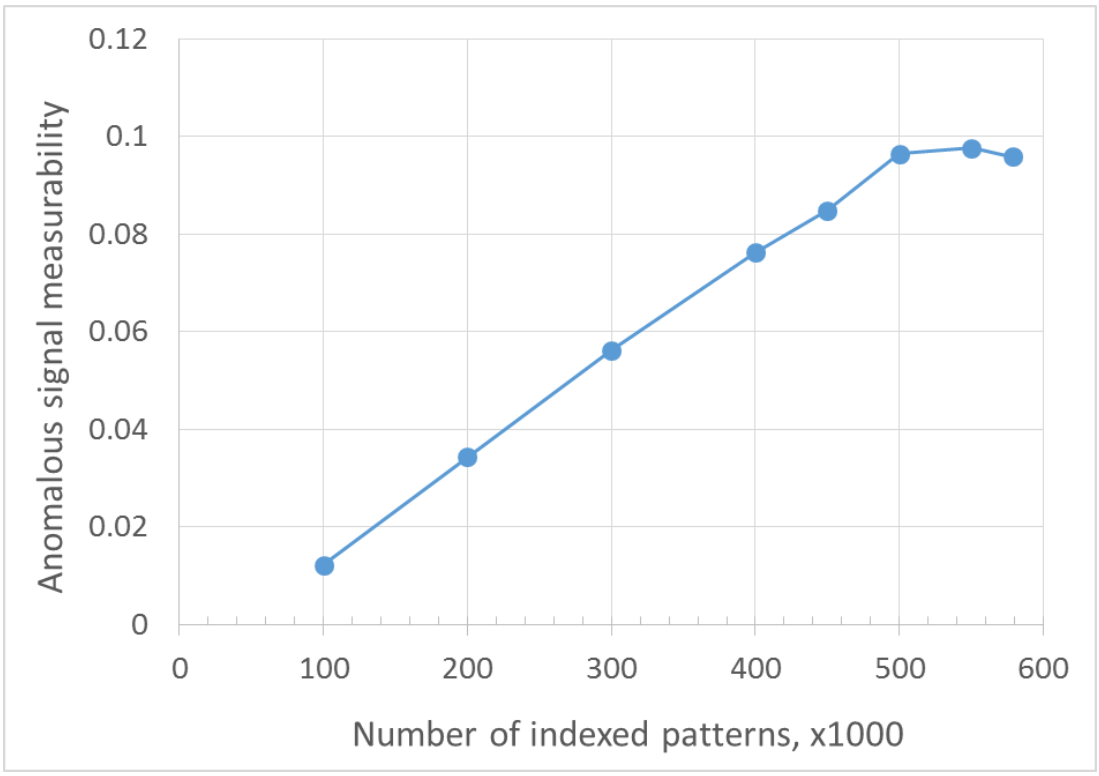

fig. S5. Dependence of anomalous signal measurability on the number of indexed patterns. The anomalous signal measurability is calculated by phenix.xtriage (31) in the resolution range 28 – 3.7 Å. The anomalous signal measurability is defined as the fraction of Bijvoet related intensity differences, for which $|I^+-I^-|/\sigma(I^+-I^-) > 3.0$, $\min[I^+/\sigma(I^+), I^-/\sigma(I^-)] > 3.0$ holds.

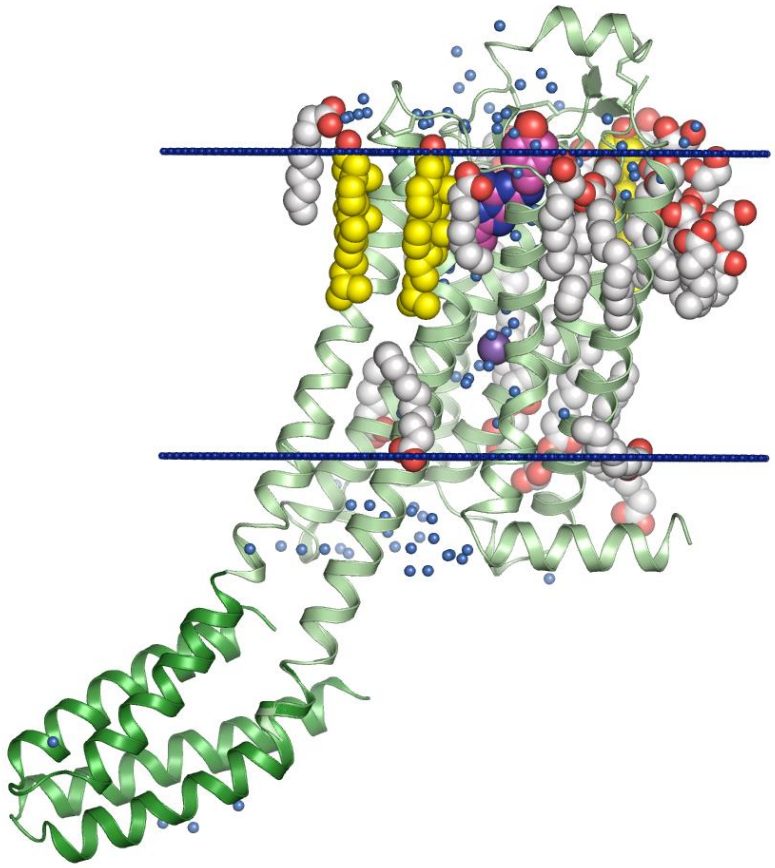

fig. S6. Final 1.9 Å XFEL room temperature A_{2A}AR-BRIL structure (A_{2A}_S-SAD_1.9). A_{2A}AR is shown as a pale green cartoon, BRIL fusion as a green cartoon. Ligand ZM241385 is shown as spheres with magenta carbons, cholesterols as spheres with yellow carbons, and other lipids as spheres with gray carbons. The sodium ion is shown as a purple sphere, and water molecules as small blue spheres. The membrane boundaries are indicated by blue lines with the extracellular side on the top and intracellular side on the bottom.

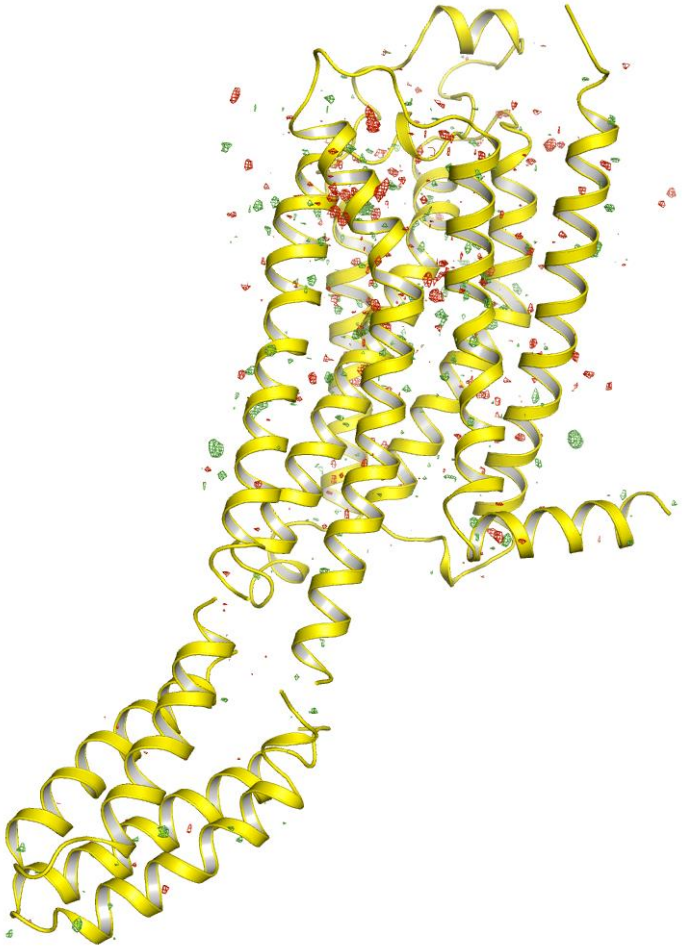

fig. S7. Structure-factor amplitude difference Fourier map between A_{2A}_S-SAD_2.5 and A_{2A}_MR_2.5 structures. Electron density is contoured at +3σ (green) and -3σ (red).

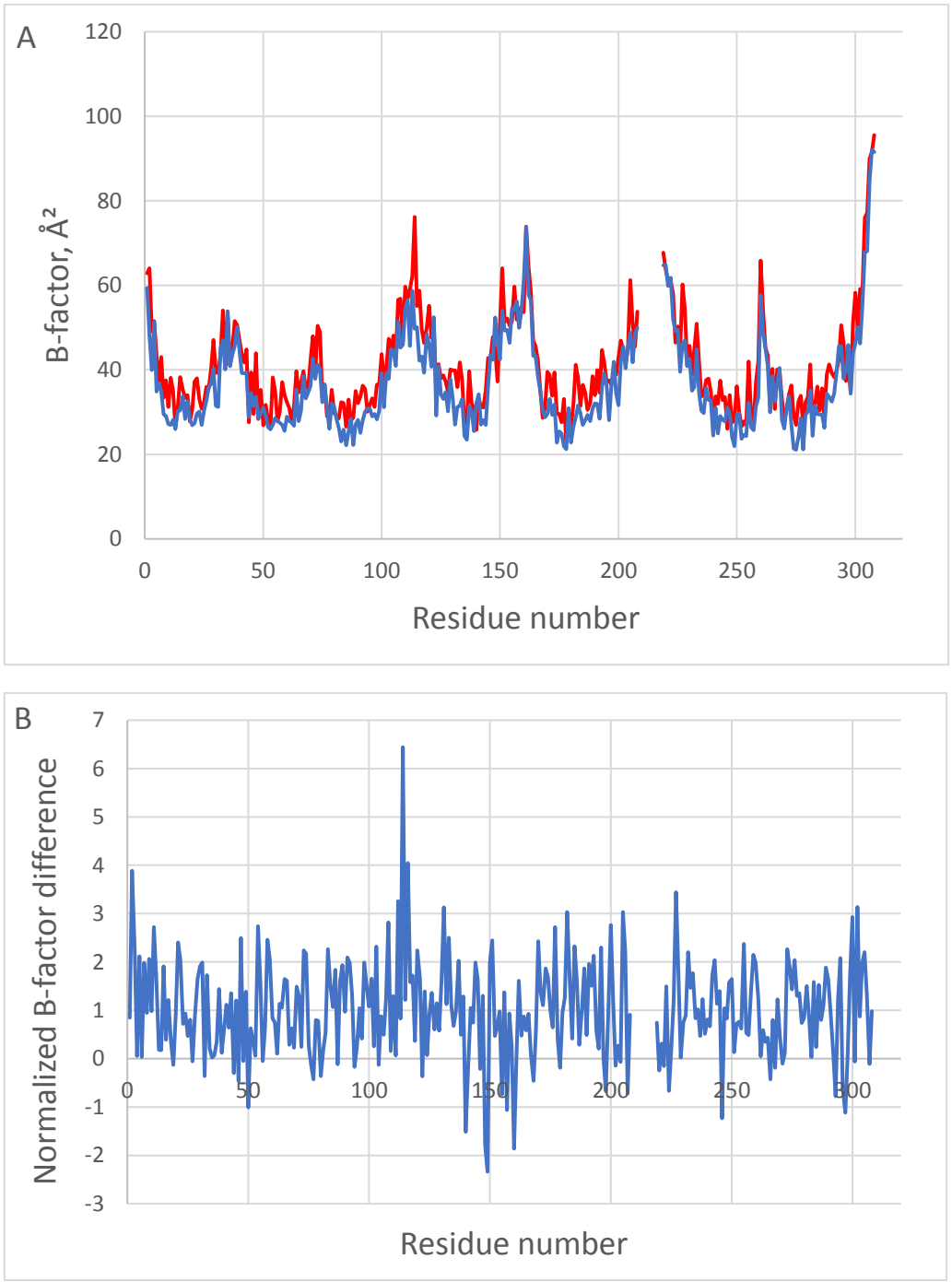

fig. S8. B factor comparison between A_{2A}_S-SAD_2.5 and A_{2A}_MR_2.5 structures. (**A**) Distribution of Cα atom B-factors for A_{2A}_S-SAD_2.5 (red) and A_{2A}_MR_2.5 (blue) vs residue number. (**B**) Normalized Cα atom B-factor difference between A_{2A}_S-SAD_2.5 and A_{2A}_MR_2.5 structures.

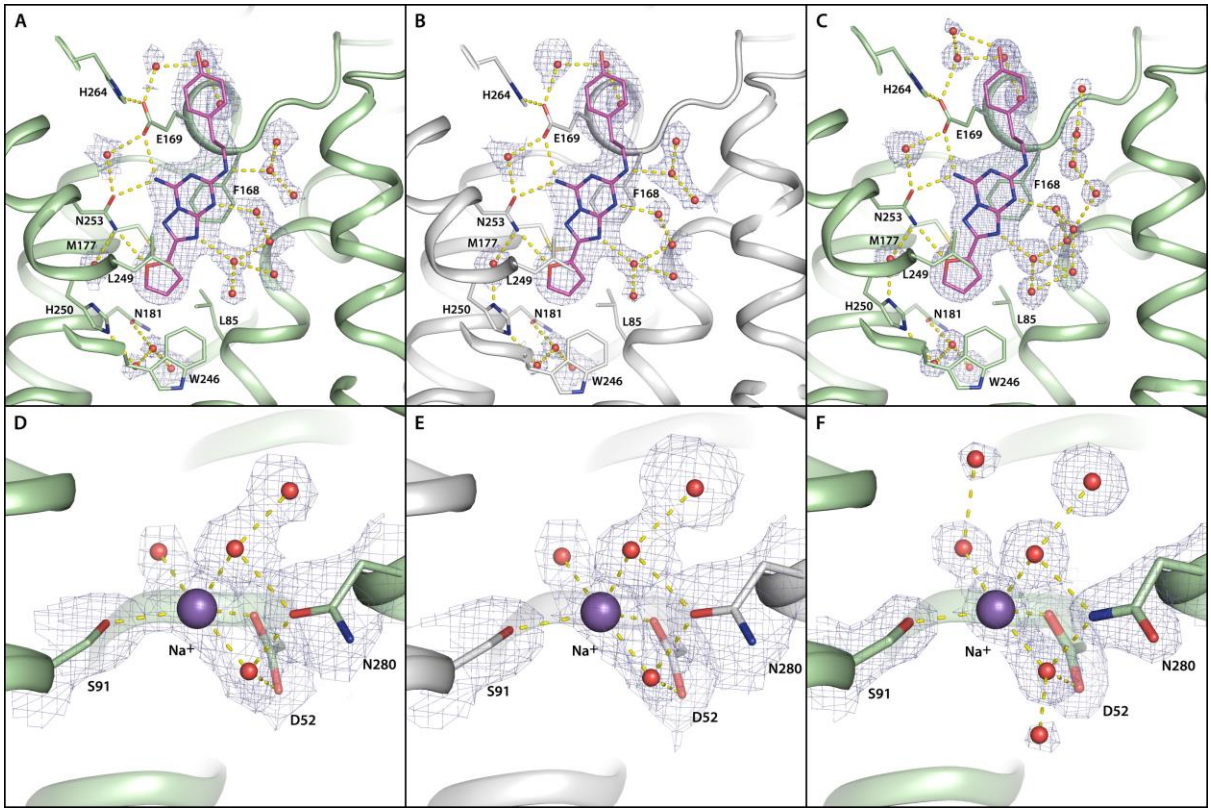

fig. S9. Comparison of 2mFo-DFc electron density maps for the ligand- and sodium-binding pockets obtained by S-SAD and MR phasing. (**A**), (**B**) and (**C**) Ligand binding pocket for A_{2A}_S-SAD_2.5 (A), A_{2A}_MR_2.5 (B) and A_{2A}_S-SAD_1.9 (C) structures. (**D**), (**E**) and (**F**) Sodium ion binding site for A_{2A}_S-SAD_2.5 (D), A_{2A}_MR_2.5 (E) and A_{2A}_S-SAD_1.9 (F) structures. Electron density is contoured at 0.7 σ around water molecules (red spheres) and at 1 σ elsewhere.

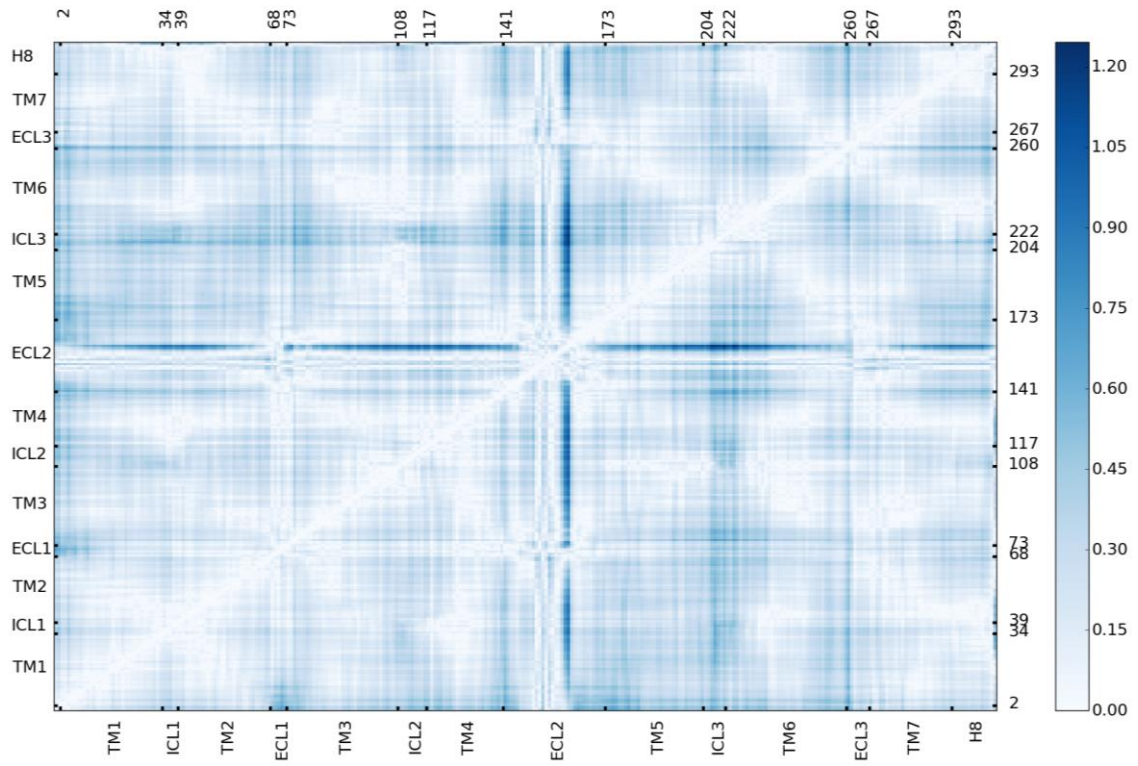

fig S10. Cα-Cα difference distance matrix between A_{2A}_S-SAD_1.9 and previously determined A_{2A}AR structure (PDB: 4EIY). Residue numbers are shown on the top and right side of the matrix. Distance color scale (in Å) is given on the right side of the matrix.

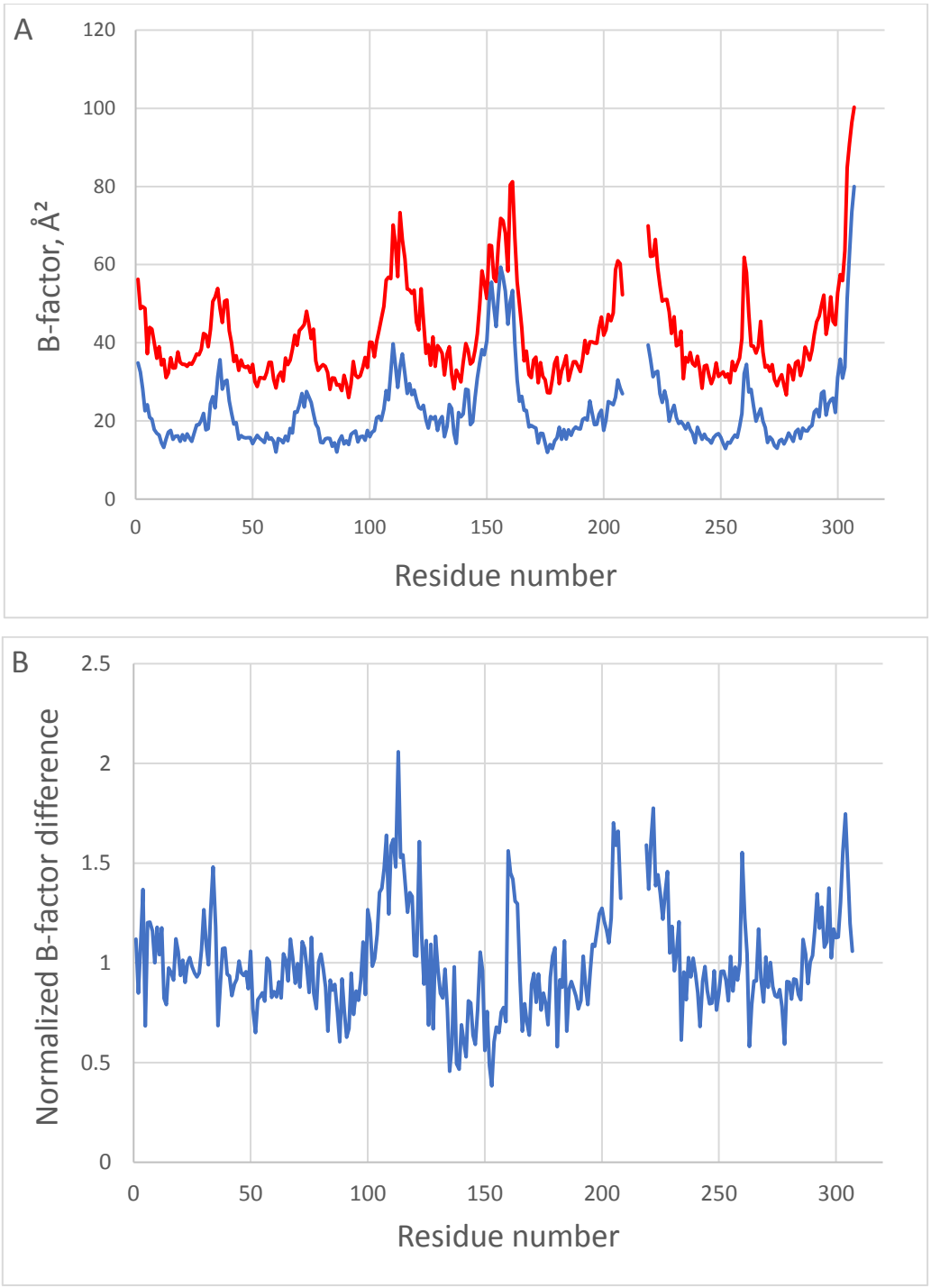

fig. S11. B factor comparison between room temperature A_{2A}_S-SAD_1.9 and previously determined cryocooled A_{2A}AR structure (PDB: 4EIY). (**A**) Distribution of Cα atom B-factors for A_{2A}_S-SAD_1.9 (red) and 4EIY (blue) vs. residue number. (**B**) Normalized Cα atom B-factor difference between A_{2A}_S-SAD_1.9 and 4EIY structures.

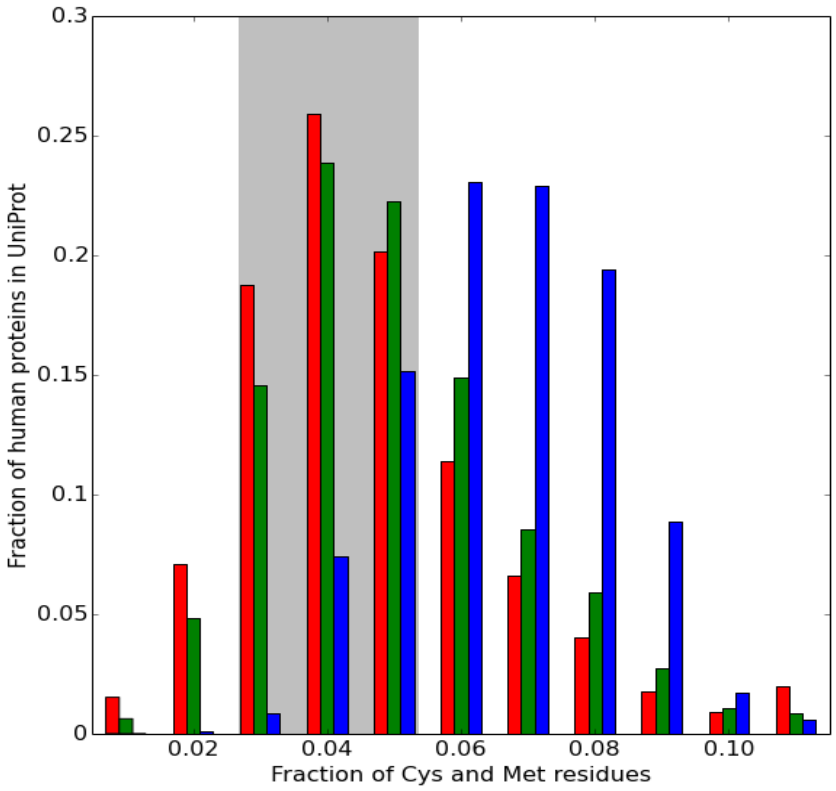

fig. S12. Distribution of Cys and Met residues in human proteins. All human proteins (red), transmembrane proteins (green), and GPCRs (blue) correspond to subsets of 20194, 5181 and 825 protein entries in UniProt (Swiss-Prot, retrieved Jan 1st 2016), respectively. The shaded area corresponds to values between the S-SAD phasing cutoff described in the main text (12 sulfur atoms) and the value obtained for the $A_{2A}AR$ (24 sulfur atoms). Proteins with a higher fraction of Cys and Met residues than the S-SAD phasing cutoff (2.7%) comprise more than 88.5% of all, 92.4% of transmembrane, and 99.9% of all GPCR proteins.

table S1. Data collection statistics. Highest-resolution shell is shown in parentheses.

Dataset	**A_{2A}_S-SAD**	**A_{2A}_High-Res**
Wavelength (Å)	2.07	1.27
Pulse duration, fs	45	35
Flux, photons/pulse	$1.7 \cdot 10^{11}$	$6.4 \cdot 10^{10}$
Max dose per crystal, MGy	140	20
Space group	$C222_1$	$C222_1$
Unit cell parameters *a,b,c* (Å)	40.4 180.5 142.7	40.4 180.7 142.8
Number of collected images	7,324,430	948,961
Number of hits / indexed images	1,797,503 / 578,620	232,283 / 72,735
Number of total / unique reflections	38,036,147 / 34,140[#]	12,215,692 / 41,882
Resolution (Å)	24 – 2.50 (2.58 – 2.50)	24 – 1.90 (2.00 – 1.90)
Mean I/σ(I)	20.3 (1.6)[#]	6.0 (0.6)
Completeness (%)	100 (83.5)[#]	100 (100)
Multiplicity	1022 (7)[#]	291 (62)
R_{split} (%)	4.1 (75.1)[#]	10.1 (197)
CC*	0.999 (0.69)[#]	0.998 (0.58)

[#]considering Friedel pairs as individual reflections

table S2. Data refinement statistics. Highest-resolution shell is shown in parentheses.

Structure	**A_{2A} S-SAD_1.9**	**A_{2A} MR_1.9**	**A_{2A} S-SAD_2.5**	**A_{2A} MR_2.5**
Resolution (Å)	24 – 1.90	24 – 1.90	24 – 2.50	24 – 2.50
No reflections / test set	41,828 / 1,988	41,828 / 1,988	18,567 / 899	18,622 / 900
R_{work} / R_{free}	0.173 / 0.208	0.174 / 0.207	0.174 / 0.228	0.173 / 0.219
Number of atoms				
All	3,632	3,636	3,508	3,484
A_{2A}AR	2,394	2,408	2,317	2,350
BRIL	717	713	718	713
ZM241385	25	25	25	25
Na^+ ion	1	1	1	1
Waters	105	93	70	63
Lipids and other	389	396	377	332
Wilson B ($Å^2$)	41.5	41.5	46.9	41.0
Mean overall B ($Å^2$)				
All	58.9	58.5	55.3	50.5
A_{2A}AR	45.4	45.3	43.7	39.5
BRIL	93.8	92.3	87.9	81.9
ZM241385	36.0	35.6	36.5	30.3
Na^+ ion	35.3	35.5	42.8	34.2
Waters	56.8	54.1	48.1	41.2
Lipids and other	79.7	80.1	67.1	64.9
R.m.s. deviations				
Bond lengths (Å)	0.010	0.010	0.002	0.002
Bond angles (°)	1.2	1.2	0.91	0.90
Clash Score[#]	3.9	3.2	1.9	2.9
Cβ outliers (%)[#]	0	0	0	0
Rotamer outliers (%)[#]	1.6	2.2	2.0	2.3
Ramanchandran plot statistics (%)[#]				
Favored regions	99.0	99.0	99.0	98.0
Allowed regions	1.0	1.0	1.0	2.0
Disallowed regions	0	0	0	0

[#]As defined in MolProbity (*44*)

table S3. Comparison of interactions involving charged residues between PDB: 4EIY and A_{2A}_S-SAD_1.9 structures. Interactions which were observed in the A_{2A}_S-SAD_1.9 to be more/equally/less favorable are highlighted in green/gray/red, respectively. Interactions involving disordered residues are highlighted in dark gray.

Residue 1	Residue 2	Distance, Å 4EIY	Distance, Å A_{2A}_S-SAD_1.9	ΔB Residue 1, Å^2 (atom)	ΔB Residue 2, Å^2 (atom)
K150	D170	3.20	2.63	24.8 (NZ)	14.7 (OD1)
E151	H155	3.73	2.84	13.6 (O)	15.7 (ND1)
K153	D170	3.24	2.76	-4.5 (NZ)	14.4 (OD2)
E294	N34	2.78	2.66	22.7(OE2)	18.5 (ND2)
D101	R102	3.01	2.92	22.0 (OD2)	16.9 (NH2)
D101	R102	2.83	2.88	22.0 (OD2)	20.6 (NE)
R120	L115	2.82	2.80	21.6 (NH2)	31.5 (O)
E169	H264	2.77	2.78	10.5 (OE1)	10.3 (NE2)
E13	H278	2.50	2.54	8.9 (OE2)	12.6 (ND1)
E13	H278	2.87	3.08	10.2 (OE1)	12.6 (ND1)
R291	S234	2.54	2.82	27.7 (NE)	17.2 (OG)
E151	N154	4.48	2.53	32.3 (OE1)	19.1 (ND2:A)
R300	Q297	8.26	5.45	4.1 (NH1:B)	30.7 (OE1)

table S4. Comparison of protein and data collection parameters for successful S-SAD phasing of XFEL data.

	Lysozyme (22)	Thaumatin (23)	A_{2A}AR-BRIL
Molecular weight (kDa)	14.4	22.4	49.9
Ordered Cys and Met residues	10[#]	17	21
Total residues	129	207	447
% Cys and Met	7.8[#]	8.2	4.7
Crystal size (μm)	7-10	3×3×5	2×5×5
Crystal solvent content (%)	40.4	57.6	52.9
Space group	P 43 21 2	P 41 21 2	C 2 2 21
X-ray energy (keV)	7	6	6
Minimum number of indexed patterns sufficient for indexing	150,000	125,000	500,000
Protein used, mg	5	35	2.7

[#] Anomalous signal from one chloride atom was used in addition to sulfurs.