



Published in final edited form as:

Science. 2012 May 25; 336(6084): 1033–1037. doi:10.1126/science.1218753.

Structures from Anomalous Diffraction of Native Biological Macromolecules

Qun Liu¹, Tassadite Dahmane², Zhen Zhang², Zahra Assur², Julia Brasch², Lawrence Shapiro², Filippo Mancia³, and Wayne A. Hendrickson^{1,2,3,4,*}

¹New York Structural Biology Center, NSLS X4, Building 725, Brookhaven National Laboratory, Upton, NY 11973

²Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032

³Department of Physiology and Cellular Biophysics, Columbia University, New York, NY 10032

⁴Howard Hughes Medical Institute, Columbia University, New York, NY 10032

Abstract

Crystal structure analyses for biological macromolecules without known structural relatives entail solving the crystallographic phase problem. Typical *de novo* phase evaluations depend on incorporating heavier atoms than those found natively; most commonly, multi- or single-wavelength anomalous diffraction (MAD or SAD) experiments exploit selenomethionyl proteins. Here we realize routine structure determination using intrinsic anomalous scattering from native macromolecules. We devised robust procedures for enhancing signal-to-noise in the slight anomalous scattering from generic native structures by combining data measured from multiple crystals at lower-than-usual x-ray energy. Using this multi-crystal SAD method (5–13 equivalent crystals), we determined structures at modest resolution (2.8Å–2.3Å) for native proteins varying in size (127–1148 unique residues) and number of sulfur sites (3–28). With no requirement for heavy-atom incorporation, such experiments provide an attractive alternative to selenomethionyl SAD experiments.

Crystallographic structure determinations for biomolecules require the retrieval of phases, which are lost when measuring x-ray diffraction patterns. For the first protein crystal structures, phase evaluation was by the method of multiple isomorphous replacement (MIR) with derivatives incorporating mercury ($Z = 80$) or other heavy atoms. Once many structures were known, phases could often be estimated by the method of molecular replacement; however, *de novo* structure determination remained essential for molecules without adequately close structural relatives. Multiwavelength anomalous diffraction (MAD) analyses (1), which exploit element-specific scattering from x-ray resonance with atomic orbitals, came to be used increasingly for *de novo* structures as tunable synchrotron beamlines developed (2). Whereas MAD gives definitive phase information, its single-wavelength counterpart SAD is ambiguous in defining only trigonometric sines of phases.

*To whom correspondence should be addressed: wayne@convex.hhmi.columbia.edu.

Supporting Online Material

www.sciencemag.org/cgi/content/full/.....

Materials and Methods

SOM Text

Figures S1–S4

Tables S1–S10

References 24–43

This phase ambiguity could be resolved once density modification procedures, based largely on molecular boundaries and symmetry, were devised (3,4); and SAD then surged (5). MAD and SAD now dominate *de novo* phasing, having the advantage that lighter atoms can be effective sources of phasing signals. Selenomethionine is easily incorporated into proteins (6), and selenium ($Z = 34$) is now by far the most used phasing element (2). With MAD and SAD, metal atoms such as iron ($Z = 26$) present in some native proteins can also suffice.

Sulfur ($Z = 16$) is the heaviest element in most native proteins. Its K-shell resonance at 2.47 keV ($\lambda = 5.02\text{\AA}$) is inaccessible to standard MAD experiments and its anomalous scattering at conventional wavelengths is slight; nevertheless, sulfur anomalous scattering can suffice for SAD phasing. The structure of crambin was the first to be determined from sulfur SAD phasing (7), although the experiment was not then identified as SAD. Later, broader effectiveness of sulfur SAD was demonstrated with tests on lysozyme (8) and in solving the structure of obelin (9). Similarly, the feasibility of phosphorous SAD was demonstrated for nucleic acids (10). The motivation for truly routine native SAD is great since heavy-atom incorporations are often problematic, even for the most reliable selenomethionine. Subsequent optimization of native-SAD experiments has included developments for low-energy measurements (11), assessments of the impact of high data redundancy (10,11), optimal wavelength selection (12), control of complications from radiation damage (13,14), and the use of home-source CrK α radiation (15).

Besides test cases and technical developments, some novel protein structures beyond crambin and obelin have been determined by sulfur SAD analyses. As compared to the swelling numbers of SAD structures in general, however, the sulfur SAD component is small and not growing. Our February 2012 compilation of light-only SAD structures (all atoms with $Z \leq 20$) as reported in the Protein Data Bank (<http://www.pdb.org>) reveals 57 novel structures determined since 1981 (Table S1). Although more than 30 novel native-SAD structures were reported between 2004 and 2007, there have been only six after 2009 (not counting deposits reported here). Most of the Table 1 native SAD structures (42 of 57) are at high resolution ($d_{\min} < 2.0\text{\AA}$); only two have $d_{\min} > 2.3\text{\AA}$. Native SAD numbers are to be compared with overwhelming SAD and MAD PDB deposits overall (>5000 each since 1996). Why is the sulfur SAD contribution so meager? Surely a major factor is the low strength of anomalous scattering signals from sulfur as compared to selenium, the mainstay of SAD and MAD experiments. With one sulfur atom for every 30 residues, the average metal-free protein will produce a Bijvoet diffraction ratio ($\langle |F_{\pm h}| \rangle / \langle |F| \rangle$; (1)) of only ~1.0% even for 7 keV x-rays as compared to ~6.5% for a corresponding average selenomethionyl protein at the Se K-edge (12.658 keV). When compounded with noise from counting statistics, diffuse scattering, absorption, radiation damage and many other sources, such feeble signals are difficult to measure with sufficient accuracy to place the dozens of atomic positions in a typical sulfur substructure and then to obtain phases for full structure analysis.

The ratio of signal-to-noise in diffraction measurements can be enhanced by increasing data multiplicity; however, radiation damage is the enemy of multiplicity as crystal deterioration limits useful redundancy (16). We previously devised a procedure for combining the SAD data from multiple crystals to increase multiplicity without added radiation damage, whereby we were able to solve a relatively large and poorly diffracting selenomethionyl protein structure (17). For light-only native-SAD phasing, the anomalous signals are much weaker than those from Se; thus, sample preparation, data collection and data analysis require much more care. Here, we describe procedures for robust structure evaluation from the small anomalous signals of only-light-atom molecules and apply these procedures in determining four protein structures.

Anomalous signals for light elements increase steadily with decreasing energy; notably, f (the determinative, imaginary component of anomalous scattering) for sulfur increases from 0.24 electrons (e) at the Se K edge (12.658 keV) to 0.73 e at 7 keV and to 1.93 e at 4 keV (Table S2). Thus, native SAD analyses benefit from low-energy experiments. Enhanced f comes at the expense of increased x-ray absorption and incoherent scattering, however, which also increases with decreasing energy (Table S2). All materials in all beam paths contribute, including the sample itself and gases between the sample and detector. Absorption diminishes diffracted signals and incoherent scattering dramatically increases the background beneath integrated Bragg intensities; both reduce the signal-to-noise ratio. To optimize the experimental design, we estimated transmitted anomalous signals as a function of x-ray energy and sample size (Fig. S1). For samples of the sizes used in these studies (100–300 microns), the optimal x-ray energy is at 6–7 keV and we chose to use 7.112 keV x-rays as defined by the Fe K edge; for smaller samples and microbeams, lower energies become optimal assuming appropriate detector geometry. To minimize diffuse scattering, we introduced a helium-filled cone in the sample-to-detector beam path and conscribed the beam size to match crystal size. To enhance signal-to-noise ratios, we adopted an inverse-beam data collection strategy (18), which reduces systematic errors by measuring Friedel mates with equivalent geometry, and we used a multi-crystal strategy (17) to mitigate random errors by greatly increasing data multiplicity. To avoid crystal variation, we devised measures to assure statistical equivalence of all included crystals. To minimize effects of radiation damage, we merged data in successive wedges and excluded highly deteriorated wedges.

We undertook to develop our multi-crystal native SAD procedures by determining four crystal structures; experimental details are given in Table 1 and in Materials and Methods (SOM). Three of these applications solved protein structures without homologs of known structure (HK9_S, CysZ, netrin G2) and one (TorT/TorS_S) ‘re-solved’ a challenging test problem solved previously by other methods. One (CysZ) is an integral membrane protein and the others are domains of membrane receptors.

Crystal variation is always a concern for structure determination from multiple crystals, especially since variations may arise upon rapid freezing (19); nevertheless, we did not observe adverse effects in our earlier eight-crystal mergings (17). Here as before, the crystals did differ somewhat in various parameters, including unit cell dimensions, $I/|I|$, R_{merge} , anomalous CC and F_0/F (Tables S3 – S6). We evaluated these crystal variations, testing for statistical equivalence in three characteristics: unit cell variations, overall diffraction dissimilarity (1.0 – pairwise intensity correlation coefficient), and the relative anomalous correlation coefficient (RACC). RACC correlates the Bijvoet differences from an individual data set with those of the data merged from all crystals. Remarkably, we found evidence for rejecting only one data set (CysZ data set 8) as an outlier.

Fig. 1 shows clustering analyses on unit cell variations (left panels), overall diffraction variations (center panels) and anomalous differences (right panels) for 6 crystals of HK9_S (panels A, B & C), 8 crystals of CysZ (panels D, E & F), 5 crystals of netrin G2 (panels G, H & I), and 13 crystals of TorT/TorS_S (panels J, K & L). Variations among crystals within an experiment were small except for data set 8 of CysZ, which deviated from other CysZ crystals both for unit cell parameters ($> 4\%$) and diffracted intensities ($> 15\%$). The next biggest deviations were from data 2 of TorT/TorS_S where deviations were within 2.2% for unit cell parameters and within 5% for overall diffraction patterns, indicating their compatibility.

Although the anomalous correlation coefficient (ACC) was a most effective measure of SAD phasing efficacy in our previous SeMet study (17), ACC proved to be unreliable for

many single native-SAD data sets, with overall ACC values approaching zero. Thus, we devised the relative ACC (RACC) to provide a more robust measure of anomalous signals. As shown in Tables S3 – S6, typical RACC values are at or above 40%; CysZ outlier data set 8 was the one exception (RACC = 24.1%). For the four crystal systems in this study, we found that rejection criteria of unit cell deviations > 3 , overall diffraction dissimilarity $> 5\%$, and RACC $< 35\%$ were satisfactory and consistent with one another for outlier identification. A combinatorial criterion could be more robust, and appropriate rejection criteria could vary with characteristics of specific native-SAD experiments. Populated sub-clusters of data sets might be identified to yield satisfactorily valid but non-isomorphous structures.

Individual data sets from accepted crystals were further checked for exclusion of radiation-damaged wedges, and for individual outlier rejections upon scaling and merging with the program *SCALA* (20). For each application, individual data sets were reordered according to overall RACC, and mergings were made by successively adding one data set at a time, both best to worst (Tables S7A-S10A) and worst to best (Tables S7B-S10B), followed by attempts at structure determination. In general, substructures were determined at reduced resolution (3.0Å to 3.9Å) and then used for phasing at the data limit. In each case, assessment parameters improved with successive mergings once substructure determination was feasible; and measures of anomalous signal strength (Fig. S2) and diffracted intensity precision (Fig. S3) were appreciably enhanced for merged data versus individual data sets. Most importantly, substructure determinations by *SHELXD* (21) were robust in all cases after data merger (Figs. 2A,C,E&G) and resulting phases were accurate as evidenced by Bijvoet-difference peak profiles (Fig. 2B,D,F&H) and electron density maps (Fig. S4). Each of the multi-crystal electron density distributions supported automated chain tracing at a high level of completion (77–93%; Table 1). Phasing efficacy increased with number of crystals and with multiplicity ordered by data wedges, but it appeared to reach an asymptotic limit (Fig. 3). Map quality continued to improve even at 150-fold data multiplicity.

The four crystal structures determined here are shown as ribbon diagrams in Fig. 4, including the sulfur, chlorine, calcium and sulfate substructures identified and used in the SAD analyses. These structures range from 127 to 1148 ordered residues, are in crystal symmetries from monoclinic to tetragonal, and have resolution limits from 2.8Å to 2.3Å. Full descriptions of the structures and associated biology are being published elsewhere, including the structure of netrin G2 as refined at higher resolution (22) and the prior structure of TorT/TorS_S (23). Each structure was unknown previously; both HK9_S and netrin G2 had resisted structure determination by other methods; CysZ is an unprecedented kind of membrane protein; and the TorT/TorS_S complex is both large and at modest resolution (2.8Å). The range of these applications indicates substantial generality and robustness.

Although the multi-crystal native SAD procedure was effective as implemented here, our experimental setup was not ideal. These results are from a bending-magnet beamline on a second-generation synchrotron source. As suggested by Fig. S1 and Table S1, substantially better performance can be expected by microdiffraction at an advanced undulator beamline that is optimized for low-energy experiments, including the provision of a detector designed to capture high-angle diffraction effectively. Improvements in scaling and weighting procedures can also be anticipated.

Notwithstanding prospects for improvement, we suggest that multi-crystal SAD phasing can facilitate robust *de novo* determination of native proteins and nucleic acids at a level rivaling for selenomethionyl proteins at present. The TorT/TorS_S complex has a size (~1200 residues) and a resolution limit (2.8 Å) that are at a complexity level exceeding more than

90% of entries in the current PDB. While difficulties may arise in obtaining sufficient numbers of adequate crystals to support multi-crystal determinations, advantages will accrue in being liberated from the need for derivatization or selenomethionine incorporation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Randy Abramowitz and John Schwanof for help with synchrotron data collection. This work was supported in part by cooperative agreement GM095315 from the Protein Structure Initiative (CysZ), and by NIH grants GM034102 (HK_{9s} and TorT/TorS_s) and GM062270 (netrin G2). Beamline X4A of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory, a DOE facility, is supported by the New York Structural Biology Center. Accession codes for PDB deposits are listed in Table 1.

References

1. Hendrickson WA. *Science*. 1991; 254:51. [PubMed: 1925561]
2. Hendrickson WA. *J. Synchrotron Rad.* 1999; 6:845.
3. Wang B-C. *Methods Enzymol.* 1985; 115:90. [PubMed: 4079800]
4. Cowtan KD, Main P. *Acta Crystallogr. D.* 1993; 49:148. [PubMed: 15299555]
5. Dauter Z, Dauter M, Dodson E. *Acta Crystallogr. D.* 2002; 58:494. [PubMed: 11856836]
6. Hendrickson WA, Horton JR, Lemaster DM. *EMBO J.* 1990; 9:1665. [PubMed: 2184035]
7. Hendrickson WA, Teeter MM. *Nature*. 1981; 290:107.
8. Dauter Z, Dauter M, de La Fortelle E, Bricogne G, Sheldrick GM. *J. Mol. Biol.* 1999; 289:83. [PubMed: 10339407]
9. Liu ZJ, et al. *Protein Sci.* 2000; 9:2085. [PubMed: 11152120]
10. Dauter Z, Adamiak DA. *Acta Crystallogr. D.* 2001; 57:990. [PubMed: 11418767]
11. Weiss MS, Sicker T, Hilgenfeld R. *Structure*. 2001; 9:771. [PubMed: 11566127]
12. Mueller-Dieckmann C, Panjikar S, Tucker PA, Weiss MS. *Acta Crystallogr. D.* 2005; 61:1263. [PubMed: 16131760]
13. Weiss MS, Panjikar S, Mueller-Dieckmann C, Tucker PA. *J. Synchrotron Rad.* 2005; 12:304.
14. Cianci M, Helliwell JR, Suzuki A. *Acta Crystallogr. D.* 2008; 64:1196. [PubMed: 19018096]
15. Yang C, Pflugrath JW, Courville DA, Stence CN, Ferrara JD. *Acta Crystallogr. D.* 2003; 59:1943. [PubMed: 14573949]
16. Garman EF. *Acta Crystallogr. D.* 2010; 66:339. [PubMed: 20382986]
17. Liu Q, Zhang Z, Hendrickson WA. *Acta Crystallogr. D.* 2011; 67:45. [PubMed: 21206061]
18. Hendrickson WA, et al. *Proc. Natl. Acad. Sci. USA.* 1989; 86:2190. [PubMed: 2928324]
19. Garman E. *Acta Crystallogr. D.* 1999; 55:1641. [PubMed: 10531512]
20. Evans P. *Acta Crystallogr. D.* 2006; 62:72. [PubMed: 16369096]
21. Schneider TR, Sheldrick GM. *Acta Crystallogr. D.* 2002; 58:1772. [PubMed: 12351820]
22. Brasch J, Harrison OJ, Ahlsen G, Liu Q, Shapiro L. *J. Mol. Biol.* 2011; 414:723. [PubMed: 22041449]
23. Moore JO, Hendrickson WA. *Structure*. citation details to be added.

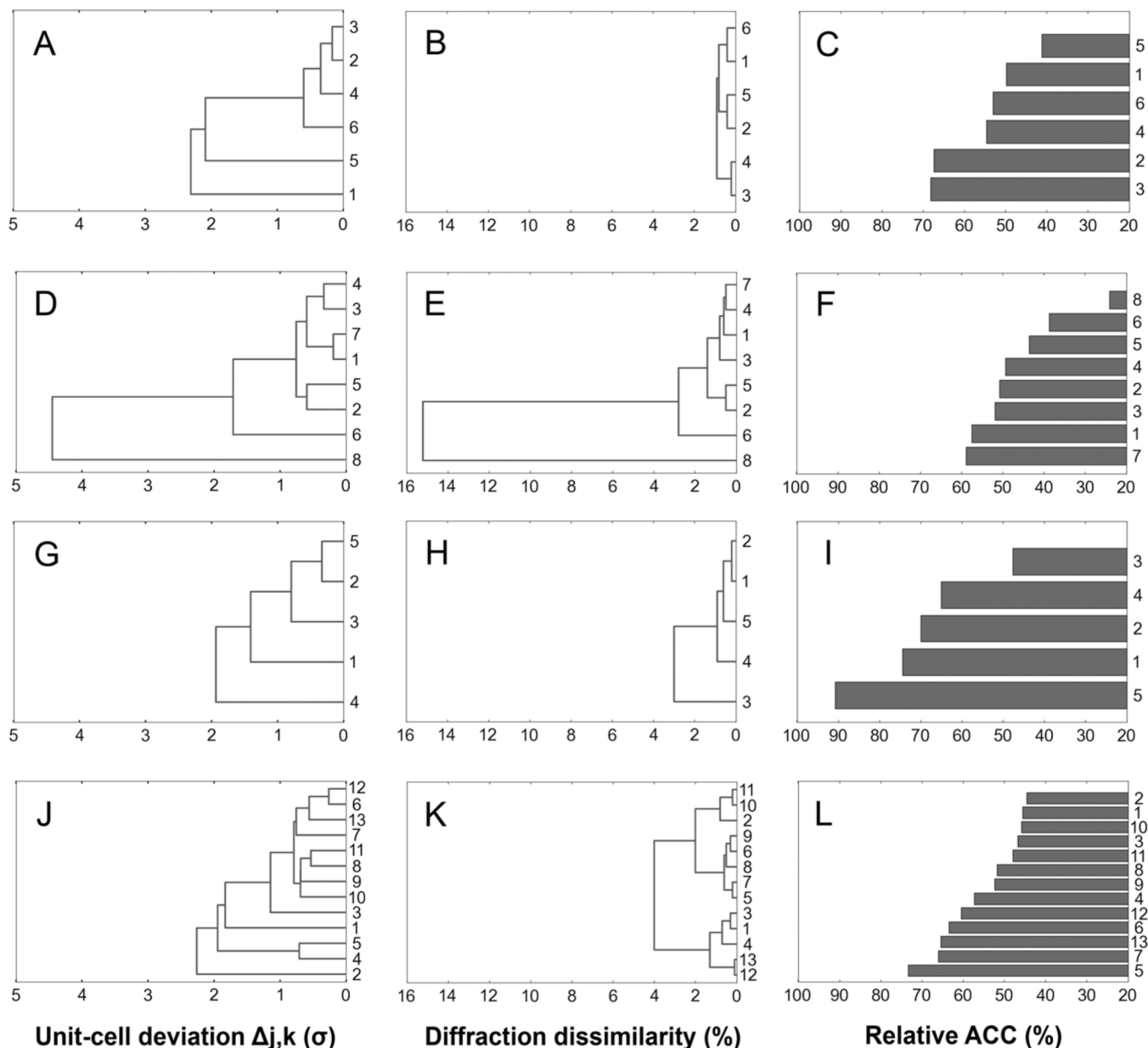


Fig. 1. Variations among crystals from multi-crystal data sets

(Left) Cluster analyses of unit cell variations. **(Center)** Cluster analyses of overall diffraction dissimilarity. **(Right)** Relative anomalous correlation coefficient. **(A)**, **(B)** and **(C)** HK9S. **(D)**, **(E)** and **(F)** CysZ. **(G)**, **(H)** and **(I)** netrin G2. **(J)** **(K)** and **(L)** TorT/TorS₅. Unit cell variations are standard Euclidean distances normalized by population variances, i.e. the distance between j and k of among N crystals is $d_{j,k} = \left(\frac{1}{N} \sum_{i=1}^N [(u_{i,j} - u_{i,k})^2 / V_i] \right)^{1/2}$ where u_i includes all variable unit cell parameters i , each having a variance of $V_i = \frac{1}{N} \sum_{k=1}^N (u_k - \bar{u}_k)^2$. Overall diffraction dissimilarity between crystals j and k is defined as $D_{i,j} = 1.0 - C_{ij}$ where C_{ij} is the correlation coefficient between all Bragg intensities in common between the two diffraction patterns. For C_{ij} calculations, high-angle data cut-offs were at 3.0 Å for HK9S, CysZ and netrin G2 and 3.9 Å for TorT/TorS₅. Clustering calculations were made in a single-linkage (minimum) mode for unit cell analyses and in a complete-linkage (maximum) mode for diffraction analyses. The linkage line connecting any two clusters

defines the 'distance' between those clusters. The relative anomalous correlation coefficient (Relative ACC) compares Bijvoet differences from an identified individual data set with those in the data set merged from all crystals.

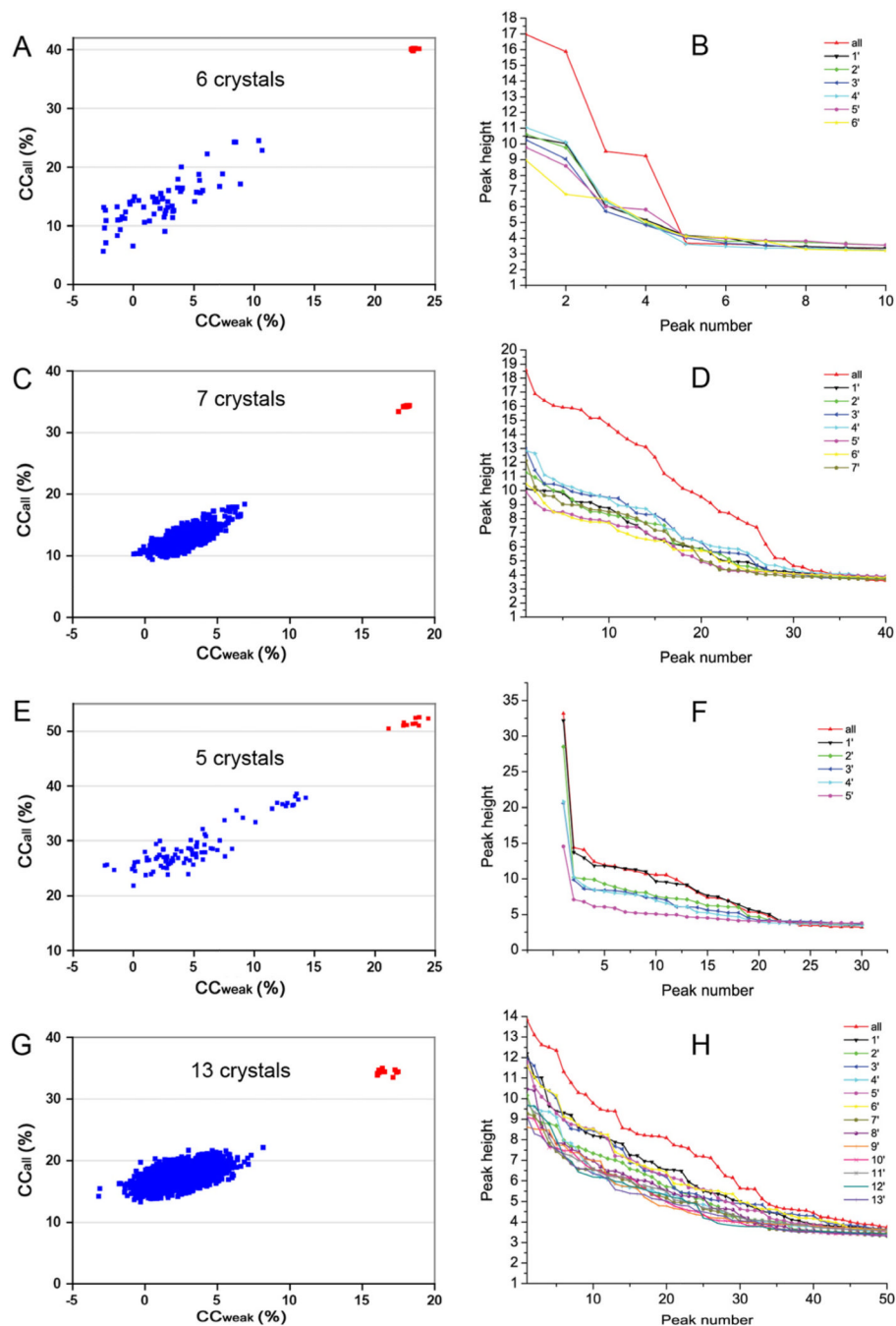


Fig. 2. Analysis of anomalously scattering substructures

(Left) Profiles of *SHELXD* correlation coefficients (CC) between observed and calculated Bijvoet differences. **(A)** HK9_S, 35 solutions in 100 tries. **(C)** CysZ, 34 solutions in 1000 tries. **(E)** Netrin G2, 10 solutions in 100 tries. **(G)** TorT/TorS_S, 14 solutions in 2000 tries. Successful solutions are colored in red; random solutions are colored in blue. **(Right)** Bijvoet-difference Fourier peak profiles. **(B)** HK9_S, 10 highest peaks. **(D)** CysZ, 40 highest peaks. **(F)** Netrin G2, 30 highest peaks. **(H)** TorT/TorS_S, 50 highest peaks. For each experiment, ordered peak-height profiles are shown for maps from each single-crystal data set (identified by inset keys) and for the map from the merged data set (red in all cases).

Peak heights are given in units of root-mean-squared deviation over the entire respective Fourier syntheses.

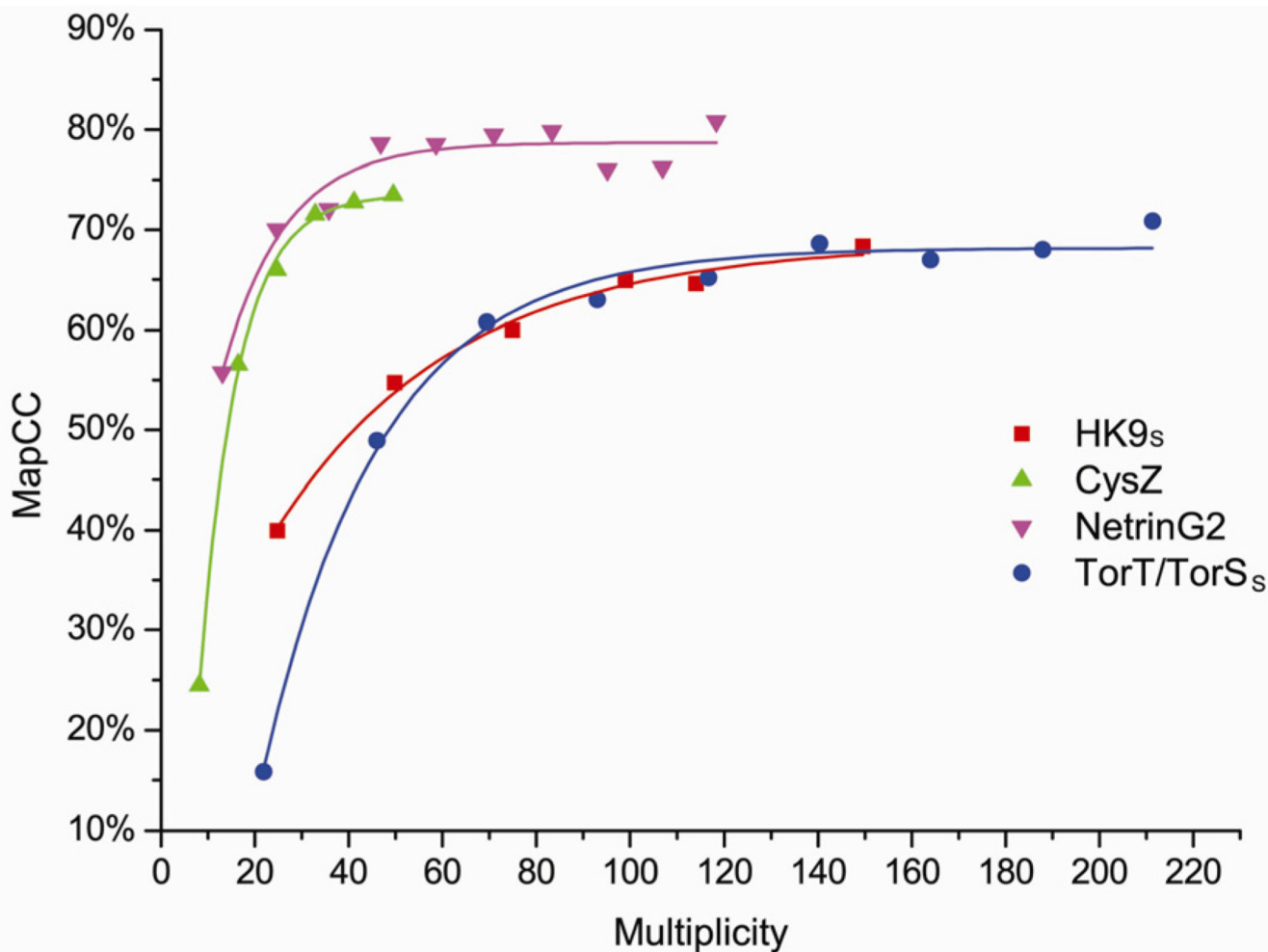


Fig. 3. Phasing efficacy dependence on data multiplicity

Phasing efficacy is measured here by the map correlation coefficient (mapCC) and average data multiplicity is given in order of wedges of accumulated diffraction data. Each individual data set was divided into wedges of sequentially measured frames (6 wedges for HK9_s, 6 wedges for CysZ, 10 wedges for netrin G2, and 9 wedges for TorT/TorS_s), and these data were then merged wedge by wedge. Accumulations from these wedges were then used for native-SAD phasing based on substructures obtained before from all data. For each accumulation, experimental electron densities after density modification were compared with model electron densities. Resultant mapCCs for these wedge structures are plotted with respect to multiplicity for the respective accumulations: HK9_s (red), CysZ (green), netrin G2 (magenta) and TorT/TorS_s (blue). Fittings are with an asymptotic formula described in SOM Text.

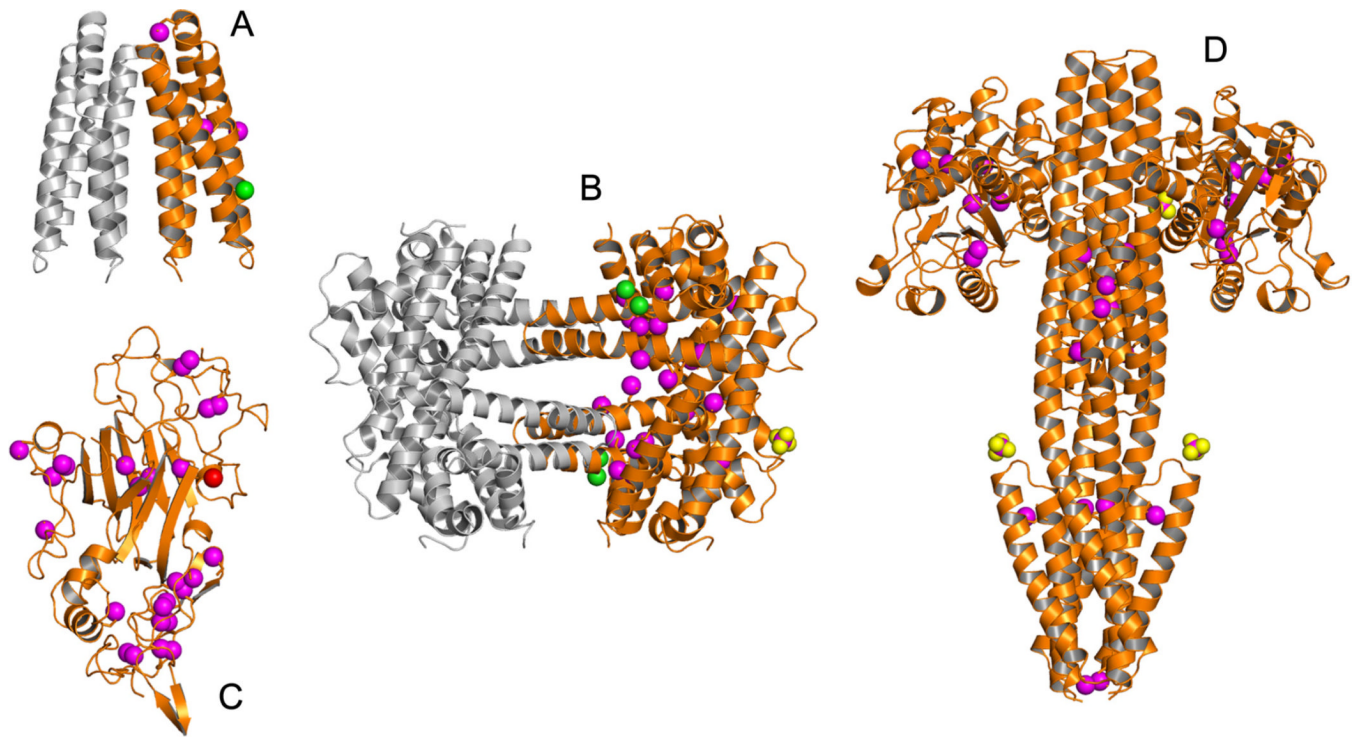


Fig. 4. Native SAD crystal structures

(A) HK9_S; a substructure of 3 S atoms and 1 Cl atom defined 127 ordered residues at 2.3 Å resolution. (B) CysZ, a substructure of 20 S atoms, 4 Cl atoms and 1 SO₄ defined 452 ordered residues at 2.3 Å resolution. (C) Netrin G2; a substructure of 26 S atoms and 1 Ca atom defined 312 ordered residues at 2.3 Å resolution. (D) TorT/TorS_S, a substructure of 28 S atoms and 3 SO₄ defined 1148 ordered residues at 2.8 Å resolution. Each molecular oligomer or complex is shown as a ribbon diagram with those residues in the asymmetric unit colored orange. Anomalous scattering substructures are shown as spheres with sulfur atoms in magenta, chloride atoms in green (HK9_S & CysZ), sulfate ions in yellow and magenta (CysZ & TorT/TorS_S) and the one calcium ion in red (netrin G2).

Table 1

Summary of native-SAD structure determinations

	HK9_s	CysZ	Netrin G2	TorT/TorS_s
Number of crystals	6	7	5	13
Resolution (Å)	2.3	2.3	2.3	2.8
Unique residues	157	492	332	1162
Unique sulfur atoms	3	22	26	28
$\langle F_{+h} \rangle / \langle F \rangle$ (%) ¹	0.74	1.12	1.49	0.83
Diffraction statistics	Table S3	Table S4	Table S5	Table S6
Substructure success (%) ²	35.0	3.4	1.1	0.7
MapCC after DM (%) ³	65.5	72.3	80.8	71.3
Auto-built residues ⁴	98/127	408/453	291/312	965/1148
R	0.187	0.200	0.203	0.184
R _{free}	0.227	0.238	0.241	0.242
PDB code ⁵	3VA9	3TX3	3TBD	3O1I

¹Estimated at zero scattering angle from *a priori* contents (6)

²*SHELXD* solution rate from 100 tries for HK9_s and netrin G2, 1000 tries for CysZ and 2000 tries for TorT/TorS_s.

³Correlation coefficient between density-modified (*DM*, no averaging) experimental- and model-phased maps

⁴Model building (successes / refined as ordered) by *ARP/wARP* except by *Buccaneer* for TorT/TorS_s

⁵PDB code 3TBD was refined at a higher resolution and 3O1I was refined previously from other data