Supplementary Material to the paper:

PAN-modular Structure of Parasite *Sarcocystis muris* Microneme Protein SML-2 at 1.95 Å Resolution and the Complex with 1-Thio-β-D-Galactose

Jürgen J. Müller,[a] Manfred Weiss,[b] Udo Heinemann[a,c]

[a]Kristallographie, Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Str. 10, 13125 Berlin, Germany

[b]AG Makromolekulare Kristallographie, Helmholtz-Zentrum Berlin für Materialien und Energie, BESSY-II, Albert-Einstein-Str. 15, 12489 Berlin, Germany

[c]Institut für Chemie und Biochemie, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany

SAD substructure determination by sulfur phasing using the SHELX suite

In the following we will describe the structure solution with SHELX C, D, E-beta (Sheldrick, 2010) by using unconventional input parameters. It must be mentioned that the primary data evaluation and scaling had to be done very carefully, removing dubious images and reflections marked as "alien" by XDS (Kabsch, 1993; version May 10, 2010), cutting very low and very high resolution bins with high R_{sym}, and using the newest software versions.

Anomalous datasets were collected with relatively low redundancy (multiplicity < 17.8) during five runs at a wavelength of 2.0000 Å, simultaneously optimizing the anomalous signal from sulfur and crystal life time. Primary data evaluation was done with XDS/XSCALE version 2010 (Kabsch, 1993). Different numbers of datasets were merged to optimize the anomalous signal. The highest signal was obtained from a combination of sets 2-4. No substructure solution could be found when set 5 was included (Table S1), because this set suffers from beginning radiation damage (see Fig. S10). The term "native" is used as a synonym for the data set collected at 1 Å wavelength. It has a higher resolution than the anomalous data and was used for phasing and density modification in SHELXE-beta. Friedel mates were not merged because of the local scaling facility of native to anomalous data.

Resolution limit [Å]	SigAno SET 2	Resolution limit [Å]	SigAno SET 3	SigAno SET 4	Resolution limit [Å]	SigAno SET 5
9.84	3.75	10.73	4.07	3.96	12.97	2.61
6.96	3.35	7.59	3.47	3.10	9.17	2.21
5.68	3.22	6.20	3.48	3.21	7.49	2.02
4.92	2.39	5.37	2.74	2.53	6.48	1.82
4.4	1.83	4.80	2.18	1.96	5.80	1.62
4.02	1.71	4.38	1.68	1.57	5.29	1.44
3.72	1.57	4.06	1.62	1.51	4.9	1.29
3.48	1.47	3.79	1.57	1.45	4.59	1.16
3.28	1.33	3.58	1.39	1.33	4.32	1.06
3.11	1.29	3.39	1.32	1.25	4.1	1.05
2.97	1.56	3.24	1.23	1.10	3.91	0.97
2.84	1.09	3.10	1.15	1.08	3.74	0.98
2.73	1.02	2.98	1.04	0.98	3.60	0.92
2.63	0.97	2.87	0.98	0.92	3.47	0.85
2.54	0.89	2.77	0.91	0.88	3.35	0.89
2.46	0.87	2.68	0.87	0.84	3.24	0.78
2.39	0.85	2.60	0.82	0.80	3.15	0.76
2.32	0.79	2.53	0.81	0.75	3.06	0.73
2.26	0.78	2.46	0.72	0.77	2.98	0.69
2.20	0.75	2.40	0.75	0.70	2.90	0.76

Table S1
Anomalous signal SigAno from XDS (Kabsch, 1993; version May 10, 2010) shows the inferior data quality of set 5 compared to sets 2-4 due to radiation damage (see also Fig. S10).

The maximal resolution region for high data quality and a useful anomalous signal was chosen according to several criteria:

(1) An $\langle I/\sigma(I)\rangle$ value over 30 for native data should be realized for good SAD-phasing conditions. A maximal resolution of 3.0 Å was chosen by that criterion (Fig. S1).

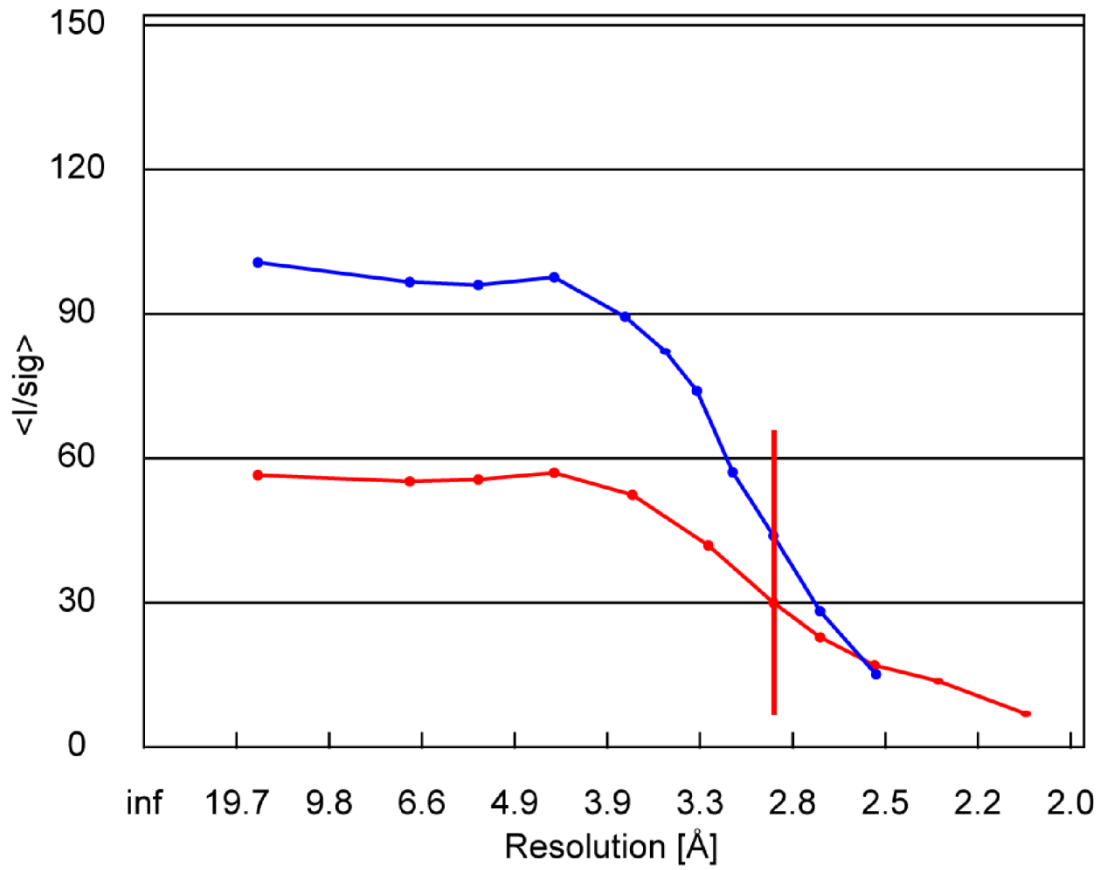

Fig. S1. SHELXC: <*I*/σ(*I*)> plot for native (red) and merged datasets 2+3+4 (blue). The red bar marks the high resolution limit.

(2) SHELXC estimates <*d''*/σ> as anomalous signal. A value larger than 1.2 is recommended for higher phasing power (0.8 corresponds to noise).

From that criterion a useful resolution of 3.1 Å was chosen for substructure determination by SHELXD (Fig. S2).

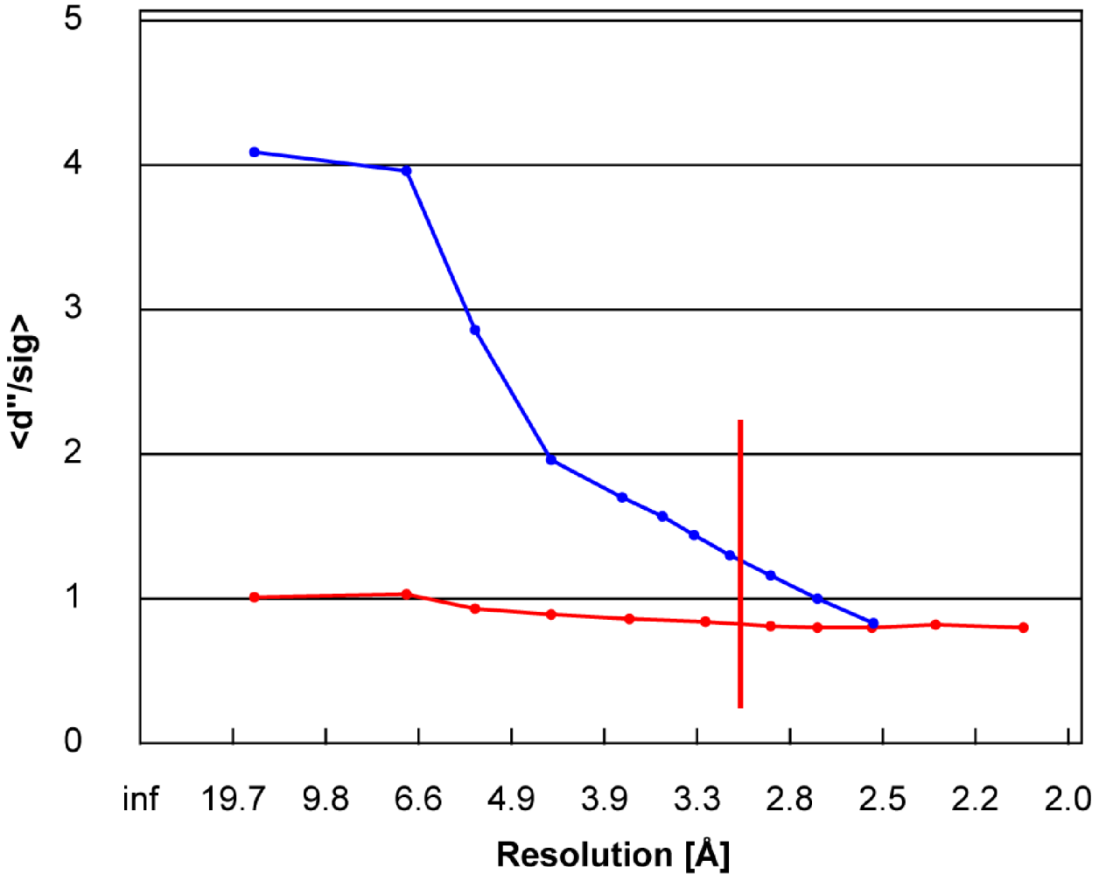

Fig. S2. SHELXC: <*d"*/σ> plot for native (red) and merged datasets 2+3+4 (blue). The red bar marks the high resolution limit.

(3) SHELXC estimates that the self-anomalous CC and the resolution cut-off for a high anomalous signal should be set at a value of 30%. The useful resolution deduced therefrom is 3.1 Å (Fig. S3).

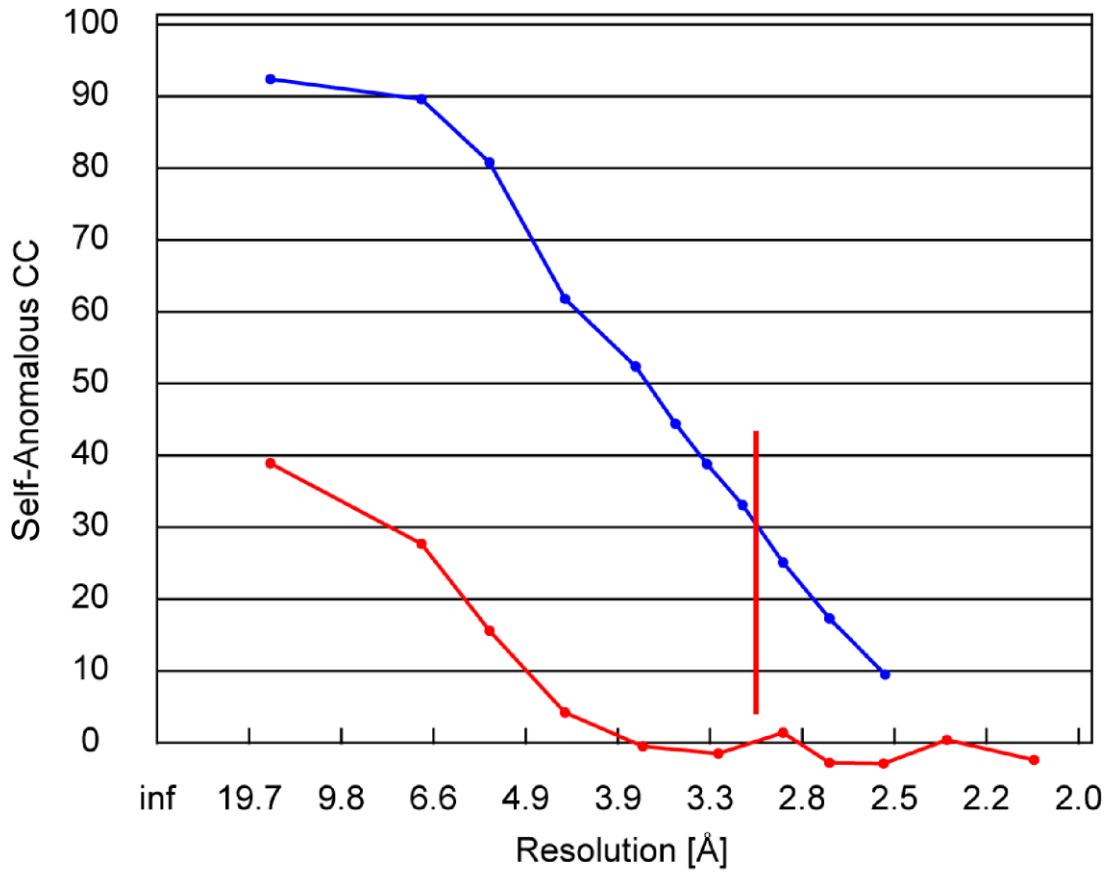

Fig. S3. SHELXC: Self-Anomalous CC plot for native (red) and merged datasets 2+3+4 (blue). The red bar marks the high-resolution limit.

SHELXD was run to determine the substructure with 66 sulfur atoms which includes 36 disulfide bridges treated as super-sulfurs (minimal distance allowed between heavy atoms 3.5 Å). Special positions are forbidden, in spite of the possible existence of ionic anomalous scatterers like SO_4^{2-}, and Cl^-. The limit number of tries to find the substructure was enhanced step by step up to 50,000 in different runs (recommended number some tens, in difficult cases some hundreds). The correct solution was found at try number 35,165 with rather low values of CCall = 35.5 and CCweak = 10.73, representing a singular point in "solution space" and not a small, clustered group of solutions (Figs. S4, S5, S6).

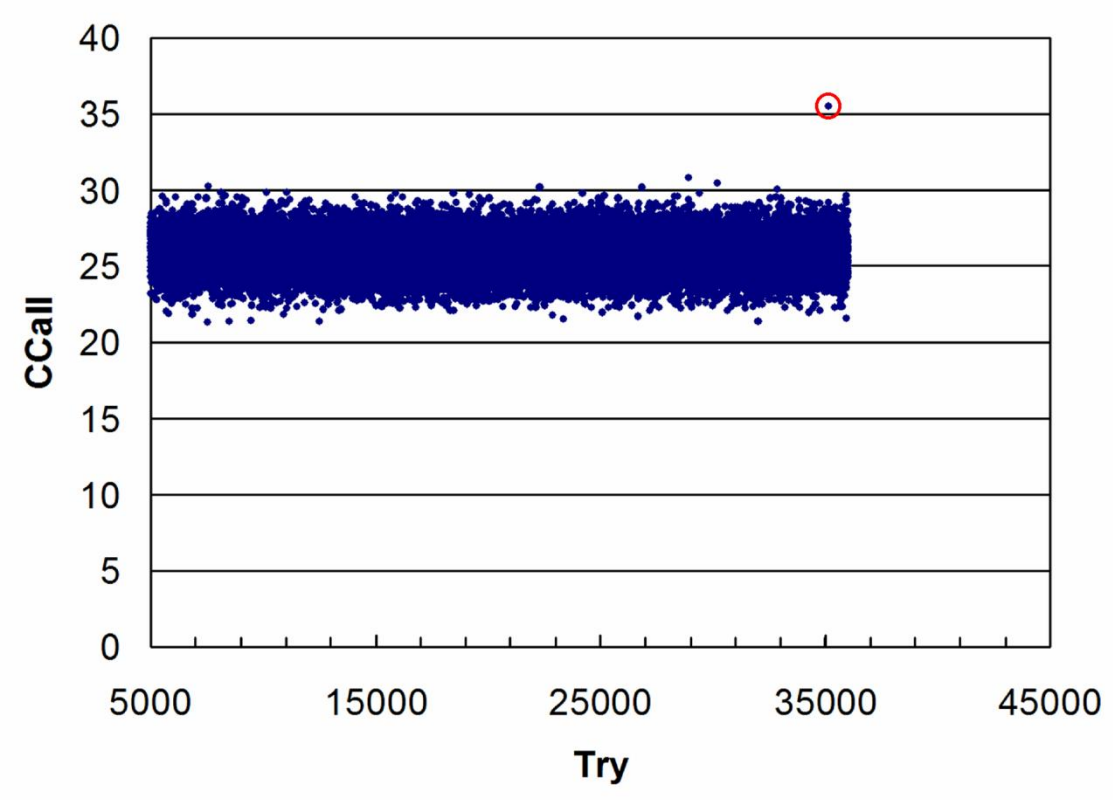

Fig. S4. SHELXD: CCall versus try. The red circle marks the solution.

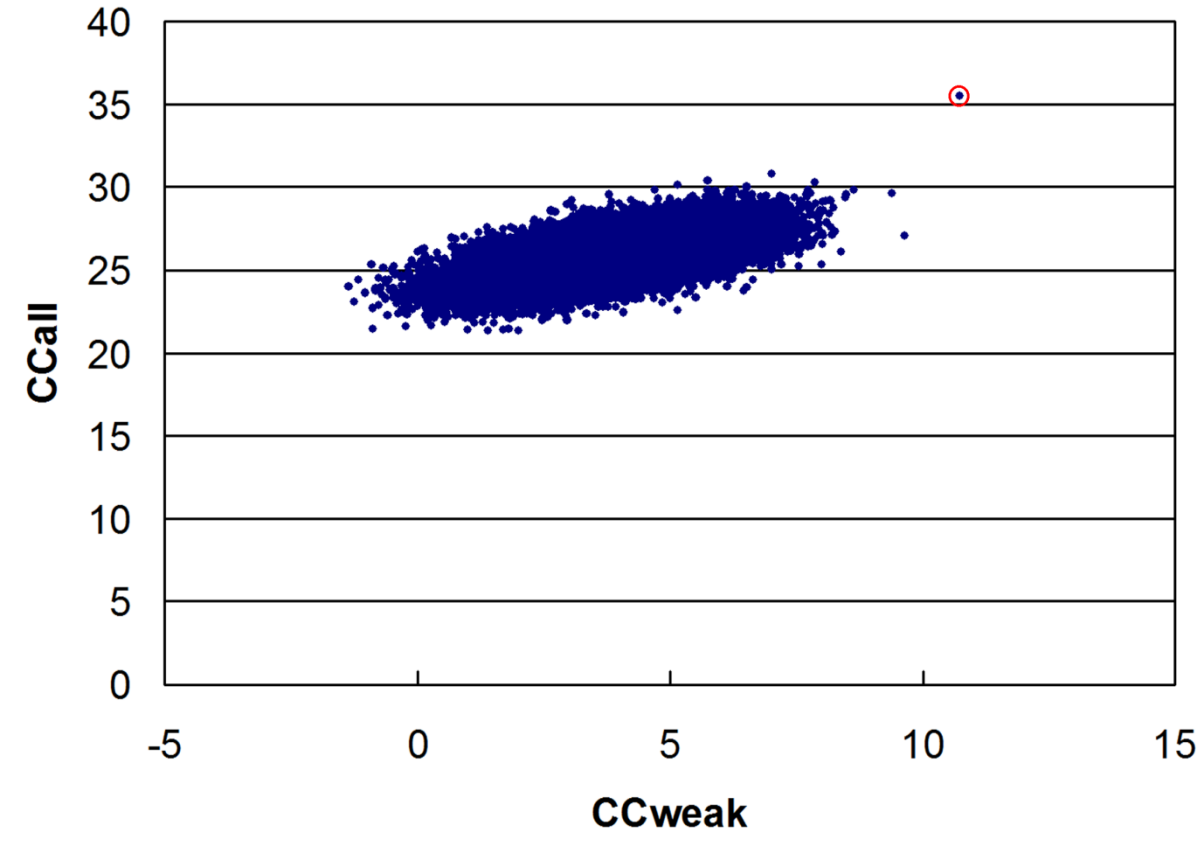

Fig. S5. SHELXD: CCall versus CCweak. The red circle marks the solution.

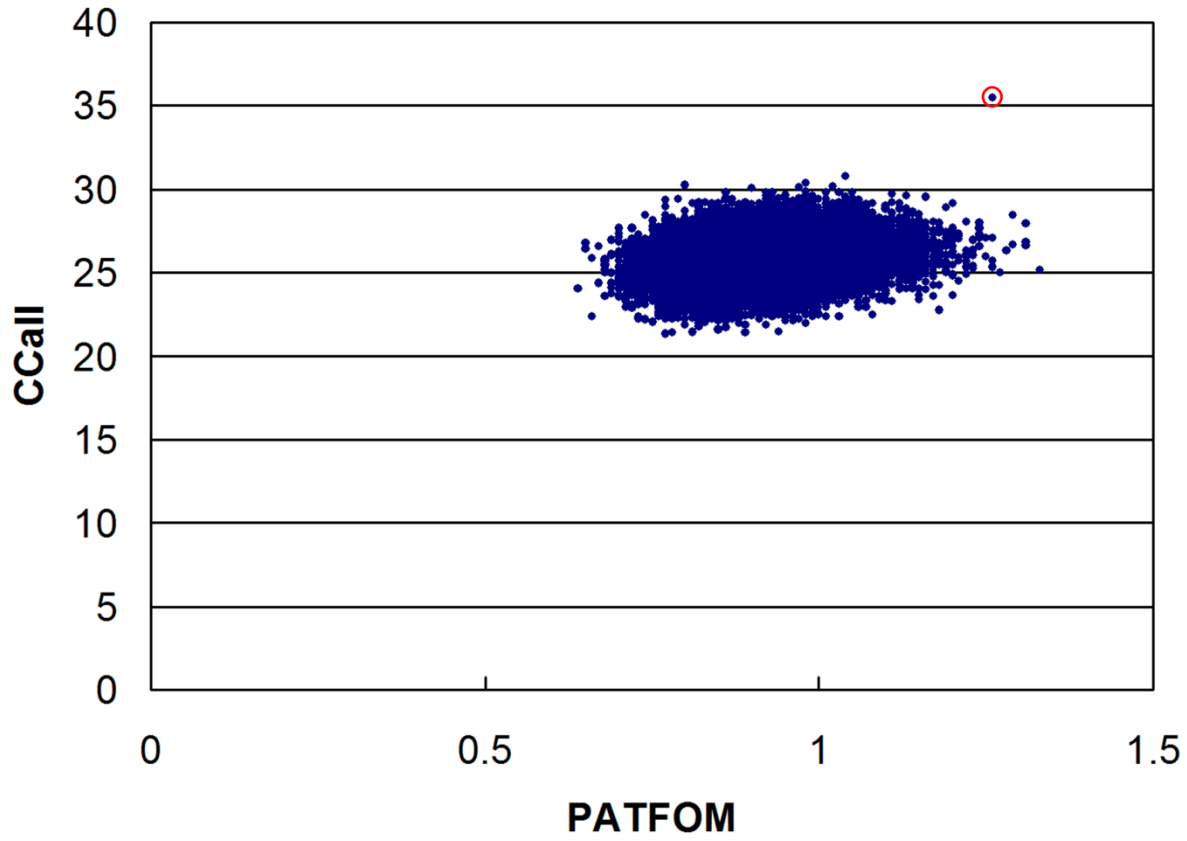

Fig. S6. SHELXD: CCall versus PATFOM. The red circle marks the solution.

In Fig. S7 the occupancy of the anomalous scatterers is shown as determined by SHELXD. 108 scatterers were found, exceeding the number of 102 sulfurs in the asymmetric unit. The output was used in SHELXE-beta, the version containing the auto-tracing module. The solvent content parameter (-s) was varied from 0.55 to 0.65, which was the best fictive disorder fraction (theoretical value from packing 0.59), and the –e1.0 extrapolation, in addition to the high resolution of 1.95 Å provided by the native data set, were added. Ten cycles of auto-tracing and final density modification ruled out the inverted structure, and three cycles of auto-tracing and density modification of the original structure determined the phases and resembled the threading in the asymmetric unit (Fig. S8). The traced poly-alanine chains consisted of 640 residues and allowed to recognize secondary structure elements. The experimental electron density permitted automatic side chain modeling (see Fig. 1 of this publication) using ArpWarp (Lamzin & Wilson, 1993).

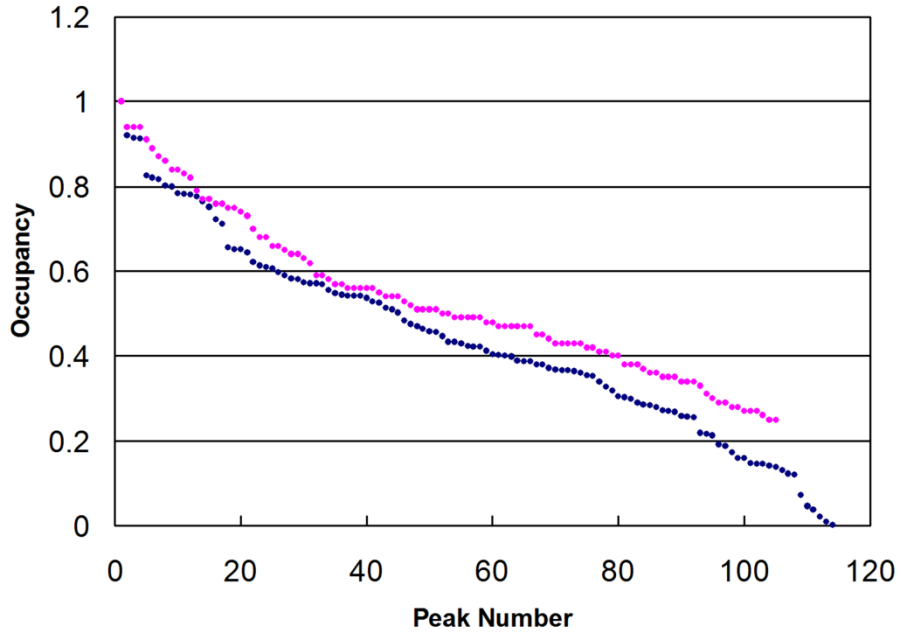

Fig. S7. SHELXD and SHELXE-beta: Occupancy of anomalous scatterers after SHELXD (colored blue) and 3 cycles of density modification with SHELXE-beta (magenta).

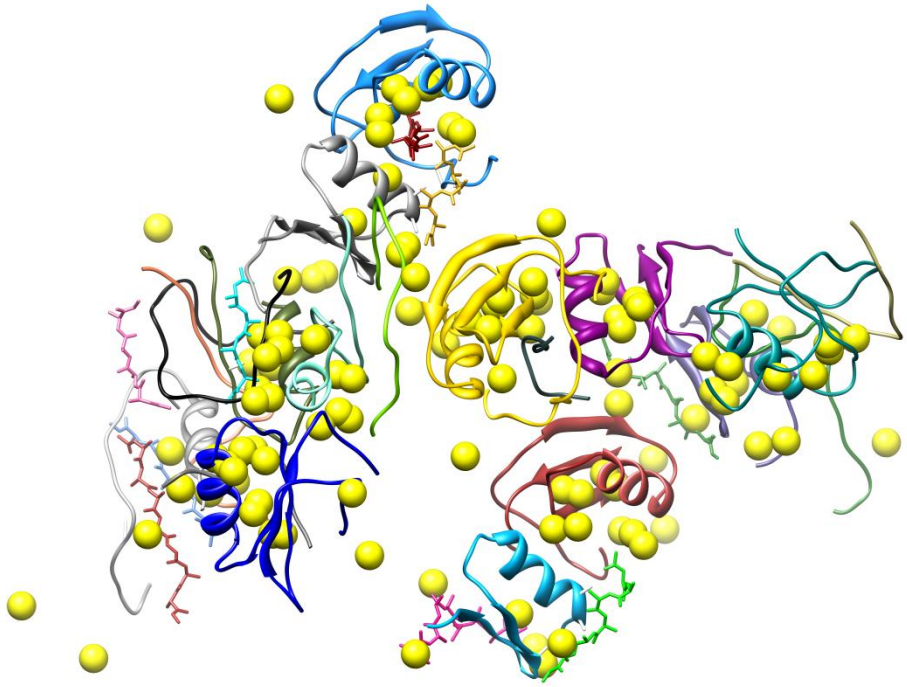

Fig. S8. Threading result of SHELXE-beta. The substructure of 105 anomalous scatterers output from SHELXE-beta (Fig. S7) after density modification are drawn as yellow spheres.

After using the phases and the preliminary model in ArpWarp (Lamzin & Wilson, 1993) for complete structure building and water search, several rounds of refinement in COOT (Emsley *et al.*, 2010) and REFMAC (CCP4, 1994) we ended up with the asymmetric unit shown in

Fig. S9. 59 out of the 105 anomalous scatterers could be identified (Dall'Antonia & Schneider, 2006) as sulfurs in the cysteine and methionine residues.

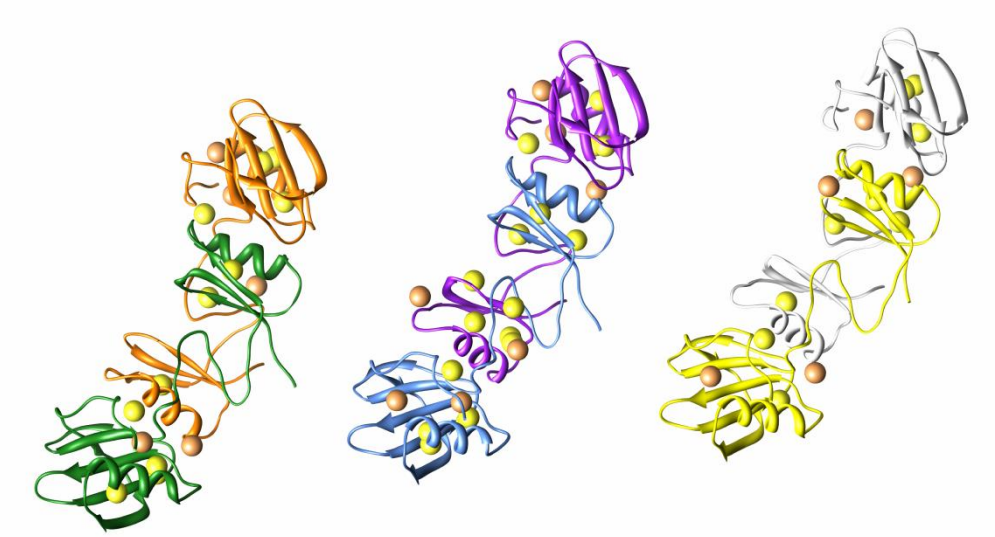

Fig. S9 Identification of 59 sulfurs of the substructure in the refined six chains in the asymmetric unit. Cysteine sulfurs are drawn as yellow spheres, methionine sulfurs in orange. The identification was done with program SITCOM (Dall'Antonia & Schneider, 2006).

After finishing the structure refinement, data set 5 was merged and used for a molecular replacement experiment using the dimers from the native structure. After several cycles of refinement with REFMAC 5 radiation damage became visible.

The solvent-exposed disulfide bridges C9-C78, C82-C86, C111-C117 show severe radiation damage (Fig. S10) whereas the three buried disulfides C34-C56, C38-C44, C107-C127 remained intact.

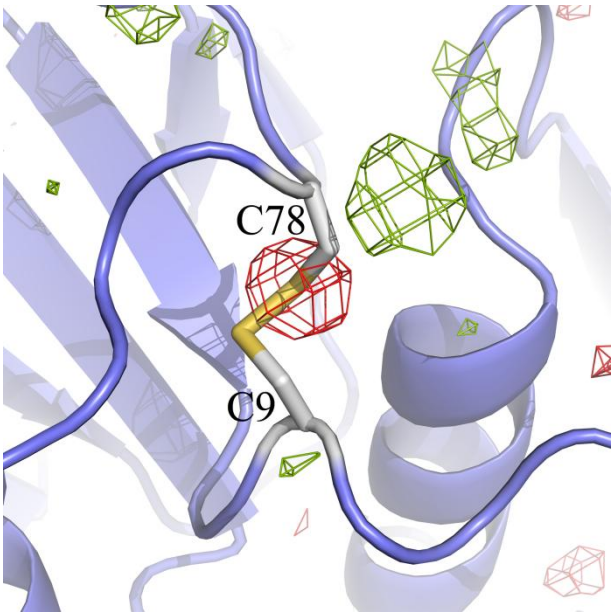

Fig. S10. Electron density around the radiation-damaged disulfide bridge C9-C78 of chain A. The positive difference density was contoured at 3σ (green), and the negative density at -3σ (red) (drawing with PYMOL (DeLano, 2002)).

References

Collaborative Computational Project, Number 4. (1994). *Acta Cryst. D* **50**, 760-763.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.

Dall'Antonia, F. & Schneider, T.R. (2006). *J. Appl. Cryst.* **39**, 618-619.

Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486-501.

Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795-800.

Lamzin, V.S. & Wilson, K.S. (1993). *Acta Cryst. D* **49**, 129-147.

Weiss, M.S. (2001) *J. Appl. Cryst*. **34**,130-135.

Sheldrick, G.M. (2010). *Acta Cryst. D* **66**, 479-485.