STRUCTURAL BIOLOGY
ISSN 2059-7983

Obtaining anomalous and ensemble information from protein crystals from 220 K up to physiological temperatures

Tzanko Doukov,[a]* Daniel Herschlag[b,c,d] and Filip Yabukarski[b,e]*

[a]SMB, Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA, [b]Deparment of Biochemistry, Stanford University, Stanford, CA 94305, USA, [c]Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA, [d]Stanford ChEM-H, Stanford University, Stanford, CA 94305, USA, and [e]Bristol-Myers Squibb, San Diego, CA 92121, USA. *Correspondence e-mail: tdoukov@slac.stanford.edu, fyabukarski@gmail.com

Received 3 November 2022
Accepted 31 January 2023

Edited by R. Steiner, King's College London, United Kingdom and University of Padova, Italy

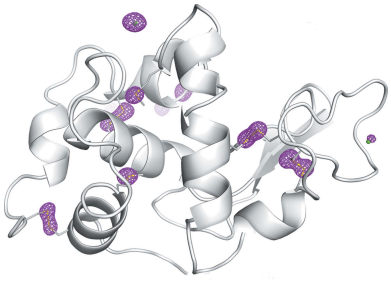
Keywords: native phasing; anomalous diffraction; room temperature; data collection.

PDB references: lysozyme, 7.1 keV, 220 K, 8ezo; 240 K, 8f0b; 260 K, 8ezp; 273 K, 8ezu; 293 K, 8ezx; 12 eV, 293 K, 8f00; thaumatin, 7.1 keV, 293 K, 8f01; 12 eV, 293 K, 8f03; proteinase K, 7.1 keV, 293 K, 8f05; 310 K, 8f06; 12 keV, 273 K, 8f07

Supporting information: this article has supporting information at journals.iucr.org/d

X-ray crystallography has been invaluable in delivering structural information about proteins. Previously, an approach has been developed that allows high-quality X-ray diffraction data to be obtained from protein crystals at and above room temperature. Here, this previous work is built on and extended by showing that high-quality anomalous signal can be obtained from single protein crystals using diffraction data collected at 220 K up to physiological temperatures. The anomalous signal can be used to directly determine the structure of a protein, *i.e.* to phase the data, as is routinely performed under cryoconditions. This ability is demonstrated by obtaining diffraction data from model lysozyme, thaumatin and proteinase K crystals, the anomalous signal from which allowed their structures to be solved experimentally at 7.1 keV X-ray energy and at room temperature with relatively low data redundancy. It is also demonstrated that the anomalous signal from diffraction data obtained at 310 K (37°C) can be used to solve the structure of proteinase K and to identify ordered ions. The method provides useful anomalous signal at temperatures down to 220 K, resulting in an extended crystal lifetime and increased data redundancy. Finally, we show that useful anomalous signal can be obtained at room temperature using X-rays of 12 keV energy as typically used for routine data collection, allowing this type of experiment to be carried out at widely accessible synchrotron beamline energies and enabling the simultaneous extraction of high-resolution data and anomalous signal. With the recent emphasis on obtaining conformational ensemble information for proteins, the high resolution of the data allows such ensembles to be built, while the anomalous signal allows the structure to be experimentally solved, ions to be identified, and water molecules and ions to be differentiated. Because bound metal-, phosphorus- and sulfur-containing ions all have anomalous signal, obtaining anomalous signal across temperatures and up to physiological temperatures will provide a more complete description of protein conformational ensembles, function and energetics.

1. Introduction



X-ray crystallography using cryocooled protein crystals and the traditional single-conformation structural models obtained from diffraction data are a cornerstone of biological sciences and have been invaluable in delivering structural information for thousands of proteins (Fersht, 1985; Berg *et al.*, 2019). Nevertheless, proteins exist as conformational ensembles with relative populations defined by the free-energy landscape (Frauenfelder *et al.*, 1991; Austin *et al.*, 1975; Alber *et al.*, 1983; Hammes *et al.*, 2011; Wei *et al.*, 2016). Thus, ensembles rather than single structures (as traditionally obtained in X-ray

crystallography) relate to the energetics of the system and there is a need for ensemble information to more deeply explore biological function (Frauenfelder *et al.*, 1979, 1988; Frauenfelder & Petsko, 1980).

The majority of X-ray diffraction data have been obtained from cryocooled protein crystals to mitigate X-ray damage and allow more diffraction data to be collected compared with crystals at room temperature (Hope, 1988). Similarly, anomalous signal has traditionally been obtained from cryocooled crystals because of their resilience towards X-ray damage compared with protein crystals at room temperature. However, there are limitations such as the time and resources, including the search for an optimal cryoprotection protocol, that are available for each type of crystal, and in some cases it was not possible to establish efficient cryogenic conditions (Ravelli & McSweeney, 2000; Rice *et al.*, 2000; Garman, 2003). In addition, cryocooling can quench motions and alter conformational states (Fraser *et al.*, 2011; Keedy *et al.*, 2014; Ringe & Petsko, 2003; Tilton *et al.*, 1992).

Few room-temperature (RT) native phasing experiments have been reported following the pioneering work by Hendrickson (Hendrickson & Teeter, 1981) and Cianci (Cianci *et al.*, 2004). With respect to SAD phasing at RT, the recent work by Greisman *et al.* (2022) is of particular interest. The data-collection approach used by Greisman and coworkers is largely based on traditional methods for anomalous data collection using large crystals inside MicroRT sleeves (MiTeGen), 5 × 30 µm (FWHM) helical data collection at 6.55 keV energy at a synchrotron and 26–35-fold redundant data sets. Here, we report an approach based on a recently developed method for data collection at synchrotrons across temperature which differs from traditional data-collection methods and has been optimized for rapid data collection with minimized X-ray damage (Doukov *et al.*, 2020; Yabukarski, Doukov, Mokhtari *et al.*, 2022). Given the recent emphasis on obtaining conformational ensemble information to understand function (Henzler-Wildman *et al.*, 2007; Klinman, 2015; Yabukarski, Doukov, Pinney *et al.*, 2022), RT and physiological temperature (PT) X-ray diffraction data can be used to obtain conformational heterogeneity information for proteins (van den Bedem *et al.*, 2009; Riley *et al.*, 2021; Lang *et al.*, 2010; Burnley *et al.*, 2012; Ploscariu *et al.*, 2021), with the conformational heterogeneity being the experimental manifestation of the ensemble nature of proteins (Keedy, Fraser *et al.*, 2015).

While it is now accepted that RT (and PT) data collection provides more physiologically relevant structural information by eliminating cryocooling artefacts (Keedy *et al.*, 2014; Keedy, 2019; Fraser *et al.*, 2011; Keedy, Fraser *et al.*, 2015), new structures for which adequate molecular-replacement models are not available are still being phased based on naturally present or added anomalous scattering atoms using cryocooled crystals, requiring first the collection of data sets from cryocooled crystals to determine the structure (for macromolecules with unknown structures) and subsequent RT data collection to obtain conformational ensemble information. An ability to obtain diffraction information directly at RT (or any PT) to both solve the structure and obtain ensemble information from the same crystal would streamline the acquisition of conformational ensemble information.

Various experimental approaches have been developed to obtain experimental phases and determine the structures of proteins: multiple isomorphous replacement (MIR; Green *et al.*, 1954), multi-wavelength anomalous dispersion (MAD; Hendrickson, 1985; Kahn *et al.*, 1985) and single-wavelength anomalous dispersion (SAD; Hendrickson & Teeter, 1981; Hendrickson, 2014; Liu & Hendrickson, 2015). Currently, SAD is the most popular approach because obtaining experimental phases requires a single crystal and no heavy-atom soaks are required. Determining a new structure via SAD relies on obtaining an anomalous signal that is strong enough to derive experimental phases (Liu & Hendrickson, 2015; Terwilliger *et al.*, 2016). The diffraction data are usually collected close to the absorption edge of elements, where the anomalous signal is maximized, and high-quality data are required to obtain a sufficiently strong anomalous signal (Hendrickson, 2014). Fine-slicing and/or multi-data-set SAD data-collection techniques increase the redundancy and improve the signal-to-noise ratio of the anomalous signal from a single crystal (Rose & Wang, 2016). For weaker signals a multi-crystal data-collection approach can be utilized if proper merging of data sets from similar crystals is applied (Foadi *et al.*, 2013; Hirata *et al.*, 2019; Zander *et al.*, 2015; Gildea *et al.*, 2022). While the standard SAD experiment is carried out at cryogenic temperatures, it has recently been shown that anomalous signal and phasing information can be obtained at RT via X-ray free-electron laser (XFEL) experiments (Nass *et al.*, 2020; Colletier *et al.*, 2016) or via serial synchrotron X-ray crystallography (SSX) on third-generation microbeam beamlines (Foos *et al.*, 2018; Botha *et al.*, 2018). Nevertheless, limitations remain: access to XFELs and SSX is limited, the resolution tends to be lower compared with the resolutions of synchrotron data sets, and both XFEL and SSX experiments require sophisticated equipment and crystal-delivery systems.

Proteins are not isolated systems and the energetics of their function are defined by both their ensembles of conformational states and the interactions that they make with water, ions and other solution components at a specific temperature. Nevertheless, structural studies that provide conformational ensemble information often focus on the protein ensemble, at least in part due to difficulties in distinguishing between water molecules and other solution components (for example ions) and accurately modelling them. It is the ensemble of protein states and their hydration and interaction with ions that fully define the energetics of the system. While the electron density for large and/or high-affinity ligands is usually sufficiently clear to model these ligands unambiguously, the electron density for smaller ligands and ligands of lower affinity is often ambiguous (Pozharski *et al.*, 2013, 2017). Distinguishing and accurately modelling physiologically relevant ions is particularly challenging because the electron density for such atoms (*i.e.* sulfur, phosphorus, chlorine, sodium, potassium, calcium and magnesium) and the geometry of their interactions with protein groups are difficult to distinguish from the electron density for water molecules and the geometry of interactions

of water molecules with protein groups. The issue is exacerbated when water molecules sample alternative states (Vitkup *et al.*, 2000; Ringe & Petsko, 2003; Schirò & Weik, 2019). Thus, there is a need to develop approaches that allow not only conformational ensemble information about proteins to be obtained but also the fuller modelling of solution components such as ions that interact with proteins.

Anomalous signal can help to position ligands more accurately in the electron density, provided that anomalous scattering atoms are present. Further, anomalous signal can differentiate between ions that scatter anomalously and water molecules, which do not. Thus, obtaining high-quality anomalous signal directly from RT data is required to model the protein solvation layer and to position bound ions and ligands more accurately (Horowitz *et al.*, 2016; Salmon *et al.*, 2018). Further, RT data can be used to obtain more complete information on protein conformational heterogeneity since the anomalous signal from the sulfur in methionine and cysteine side chains can be used to better position them. More accurate models of the interactions of the solvation layer with proteins, together with modelling of the ensemble of protein states and determining how these change with ion conditions and temperature, will allow us to better understand protein structural properties and energetics. Thus, there is a need for approaches to obtain high-quality anomalous signal and ensemble information from protein crystals across physiological temperatures. Ideally, these approaches would be easy to implement, broadly applicable across systems and carried out on commonly used synchrotron beamlines. Here, we build on recent methodological and technical advances (Casanas *et al.*, 2016; Leonarski *et al.*, 2018; Doukov *et al.*, 2020) and show that high-quality anomalous signal can be obtained in minimal diffraction data sets obtained across physiological temperatures and can be used to obtain experimental phases and directly solve a new structure experimentally. The anomalous signal also allowed us to distinguish between ions and water molecules and identify ion types.

2. Methods

2.1. Crystallization

Hen egg-white lysozyme (catalogue No. L4919), *Tritirachium album* proteinase K (catalogue No. P2308) and *Thaumatococcus daniellii* thaumatin (catalogue No. T7638) were procured from Sigma and crystallized at room temperature using hanging-drop (proteinase K and lysozyme) and sitting-drop (thaumatin) setups.

2.1.1. Lysozyme crystallization. Hen egg-white lysozyme powder was dissolved in 0.1 *M* sodium acetate buffer pH 4.5 to a concentration of 100 mg ml^{-1}. 5 µl protein solution was mixed with 5 µl 0.6 *M* sodium chloride in 100 m*M* sodium acetate buffer and 25% ethylene glycol on a cover slip in a hanging-drop setup and equilibrated against 0.7 ml saturated ammonium sulfate solution (4.1 *M*). Crystals appeared within hours and grew to maximum size within two days.

2.1.2. Proteinase K crystallization. *T. album* proteinase K was dissolved to 10 mg ml^{-1} in 0.05 *M* Tris–HCl buffer pH 7.5. 2 µl proteinase K was mixed with 2 µl 1 *M* ammonium sulfate in a hanging-drop setup.

2.1.3. Thaumatin crystallization. *T. daniellii* thaumatin was dissolved to 20 mg ml^{-1} in distilled water. 5 µl protein solution was mixed with 5 µl 24% potassium sodium tartrate tetrahydrate, 15%(*v*/*v*) ethylene glycol, 0.1 *M* bis-Tris propane buffer pH 6.6 in a sitting-drop tray.

2.2. Mounting

Protein crystals with sizes of 250–500 µm were selected and fished out under Paratone N oil. Extensive removal of the water-based layer around the crystal was achieved under puddles of oil using a crystal loop of 300 µm in size as a scraper. The crystal was mounted on a double-thickness MiTeGen loop and the excess oil was tapped away until nothing but the loop and the crystal could be seen. Even a very thin oil layer served as an adhesive between the crystal and the loop. This process assured very good adhesion of the crystal to the loop and no movement (or sliding) during data collection, as well as lower overall X-ray scattered background.

2.3. Data collection

All data sets were collected on BL14-1 at Stanford Synchrotron Radiation Lightsource (SSRL) equipped with a home-built fast accurate goniometer, an EIGER 16M PAD detector and a coaxially installed Oxford Cryosystems 850 (70–400 K) cryogenic unit operating at 220 K (−53°C), 240 K (−33°C), 260 K (−13°C), 273 K (0°C), 293 K (20°C) or 310 K (37°C). An energy of 7.1 keV close to the iron absorption edge was used as a compromise between achievable high resolution (∼1.9 Å at the closest detector distance) and sufficient flux for the fast collection of excellent quality data, while sacrificing the S, Cl^- and K^+, Ca^{2+} anomalous signal strength relative to the lowest available X-ray energy (6 keV) at BL14-1. The beam stop was placed at a distance limiting the lower resolution to ∼40 Å, which is a compromise between collecting strong anomalous signal at low resolution and low background scattering. All data collections were carried out with the largest achievable beam size (250 × 80 µm) and a rotation speed depending on the crystal size: 0.04 s for larger crystals to 0.1 s for smaller crystals per 1° frame (10–25 Hz) in a shutterless mode. Such a rotation speed is close to the observed practical rotational 'speed limit' of 20° per second for obtaining high data quality at the Swiss Light Source (SLS; Casanas *et al.*, 2016). At least 1440° of data were collected per crystal with a total collection time of 57.6 s using shutterless mode while constantly rotating the sample without the usually beneficial 'inverse-beam' geometry setup, which adds nonproductive time to rotate the goniometer to the +180° step while the X-ray damage process still takes place. We found that increasing the multiplicity quickly improved the anomalous signal more than collecting Friedel pairs close in time. Individual crystal quality also mattered for the strength of the anomalous signal. The lysozyme crystal measured at 273 K

Table 1
Data collection, processing and refinement at 7.1 keV.

Protein	Proteinase K		Lysozyme					Thaumatin
Temperature (K)	293	310	220	240	260	273	293	293
Data collection								
Crystal dimensions (mm)	0.5, 0.4, 0.4	0.5, 0.3, 0.3	0.3, 0.25, 0.2	0.35, 0.35, 0.3	0.35, 0.3, 0.3	0.45, 0.4, 0.35	0.45, 0.4, 0.4	0.55, 0.25, 0.25
Wavelength (Å)	1.7462	1.7462	1.7462	1.7462	1.7462	1.7462	1.7462	1.7462
Resolution range (Å)	35.06–1.80 (1.84–1.80)	34.12–1.80 (1.84–1.80)	38.45–1.90 (1.94–1.90)	37.04–1.90 (1.94–1.90)	38.78–1.90 (1.94–1.90)	38.66–1.90 (1.94–1.90)	38.85–1.90 (1.94–1.90)	38.32–1.80 (1.84–1.80)
Rotation range (°)	120	240	240	240	180	80	80	120
ADWD dose (kGy)	7	13	53	12	4	6	6	7
MaxDose (kGy)	26	55	134	40	12	28	25	30
Collection time (s)	4.8	9.6	24	9.6	7.2	3.2	3.2	4.8
MaxDose rate (kGy s^{-1})	5.4	5.7	5.6	4.2	1.7	8.8	7.8	6.2
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_12_12$
a, b, c (Å)	68.01, 68.01, 102.46	67.95, 67.95, 102.39	76.91, 76.91, 36.96	77.11, 77.11, 37.06	77.50, 77.50, 37.18	77.32, 77.32, 37.08	77.71, 77.71, 37.12	58.96, 58.96, 151.24
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Total reflections	169716 (3770)	327762 (5707)	143956 (6712)	146595 (7364)	112154 (5574)	49317 (2433)	48911 (2347)	187201 (3759)
Multiplicity	7.5 (3.3)	14.4 (4.9)	15.6 (11.6)	15.8 (12.5)	11.9 (9.7)	5.3 (4.2)	5.6 (5.0)	7.3 (2.7)
Mosaicity (°)	0.08	0.32	0.23	0.15	0.12	0.22	0.16	0.08
Completeness (%)	98.7 (85.4)	98.8 (87.0)	99.9 (98.2)	100 (100)	99.9 (97.9)	99.5 (97.3)	92.9 (78.8)	99.5 (93.8)
Mean $I/\sigma(I)$	39.4 (12.1)	25.6 (3.7)	37.2 (8.5)	35.4 (12.6)	40.2 (19.7)	29.7 (9.1)	30.7 (9.9)	20.6 (1.9)
Wilson B factor (Å^2)	17.3	25.8	30.8	27.6	21.5	26.7	25.9	30.9
R_{merge}	0.033 (0.043)	0.060 (0.328)	0.046 (0.251)	0.053 (0.169)	0.048 (0.089)	0.028 (0.104)	0.033 (0.119)	0.046 (0.481)
$R_{p.i.m.}$	0.013 (0.044)	0.016 (0.170)	0.015 (0.102)	0.018 (0.067)	0.019 (0.030)	0.019 (0.082)	0.022 (0.084)	0.019 (0.334)
$CC_{1/2}$	0.999 (0.991)	0.998 (0.886)	0.999 (0.992)	0.999 (0.993)	0.998 (0.997)	0.999 (0.988)	0.997 (0.984)	0.999 (0.742)
ISa	31.6	27.5	37.2	25.9	21.4	51.6	35.2	37.1
Refinement								
Resolution (Å)	35.06–1.80 (1.85–1.80)	34.12–1.80 (1.85–1.80)	38.48–1.90 (1.95–1.90)	34.51–1.90 (1.95–1.90)	38.78–1.90 (1.95–1.90)	38.69–1.90 (1.95–1.90)	38.88–1.90 (1.95–1.90)	38.35–1.80 (1.85–1.80)
R_{work} (%)	10.9 (14.2)	11.8 (24.4)	15.7 (16.0)	13.7 (14.0)	13.0 (12.3)	13.4 (15.5)	13.0 (14.6)	14.1 (28.0)
R_{free} (%)	14.0 (21.4)	14.7 (34.6)	18.9 (23.0)	19.8 (20.9)	17.0 (19.8)	17.0 (25.7)	16.7 (23.7)	16.1 (27.4)
No. of non-H atoms								
Total	2422	2387	1141	1153	1166	1150	1151	1715
Macromolecules	2166	2128	1027	1032	1049	1053	1065	1582
Ligands	36	16	16	16	12	17	13	11
Water	220	240	97	105	108	80	76	122
R.m.s. deviations								
Bond lengths (Å)	0.011	0.010	0.009	0.007	0.008	0.008	0.008	0.008
Angles (°)	1.60	1.56	1.61	1.48	1.55	1.56	1.55	1.42
Average B factor (Å^2)								
Overall	11.46	19.44	28.92	25.11	20.37	24.85	25.65	32.62
Macromolecules	11.46	17.49	28.46	23.83	18.94	23.85	24.75	31.77
Ligands	19.85	55.88	33.31	32.87	29.09	31.17	31.28	30.88
Water	27.27	34.22	38.32	36.51	32.40	36.62	37.26	43.92
Clashscore	4.85	3.79	3.43	4.37	8.61	4.28	4.69	2.25
Ramachandran plot								
Favoured (%)	93.6	95.4	95.1	93.4	95.8	95.8	95.7	96.1
Allowed (%)	5.6	3.8	4.9	6.6	4.2	4.2	4.3	3.9
Outliers (%)	0.8	0.8	0	0	0	0	0	0
PDB code	8f05	8f06	8ezo	8f0b	8ezp	8ezu	8ezx	8f01

was of exceptional quality, with Isa > 50, and could be phased with a barely complete data set (multiplicity of ∼5.3) composed of only 80° of data; the whole 5760° range, collected over 230.4 s (multiplicity of 376), could also be experimentally phased. Additional crystals measured at the same temperature could be phased using a redundant data set, supporting the robustness of this method. Measurable decay was observed towards the end of each data collection. Since the goal of the experiment was phasing, which depended on the collection speed, high data quality and the anomalous diffraction strength, the empirically derived optimal data-collection parameters were kept constant for the three test proteins.

Data collection at 12 keV, 293 K and using the closest accessible detector distance allowed atomic resolution structures of the three proteins to be obtained: lysozyme at 1.20 Å resolution, thaumatin at 1.40 Å and proteinase K at 1.05 Å. Data were collected at 293 K except for proteinase K, for which data were collected at 273 K.

2.4. Data processing

All data were processed with *XDS*, *AIMLESS* and *TRUNCATE* as implemented in the *autoXDS* script developed at SSRL (Ana Gonzalez; https://smb.slac.stanford.edu/facilities/software/xds/#autoxds_script). Each 120° of data were processed with the `-friedel false` anomalous option separately in a stepwise function, and the subsequent steps were added to the previous steps in a cumulative way.

Table 2
Data collection, processing and refinement at 12 keV.

Protein	Lysozyme	Proteinase K	Thaumatin
Temperature (K)	293	273	293
Data collection			
Crystal dimensions (mm)	0.35, 0.32, 0.3	0.5, 0.4, 0.4	0.45, 0.4, 0.4
Wavelength (Å)	1.0332	1.0332	1.0322
Resolution range (Å)	35.06–1.20 (1.22–1.20)	35.34–1.05 (1.07–1.05)	38.32–1.39 (1.41–1.39)
Rotation range (°)	720	360	720
ADWD dose (kGy)	44	81	43
MaxDose (kGy)	161	286	155
Collection time (s)	36	72	36
MaxDose rate (kGy s^{-1})	4.5	4.0	4.3
Space group	$P4_32_12$	$P4_32_12$	$P4_12_12$
a, b, c (Å)	77.79, 77.79, 37.13	68.38, 68.38, 103.56	58.93, 58.93, 151.30
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Total reflections	1847186 (82796)	2570164 (36042)	2866372 (13690)
Multiplicity	53.9 (52.8)	22.7 (7.6)	52.4 (52.3)
Mosaicity (°)	0.16	0.09	0.08
Completeness (%)	95.0 (90.1)	98.9 (84.8)	99.9 (98.7)
Mean $I/\sigma(I)$	27.9 (1.6)	34.5 (4.1)	29.7 (1.3)
Wilson B factor (Å^2)	22.2	11.8	27.4
R_{merge}	0.092 (5.750)	0.049 (0.398)	0.093 (5.181)
$R_{p.i.m.}$	0.013 (0.790)	0.014 (0.222)	0.013 (0.713)
$CC_{1/2}$	1.000 (0.622)	1.000 (0.920)	1.000 (0.686)
ISa	22.7	29.1	26.2
Refinement			
Resolution range (Å)	33.53–1.20 (1.23–1.20)	35.37–1.05 (1.08–1.05)	38.32–1.39 (1.43–1.39)
R_{work} (%)	11.4 (25.8)	8.7 (13.0)	11.1 (30.6)
R_{free} (%)	14.4 (29.4)	9.7 (12.7)	13.7 (34.2)
No. of non-H atoms			
Total	1207	2512	1784
Macromolecules	1105	2154	1616
Ligands	15	46	11
Water	90	312	157
R.m.s. deviations			
Bond lengths (Å)	0.011	0.011	0.011
Angles (°)	1.86	1.66	1.62
Average B factor (Å^2)			
Overall	23.38	12.35	26.23
Macromolecules	22.31	9.48	24.07
Ligands	30.13	31.71	23.78
Water	35.40	29.32	48.58
Clashscore	12.49	3.01	4.04
Ramachandran plot			
Favoured (%)	96.4	94.8	96.4
Allowed (%)	3.6	4.4	3.6
Outliers (%)	0	0.8	0
PDB code	8f00	8f07	8f03

Cumulative steps contained 120°, 240°, 360°, 480°, ..., 1440° data sets, while the stepwise data sets consisted of 1–120°, 121–240°, 241–360°, ..., 1321–1440° of data. Such a dual treatment allows the contribution of each interval of data to be monitored while observing the inevitable decay of each subset. This approach can also help to reveal crystal inhomogeneity in (rotation) space, while still allowing comparison of spatially related entries: 1–120° and 361–480°, for example.

To access the effects of temperature on crystal decay, equivalent ranges for the same crystals were evaluated with increasing steps of 360° per specific temperature. The observed time at which the intensity drops below 70% from the initial intensity was identified for each temperature series.

All three proteins diffracted beyond 1.7 Å resolution with excellent statistics (see Table 1 for the 7.1 keV data sets and Table 2 for the 12 keV data sets), but the data were truncated to 1.8–1.9 Å resolution when the completeness of the outer shell was >80% for the 7.1 keV data sets. None of the detrimental effects usually observed with cryogenic cooling were present; the mosaicities of the thaumatin and proteinase K crystals were below 0.09° and that for lysozyme was 0.23°.

2.5. Phasing and model building

The anomalous data at low energy (7.1 keV) were used as input for the phasing and model-building software *SHELXC/D/E* (Sheldrick, 2010) as implemented in *HKL2MAP* (Pape & Schneider, 2004). Data sets with $CC_{1/2,building}$ > 25% were considered to be phased successfully. The output polyalanine model was entered into *ARP/wARP* (Langer *et al.*, 2008) for complete model building, resulting in 95% of the model being built to match the correct sequence. Manual rebuilding and minor adjustments were used to complete each model. These final models were refined with *REFMAC* (Murshudov *et al.*, 2011) to 1.8–1.9 Å resolution. Initially, the complete 1440° data sets were used to experimentally solve each of the three test proteins at 293 K, but further analysis showed that fewer data were needed. Thaumatin and proteinase K phasing was achieved with only 120° of data at 293 K, while lysozyme needed only 80°. For the proteinase K data measured at 310 K the observed diffraction decay was faster because of the higher temperature and the minimum data set that was needed to find the sites and for phasing was 240°.

Next, we obtained anomalous data from lysozyme crystals by varying the temperature down to 220 K. Each data collection led to successful structure solution, while allowing longer crystal lifetimes.

The phase problem was also solved at higher energy (12 keV) for lysozyme and thaumatin at 293 K and for proteinase K at 273 K. At this X-ray energy the native anomalous signals are weaker (Henke *et al.*, 1993), but obtaining high-quality and high-redundancy data sets favoured detection of the anomalous sites and resulted in an improved resolution that helped in automated model building and ensemble generation.

2.6. Anomalous signal estimation

The final models were used for phase generation and for calculating the positions and peak heights of anomalous sites with *ANODE* (Thorn & Sheldrick, 2011). These peak heights were treated as an objective representation of the anomalous signal strength. Since the peak heights are qualitative (redundancy/multiplicity-dependent) values, some of the peaks with low height were confirmed by inspecting the corresponding highest redundant data set from the same crystal. Identified ions had lower *ANODE* values, indicating more flexibility in the solvent area compared with the protein matrix.

2.7. Absorbed dose calculations

The average diffraction-weighted dose (ADWD) and the maximum dose (MaxDose) calculated with *RADDOSE-3D* (Bury *et al.*, 2018) at 293 K and 7.1 keV were 6 and 25 kGy, respectively, per 80° of data for lysozyme, 7 and 30 kGy per 120°, respectively, for thaumatin and 7 and 26 kGy per 120°, respectively, for proteinase K; the values for 240° of data for proteinase K at 310 K were 13 and 55 kGy, respectively, at 7.1 keV energy. The absorbed ADWDs and MaxDoses at 7.1 keV for optimized anomalous signal lysozyme data sets ($I/I_0 \geq 0.7$) at lower temperatures were 668 and 1657 kGy at 220 K (2880°), 182 and 741 kGy at 240 K (3240°), 115 and 516 kGy at 273 K (1800°) and 94 and 440 kGy at 293 K (1560°), respectively. ADWD and MaxDose for the 12 keV data sets were 44 and 161 kGy for lysozyme (720°), 43 and 155 kGy for thaumatin (720°) and 81 and 286 kGy for proteinase K (360°), respectively.

2.8. Anomalous Bijvoet ratio calculations

The Bijvoet ratios at 7.1 keV X-ray energy for the well ordered sites, using *ANODE* values of >8 (Mou *et al.*, 2022), for lysozyme (10 S, 1 Cl^-), thaumatin (17 S, 1 K^+) and proteinase K (11 S, 1 Ca^{2+}) were calculated to be 1.49%, 1.63% and 1.21%, respectively, while these ratios decrease to 0.54%, 0.60% and 0.44% at 12 keV (Hendrickson & Teeter, 1981; Dauter *et al.*, 2002; Wang *et al.*, 2006).

3. Results

3.1. Obtaining high-quality anomalous signal across physiological temperatures from single protein crystals using 7.1 keV X-rays

To determine whether quality anomalous signal can be collected at and above RT, we collected diffraction data at 293 K and 310 K from single thaumatin, proteinase K and lysozyme crystals using a recent approach for single-crystal X-ray diffraction data collection (Doukov *et al.*, 2020). We observed a strong anomalous signal at room temperature (293 K) for thaumatin, proteinase K and lysozyme crystals at 7.1 keV. These data allowed us to experimentally phase the protein structures based on their naturally occurring sulfur-containing amino acids and bound ions (see below).

We used an energy of 7.1 keV because our experience on BL14-1 at SSRL has shown that this energy represents an optimal compromise between achieving relatively high resolution (1.8–1.9 Å) and relatively strong anomalous signal, while decreasing the X-ray absorption at lower X-ray energies (at 6 keV). The measured data were of excellent quality (Table 1). The resolution was limited by the data-collection setup, but no significant dehydration was observed when V_{cell} from the 120° data set is compared with that from the 360 + 120° data set, results that are in line with those observed previously (Doukov *et al.*, 2020).

The highest quality room-temperature data were obtained using relatively large crystals (*i.e.* dimensions larger than 200–250 µm). This is because at RT crystals are 50–100 times more sensitive to X-ray damage (Southworth-Davies *et al.*, 2007; Warkentin *et al.*, 2011). Therefore, achieving resolutions similar to those obtained at cryogenic temperatures requires a larger number of unit cells and therefore larger crystals. Obtaining higher resolution data from smaller crystals would entail using higher X-ray doses that lead to diffraction decaying before a complete data set with sufficient anomalous signal multiplicity can be recorded. Experimental native phasing has traditionally been performed using high-redundancy data sets at cryogenic temperatures (Terwilliger *et al.*, 2016). Analysis of the data sets collected indicates that data with relatively low redundancy (*i.e.* 80–120° for the tetragonal space groups of the crystals used in this work) collected at RT yield anomalous signal of sufficiently high quality to allow SAD phasing. Fig. 1 shows the anomalous atom sites found by *SHELXD* for lysozyme (Fig. 1*a*), thaumatin (Fig. 1*b*) and proteinase K (Fig. 1*c*) and the electron density of the experimentally phased thaumatin protein structure (Fig. 1*d*).

3.2. Obtaining anomalous signal from single-crystal X-ray diffraction data at 7.1 keV and in the 220–293 K temperature range

For many proteins it will not be possible to obtain the relatively large crystals that are required to collect high-resolution data at and above RT. However, because the tolerance of crystals to X-ray damage increases with decreasing temperature (Warkentin & Thorne, 2010; Warkentin *et al.*, 2011), it should still be possible to obtain useful anomalous data for smaller crystals and at lower temperatures. The proposed method allows the collection of data at such lower temperatures and allows anomalous signal to be obtained at temperatures below RT but above the so-called average glass transition temperature (180–220 K) where motions similar to those in solution remain (Keedy, Fraser *et al.*, 2015; Lewandowski *et al.*, 2015; Tilton *et al.*, 1992; Vitkup *et al.*, 2000; Ringe & Petsko, 2003).

Crystal lifetimes, measured as 'the time until the I/I_0 ratio decreases below 0.7', increase with cooling. Based on work on X-ray damage at cryotemperatures (Owen *et al.*, 2006) and at RT (Yabukarski, Doukov, Mokhtari *et al.*, 2022; de la Mora *et al.*, 2020), this value of I/I_0 was proposed as a limit for intensity decay due to X-ray damage. For lysozyme at 293 K it took 62.4 s (1560°), at 273 K it took 72 s (1800°) and at 220 K it took 288 s (2880°). The measured anomalous multiplicity was 58 at 293 K, 65 at 273 K and 161 at 220 K, notably larger than the value of 2.9 for the 80° data sets at 273 K and 293 K. The 220 K data set could be phased with 240° of data and an anomalous multiplicity of 8.5. Furthermore, processing the data beyond the $I/I_0 < 0.7$ limit did not dramatically decrease the ability to phase the structure. At 220 K, data up to 5760° with an anomalous multiplicity of 185 could be directly phased despite the accumulated X-ray damage.

Collecting data from lysozyme crystals cooled to 220 K increased the number of data frames (relative to the 293 K data set) by a factor of ∼1.8 as measured by the time taken for the intensity to decay below the 0.7 I/I_0 limit. Since the 220 K

data were collected 2.5-fold more slowly (at 0.1 s versus 0.04 s per frame), the potential crystal lifetime could be extended 4.6-fold compared with data collection at 293 K. In addition, the data collected beyond the proposed I/I_0 limit of 0.7 (at 288 s) do not prevent successful phasing with the whole 16 × 360 data sets (5760° in 576 s or almost 10 min). Such a long lifetime would increase the data multiplicity up to 370-fold [with a potential increase of up to 926-fold by data collection at 25 Hz (0.04 s per 1°) instead of 10 Hz] and thus tremendously increase the chance of successful phasing. In addition, data collection across temperature might allow potential changes in the distribution of protein conformational states to be detected (Keedy, Kenner *et al.*, 2015; Keedy, 2019). The complete lysozyme data sets were easily experimentally phased, but success varied for the low-redundancy data sets: 80° of data at 273 and 293 K and 240° of data at 220 K were needed to phase the structure, suggesting an important role of the individual crystal quality. Fig. 2 shows that the ADWD at 220 K is higher than that expected from the linear ADWD trend observed in the 240–293 K temperature range, suggesting that the glass transition for lysozyme might be occurring between 220 and 240 K. The glass transition temperature reflects the intrinsic temperature dependence of the motions of protein atoms, the bound solvent or a combination of both. To directly determine the exact transition temperature, future work will determine and compare conformational heterogeneity across temperature.

To determine whether there are any changes in the distribution of anomalous signal with temperature, we analysed the lysozyme data sets collected across temperature. Comparison of the *REFMAC*-derived $2F_o - F_c$ maps indicated a novel sodium site near active-site residue Glu35 of lysozyme in the 220–240 K temperature range, but not at 260 K and above (Fig. 3).

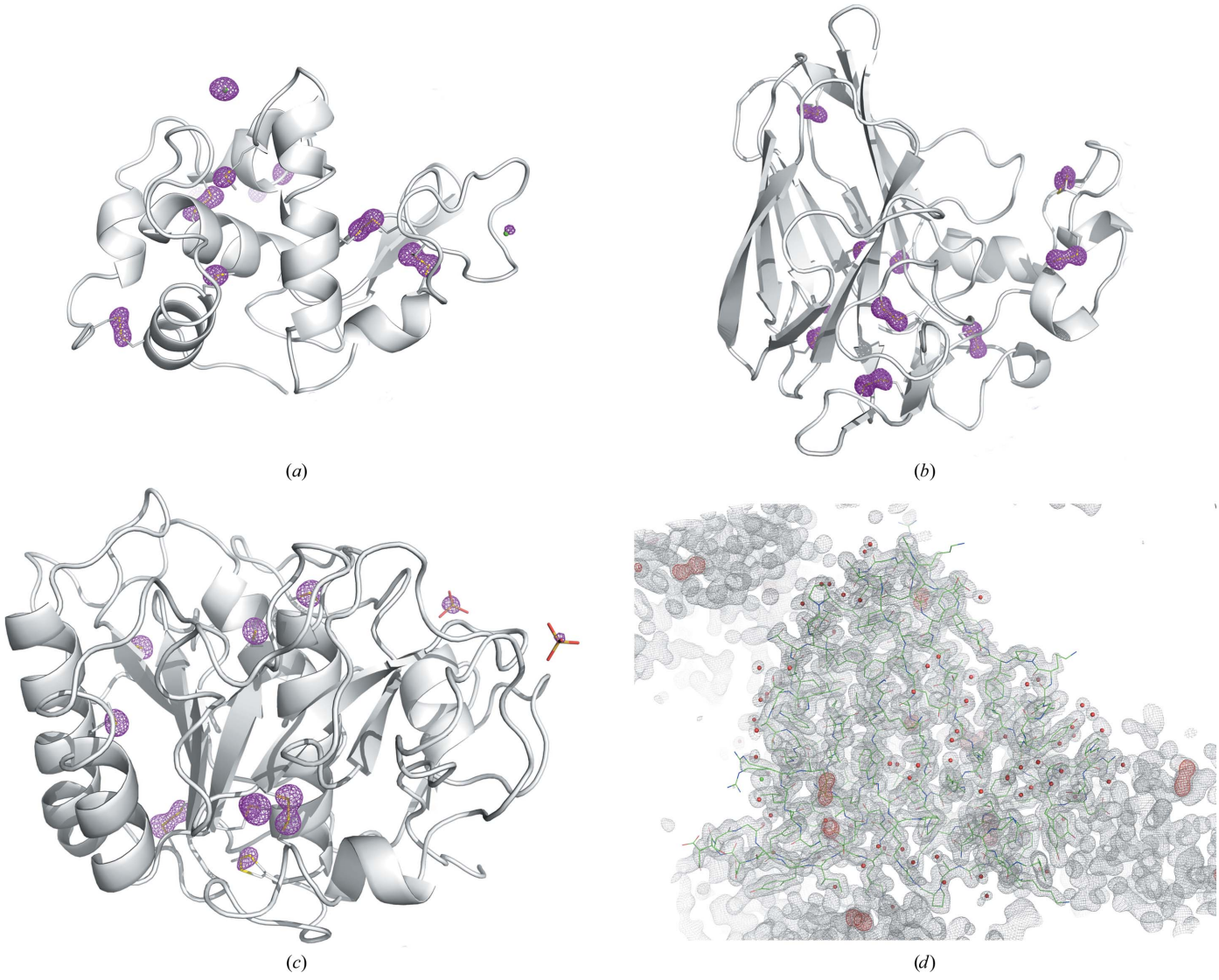


Figure 1
Anomalous sites found by *SHELXD* for lysozyme (*a*), thaumatin (*b*) and proteinase K (*c*) and the electron density of the experimentally phased thaumatin protein structure (*d*); the *ANODE*-calculated anomalous map contoured at the 4σ level is shown as a magenta mesh, while the $2F_o - F_c$ electron-density map of the final model contoured at the 1σ level is shown as a grey mesh. Images were created with *PyMOL* (DeLano, 2002).

3.3. Analysis of the anomalous signal at room temperature (293 K): the role of redundancy and effect of X-ray damage

To determine the minimal and optimal amount of data required to obtain a meaningful anomalous signal and the effects of X-ray damage, we processed diffraction data until the diffraction intensity decreased to about 0.7 of its initial value (I/I_0) and systematically analysed the strength of the anomalous signal. We analysed the diffraction data using two different approaches: stepwise and cumulative (Fig. 4). Results from a lysozyme crystal measured at 293 K and 7.1 keV with each 120° of data processed with the '`-friedel false`' anomalous option separately (stepwise data-set collection) and in a cumulative manner (in which the subsequent steps were added to the previous steps, resulting in a data set with increased multiplicity) are shown in Fig. 4 for each type of anomalous atom: the sulfur of cysteine (Fig. 4*a*, SG), the sulfur of methionine (Fig. 4*b*, SD) and a localized chloride (Fig. 4*c*, Cl). Anomalous peaks calculated with *ANODE* (Thorn & Sheldrick, 2011; *y* axis) are plotted versus 120° data-set steps. Seven of the eight observed chloride sites were partially occupied and were found in more flexible solvent regions, resulting in an average threefold lower anomalous peaks compared with the sulfur peaks. The cumulative signal at the anomalous positions increases up to 360–480° and then plateaus. Each 120° anomalous signal step is lower than the preceding step. Each of the 120° lysozyme mini data sets could be individually phased up to 1440° (12 steps), despite the anomalous signal decreasing at lower resolution. Cumulative data sets also allowed phasing up to the end of data collection: all 1440° of data could also be phased. Higher redundancy instead of low-redundancy data sets were beneficial in identifying ions due to their increased mobility and weaker signal. The X-ray damage and decay towards the end of data collection was compensated by the scaling procedure and the increased multiplicity.

Our results indicate that X-ray damage systematically decreases the amount of anomalous signal (stepwise data sets). However, in the initial stages of data collection, accumulating increasingly damaged data allows the anomalous signal to increase to a certain extent while further data collection does not lead to any further increase (cumulative data sets).

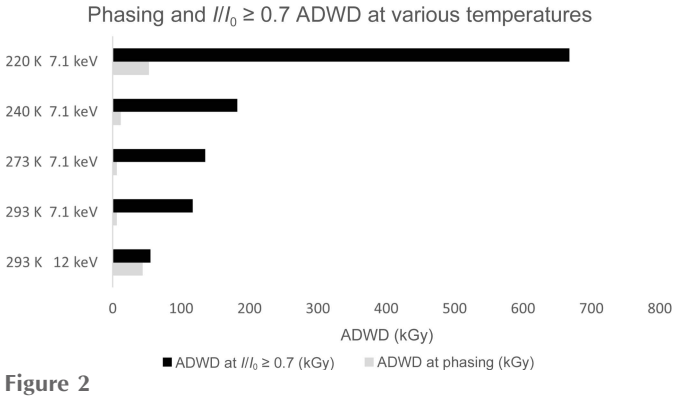


Figure 2
The benefits of lowering the data-collection temperature for collecting higher multiplicity phasing data expressed as ADWD doses.

3.4. Anomalous signal allows water molecules and ions to be distinguished in lysozyme, thaumatin and proteinase K data sets across physiological temperatures

In addition to experimentally phasing and solving a new structure, single-crystal data collection is expected to help distinguish between water molecules and ions, allowing more complete modelling of the protein solvent system.

Examination of the electron-density maps for distances between ions, surrounding protein residues and waters and anomalous signal peaks in our RT data unambiguously

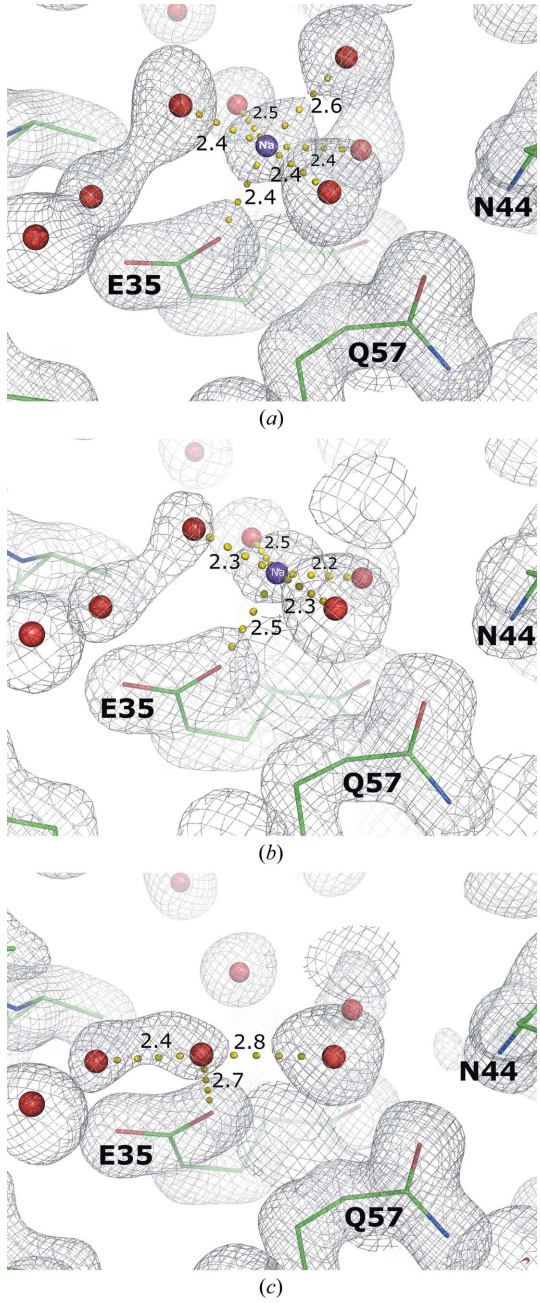


Figure 3
Temperature-dependent octahedral sodium binding site in lysozyme (Zheng *et al.*, 2014; Harding, 2002). A well ordered sodium binding site with $Na^{+}\cdots O$ distances close to the expected 2.4 Å is seen at 220 K (*a*), is less ordered at 240 K (*b*) and is missing at 260 K (*c*) and higher temperatures in the $2F_o - F_c$ maps (contoured at 1σ as grey mesh).

identified a number of ions: eight chlorides and one sodium in lysozyme, one potassium in thaumatin and four sulfate ions in proteinase K (Zheng *et al.*, 2014; Harding, 2002).

The anomalous data allowed us to identify a novel potassium site from the thaumatin data set at 293 K based on an anomalous peak inside $2F_o - F_c$ density (Fig. 5*a*). Fig. 5(*b*) shows that a clear anomalous peak for a bound sulfate ion could be identified even in the proteinase K data set at 310 K. Without a carefully planned experiment and at different wavelengths such ions could routinely be modelled inaccurately as waters.

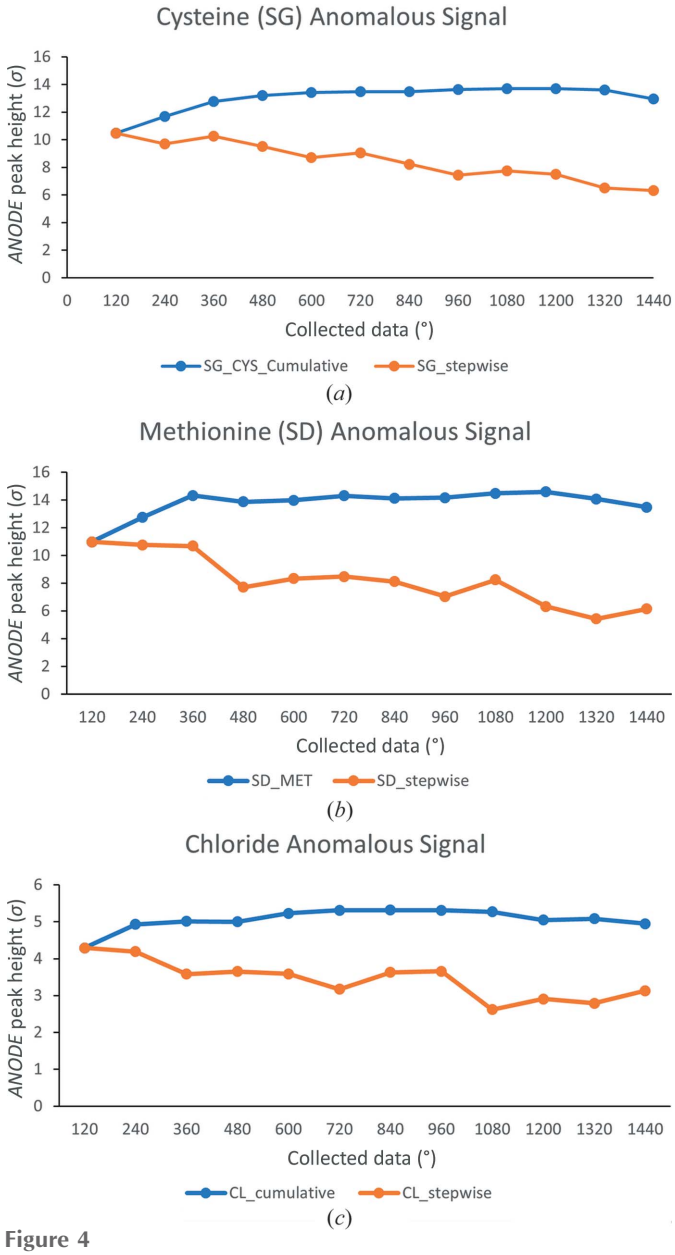


Figure 4
Analysis of 120° stepwise (orange) and cumulative (blue) (up to 1440° on the *x* axis) anomalous signals as measured in *ANODE* (*y* axis) for the lysozyme data set at 293 K: cysteine SG atoms (*a*), methionine SD atoms (*b*) and chloride ions (*c*). The cysteine sulfur anomalous signal decays faster, presumably due to its higher susceptibility to X-ray damage and subsequent mobility. Chloride signals are weaker due to their higher mobilities.

3.5. Single-crystal X-ray data collection and phasing at 12 keV and at 273 K and 293 K

The above data sets were collected at 7.1 keV, an energy traditionally used for anomalous data collection for iron and lighter elements in macromolecular X-ray crystallography (Hendrickson, 2014). While lower energies (and thus longer wavelengths) provide a stronger sulfur anomalous signal, in practice longer wavelengths also decrease the attainable dataset resolution. Thus, the collection of data with strong anomalous signal occurs at the expense of resolution for better diffracting crystals.

To overcome the above limitation, we asked whether an anomalous signal of strength sufficient for SAD phasing and ion identification could be obtained by collecting diffraction data at shorter wavelengths, which are typically used at synchrotron beamlines for routine high-resolution data collection. We collected 12 keV X-ray diffraction data from lysozyme and thaumatin crystals at 293 K and a proteinase K crystal at 273 K (Table 2) and determined the structures via

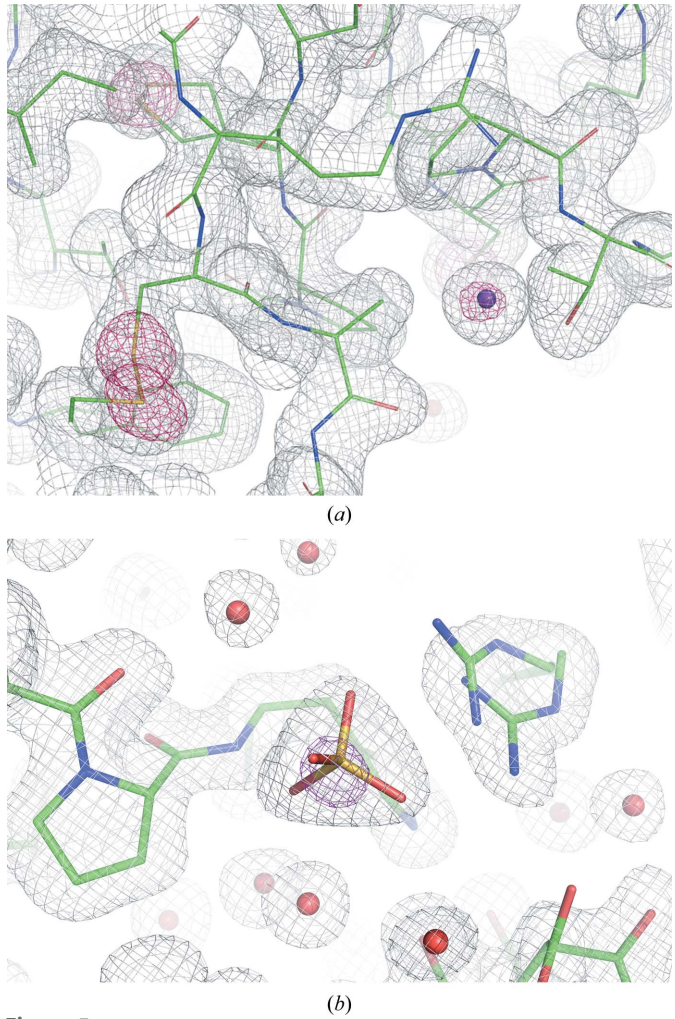


Figure 5
Specific anomalous binding sites from 7.1 keV data: potassium ion (violet) in thaumatin at 293 K (*a*) and sulfate ion in proteinase K at 310 K (*b*). The *ANODE*-calculated anomalous map at the 4σ level is shown as a magenta mesh, while the $2F_o - F_c$ electron-density map of the final model contoured at the 1σ level is shown as a grey mesh.

SAD. Analysis of the electron density and anomalous signal revealed the following. For proteinase K, 26 alternative side-chain rotamers and nine sulfate ions were found in the 273 K 12 keV data set at 1.05 Å resolution. Fig. 6(*a*) shows the electron density after phasing and modelling, while Fig. 6(*b*) shows the quality of the phases from the anomalous substructure alone.

Thus, using high-quality crystals a high-resolution data set can be collected at RT and can be used to solve the structure, to obtain conformational ensemble information (van den Bedem *et al.*, 2009; not explicitly modelled in this work) and distinguish between water and ions. It should be possible to collect data sets using X-rays of 12 keV energy at temperatures down to 220 K, as was done above for lysozyme using an X-ray energy of 7.1 keV. Such data would allow information to be obtained about solvation and ion binding by using X-ray data to obtain the temperature dependence of binding position, affinity and specificity.

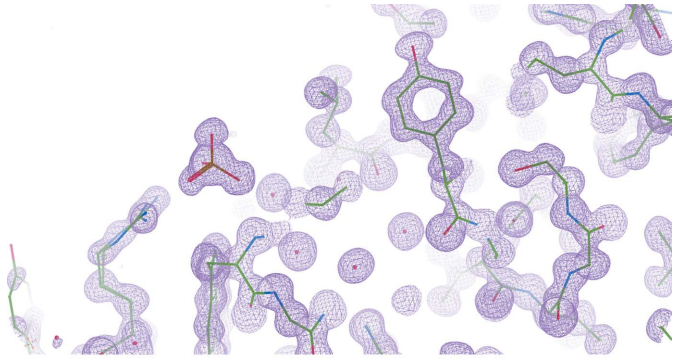
4. Discussion and conclusions

Here, using a recently developed approach for data collection from single protein crystals (Doukov *et al.*, 2020), we collected data sets at and above RT and showed that high-quality anomalous signal can be obtained from single-crystal diffraction data with relatively low data redundancy. The data can be collected using X-rays at 12 keV, an energy that is widely used for routine data collection at synchrotrons and which allows higher resolution data collection compared with lower energy X-rays for a given experimental setup. The data can be used for native SAD phasing, and both experimental phasing and structural and ensemble information can be obtained in a single experiment, thus streamlining the entire process.

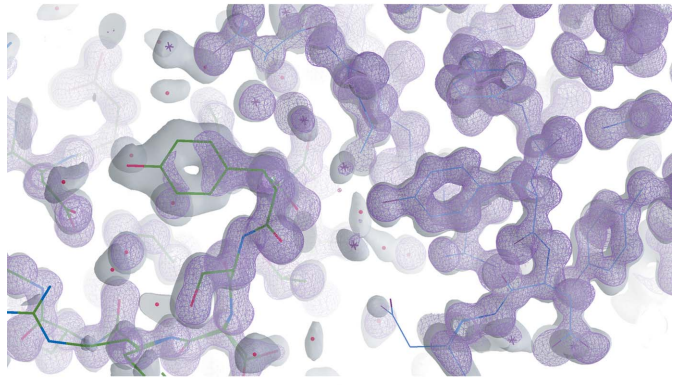
Importantly, obtaining anomalous signal at and above RT allows ions to be identified and water molecules and ions to be differentiated without any cryocooling-associated artefacts. Ions are often missed in structural models as the $2F_o - F_c$ electron-density maps do not provide clear differentiation between partially occupied ions and fully occupied water molecules. Diffraction data at or above the X-ray absorption edge of the element provide anomalous signal. This signal may be normalized via refinement versus the expected anomalous signal for a fully occupied and ordered ion at this energy and refined anomalous and atomic occupancies can be obtained. For well ordered sites and amino acids the occupancies are unity, while lower occupancies are expected for ions bound to the solvent-exposed site. More accurate modelling of solvation ions allows a more complete model of protein systems to be obtained by combining protein conformational ensemble information with more accurate solvation information.

Additional chloride binding sites were observed with increasing ion concentration during crystallization. For example, up to 17 chloride sites were located in lysozyme grown in 2 *M* NaCl (PDB entry 6g8a; Leonarski *et al.*, 2018) compared with the eight Cl atoms observed in crystals grown at a 0.6 *M* NaCl concentration here. All eight sites occupied at the lower NaCl concentration are also occupied at 2 *M* NaCl, while the other nine sites only appear at 2 *M*. However, because the structure at higher NaCl concentration (PDB entry 6g8a) was obtained at cryogenic temperature, it is also possible that a fraction or all of the additional binding sites appear as a result of cryocooling. Obtaining high-quality anomalous signal will enable future work to decipher the relationship between salt concentration-dependent protein stability effects and ion binding, effectively relating structural information to protein energetics. For proteins crystallizing under broad salt concentrations such a propensity to bind ions at specific sites could be explored with elements with stronger anomalous signal (bromide and iodide ions instead of chloride and potassium, or rubidium or caesium instead of sodium ions) and competition experiments could be used to obtain further insights into the energetics of protein systems. Learning about local electrostatic environments will also allow the force fields in molecular dynamics (MD) simulations to be improved.

For proteins that do not form the relatively large crystals required for high-resolution data collection at and above RT, anomalous signal and meaningful conformational ensemble information can be obtained at lower temperatures that are still above the average protein glass-transition temperature. For challenging cases such as multi-domain proteins, low-



(*a*)



(*b*)

Figure 6
High-quality SAD phasing maps at 12 keV shown for proteinase K at 273 K. (*a*) The quality of the experimental phases around Y82 and a sulfate ion. (*b*) The map from the anomalous atom substructure calculated using *SHELXE* is shown as a violet mesh, while the final $2F_o - F_c$ map calculated in *REFMAC* is presented as a grey isosurface in the vicinity of the surface-exposed Y61. Both maps are contoured at the 1σ level.

symmetry crystal forms, highly radiation-sensitive proteins and proteins that do not cryocool well, the use of temperatures lower than RT and low energy (7.1–6.0 keV) ensure lower X-ray damage rates, higher multiplicity and stronger anomalous signal from native scatterers, all of which result in an improved probability of successful phasing. In addition to improved crystal lifetimes, these structures would contain novel temperature-dependent ensemble information. For well diffracting crystals and well ordered multiple native scatterers a higher energy, such as the 12 keV used here, could be applied. The anomalous signal should be used as an additional constraint in generating protein ensembles using the residual electron and anomalous density (READ) approach (Salmon *et al.*, 2018). Temperature-dependent solvent and ion ligation observations could be used to improve force fields for MD simulations.

Single-crystal data collection on a standard synchrotron beamline has inherent advantages in simplicity, time, materials and equipment utilization over serial synchrotron X-ray crystallography (SSX) approaches and serial X-ray crystallography at the very few and currently oversubscribed X-ray free-electron laser (XFEL) facilities. The role of *de novo* phasing has decreased in the era of artificial intelligence (AI) protein-modelling software such as *AlphaFold* (Jumper *et al.*, 2021), which often provides models for molecular replacement. However, increasing the accuracy of protein conformation ensemble models by incorporating anomalous signal and providing a more accurate description of protein solvation by identifying ions and by discriminating between ions and water molecules will be of significant value for future studies aimed at obtaining conformational ensemble and solvation information at RT or PT.

Acknowledgements

Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health (NIH), National Institute of General Medical Sciences (NIGMS) (P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

Funding information

This work was funded by a National Science Foundation (NSF) Grant (MCB-1714723) to DH. FY was supported in part by a long-term Human Frontiers Science Program post-doctoral fellowship.

References

Alber, T., Gilbert, W. A., Ponzi, D. R. & Petsko, G. A. (1983). *Ciba Found. Symp.* **93**, 4–24.
Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. (1975). *Biochemistry*, **14**, 5355–5373.
Bedem, H. van den, Dhanik, A., Latombe, J.-C. & Deacon, A. M. (2009). *Acta Cryst.* D**65**, 1107–1117.
Berg, J. M., Gatto, G. J., Tymoczko, J. L. & Stryer, L. (2019). *Biochemistry*. New York: Macmillan Learning.
Botha, S., Baitan, D., Jungnickel, K. E. J., Oberthür, D., Schmidt, C., Stern, S., Wiedorn, M. O., Perbandt, M., Chapman, H. N. & Betzel, C. (2018). *IUCrJ*, **5**, 524–530.
Burnley, B. T., Afonine, P. V., Adams, P. D. & Gros, P. (2012). *eLife*, **1**, e00311.
Bury, C. S., Brooks-Bartlett, J. C., Walsh, S. P. & Garman, E. F. (2018). *Protein Sci.* **27**, 217–228.
Casanas, A., Warshamanage, R., Finke, A. D., Panepucci, E., Olieric, V., Nöll, A., Tampé, R., Brandstetter, S., Förster, A., Mueller, M., Schulze-Briese, C., Bunk, O. & Wang, M. (2016). *Acta Cryst.* D**72**, 1036–1048.
Cianci, M., Helliwell, J. R., Moorcroft, D., Olczak, A., Raftery, J. & Rizkallah, P. J. (2004). *J. Appl. Cryst.* **37**, 555–564.
Colletier, J.-P., Sawaya, M. R., Gingery, M., Rodriguez, J. A., Cascio, D., Brewster, A. S., Michels-Clark, T., Hice, R. H., Coquelle, N., Boutet, S., Williams, G. J., Messerschmidt, M., DePonte, D. P., Sierra, R. G., Laksmono, H., Koglin, J. E., Hunter, M. S., Park, H.-W., Uervirojnangkoorn, M., Bideshi, D. K., Brunger, A. T., Federici, B. A., Sauter, N. K. & Eisenberg, D. S. (2016). *Nature*, **539**, 43–47.
Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.
DeLano, W. L. (2002). *CCP4 Newsl. Protein Crystallogr.* **40**, 82–92.
Doukov, T., Herschlag, D. & Yabukarski, F. (2020). *J. Appl. Cryst.* **53**, 1493–1501.
Fersht, A. (1985). *Enzyme Structure and Mechanism*, 2nd ed. New York: W. H. Freeman.
Foadi, J., Aller, P., Alguel, Y., Cameron, A., Axford, D., Owen, R. L., Armour, W., Waterman, D. G., Iwata, S. & Evans, G. (2013). *Acta Cryst.* D**69**, 1617–1632.
Foos, N., Seuring, C., Schubert, R., Burkhardt, A., Svensson, O., Meents, A., Chapman, H. N. & Nanao, M. H. (2018). *Acta Cryst.* D**74**, 366–378.
Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N. & Alber, T. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 16247–16252.
Frauenfelder, H., Parak, F. & Young, R. D. (1988). *Annu. Rev. Biophys. Biophys. Chem.* **17**, 451–479.
Frauenfelder, H. & Petsko, G. A. (1980). *Biophys. J.* **32**, 465–483.
Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. (1979). *Nature*, **280**, 558–563.
Frauenfelder, H., Sligar, S. & Wolynes, P. (1991). *Science*, **254**, 1598–1603.
Garman, E. (2003). *Curr. Opin. Struct. Biol.* **13**, 545–551.
Gildea, R. J., Beilsten-Edmands, J., Axford, D., Horrell, S., Aller, P., Sandy, J., Sanchez-Weatherby, J., Owen, C. D., Lukacik, P., Strain-Damerell, C., Owen, R. L., Walsh, M. A. & Winter, G. (2022). *Acta Cryst.* D**78**, 752–769.
Green, D. W., Ingram, V. M., Perutz, M. F. & Bragg, W. L. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287–307.
Greisman, J. B., Dalton, K. M., Sheehan, C. J., Klureza, M. A., Kurinov, I. & Hekstra, D. R. (2022). *Acta Cryst.* D**78**, 986–996.
Hammes, G. G., Benkovic, S. J. & Hammes-Schiffer, S. (2011). *Biochemistry*, **50**, 10422–10430.
Harding, M. M. (2002). *Acta Cryst.* D**58**, 872–874.
Hendrickson, W. A. (1985). *Trans. Am. Crystallogr. Assoc.* **21**, 11–21.
Hendrickson, W. A. (2014). *Q. Rev. Biophys.* **47**, 49–93.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature*, **290**, 107–113.
Henke, B. L., Gullikson, E. M. & Davis, J. C. (1993). *At. Data Nucl. Data Tables*, **54**, 181–342.

Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hübner, C. G. & Kern, D. (2007). *Nature*, **450**, 838–844.
Hirata, K., Yamashita, K., Ueno, G., Kawano, Y., Hasegawa, K., Kumasaka, T. & Yamamoto, M. (2019). *Acta Cryst.* D**75**, 138–150.
Hope, H. (1988). *Acta Cryst.* B**44**, 22–26.
Horowitz, S., Salmon, L., Koldewey, P., Ahlstrom, L. S., Martin, R., Quan, S., Afonine, P. V., van den Bedem, H., Wang, L., Xu, Q., Trievel, R. C., Brooks, C. L. & Bardwell, J. C. A. (2016). *Nat. Struct. Mol. Biol.* **23**, 691–697.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). *Nature*, **596**, 583–589.
Kahn, R., Fourme, R., Bosshard, R., Chiadmi, M., Risler, J. L., Dideberg, O. & Wery, J. P. (1985). *FEBS Lett.* **179**, 133–137.
Keedy, D. A. (2019). *Acta Cryst.* D**75**, 123–137.
Keedy, D. A., Fraser, J. S. & van den Bedem, H. (2015). *PLoS Comput. Biol.* **11**, e1004507.
Keedy, D. A., Kenner, L. R., Warkentin, M., Woldeyes, R. A., Hopkins, J. B., Thompson, M. C., Brewster, A. S., Van Benschoten, A. H., Baxter, E. L., Uervirojnangkoorn, M., McPhillips, S. E., Song, J., Alonso-Mori, R., Holton, J. M., Weis, W. I., Brunger, A. T., Soltis, S. M., Lemke, H., Gonzalez, A., Sauter, N. K., Cohen, A. E., van den Bedem, H., Thorne, R. E. & Fraser, J. S. (2015). *eLife*, **4**, e07574.
Keedy, D. A., van den Bedem, H., Sivak, D. A., Petsko, G. A., Ringe, D., Wilson, M. A. & Fraser, J. S. (2014). *Structure*, **22**, 899–910.
Klinman, J. P. (2015). *Acc. Chem. Res.* **48**, 449–456.
Lang, P. T., Ng, H.-L., Fraser, J. S., Corn, J. E., Echols, N., Sales, M., Holton, J. M. & Alber, T. (2010). *Protein Sci.* **19**, 1420–1431.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nat. Protoc.* **3**, 1171–1179.
Leonarski, F., Redford, S., Mozzanica, A., Lopez-Cuenca, C., Panepucci, E., Nass, K., Ozerov, D., Vera, L., Olieric, V., Buntschu, D., Schneider, R., Tinti, G., Froejdh, E., Diederichs, K., Bunk, O., Schmitt, B. & Wang, M. (2018). *Nat. Methods*, **15**, 799–804.
Lewandowski, J. R., Halse, M. E., Blackledge, M. & Emsley, L. (2015). *Science*, **348**, 578–581.
Liu, Q. & Hendrickson, W. A. (2015). *Curr. Opin. Struct. Biol.* **34**, 99–107.
Mora, E. de la, Coquelle, N., Bury, C. S., Rosenthal, M., Holton, J. M., Carmichael, I., Garman, E. F., Burghammer, M., Colletier, J.-P. & Weik, M. (2020). *Proc. Natl Acad. Sci. USA*, **117**, 4142–4151.
Mou, T.-C., Zeng, B., Doukov, T. I. & Sprang, S. R. (2022). *Acta Cryst.* D**78**, 1021–1031.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nass, K., Cheng, R., Vera, L., Mozzanica, A., Redford, S., Ozerov, D., Basu, S., James, D., Knopp, G., Cirelli, C., Martiel, I., Casadei, C., Weinert, T., Nogly, P., Skopintsev, P., Usov, I., Leonarski, F., Geng, T., Rappas, M., Doré, A. S., Cooke, R., Nasrollahi Shirazi, S., Dworkowski, F., Sharpe, M., Olieric, N., Bacellar, C., Bohinc, R., Steinmetz, M. O., Schertler, G., Abela, R., Patthey, L., Schmitt, B., Hennig, M., Standfuss, J., Wang, M. & Milne, C. J. (2020). *IUCrJ*, **7**, 965–975.
Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4912–4917.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Ploscariu, N., Burnley, T., Gros, P. & Pearce, N. M. (2021). *Acta Cryst.* D**77**, 1357–1364.
Pozharski, E., Deller, M. C. & Rupp, B. (2017). *Methods Mol. Biol.* **1607**, 611–625.
Pozharski, E., Weichenberger, C. X. & Rupp, B. (2013). *Acta Cryst.* D**69**, 150–167.
Ravelli, R. B. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.
Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.
Riley, B. T., Wankowicz, S. A., de Oliveira, S. H. P., van Zundert, G. C. P., Hogan, D. W., Fraser, J. S., Keedy, D. A. & van den Bedem, H. (2021). *Protein Sci.* **30**, 270–285.
Ringe, D. & Petsko, G. A. (2003). *Biophys. Chem.* **105**, 667–680.
Rose, J. P. & Wang, B.-C. (2016). *Arch. Biochem. Biophys.* **602**, 80–94.
Salmon, L., Ahlstrom, L. S., Bardwell, J. C. A. & Horowitz, S. (2018). *Methods Mol. Biol.* **1764**, 491–504.
Schirò, G. & Weik, M. (2019). *J. Phys. Condens. Matter*, **31**, 463002.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Southworth-Davies, R. J., Medina, M. A., Carmichael, I. & Garman, E. F. (2007). *Structure*, **15**, 1531–1541.
Terwilliger, T. C., Bunkóczi, G., Hung, L.-W., Zwart, P. H., Smith, J. L., Akey, D. L. & Adams, P. D. (2016). *Acta Cryst.* D**72**, 346–358.
Thorn, A. & Sheldrick, G. M. (2011). *J. Appl. Cryst.* **44**, 1285–1287.
Tilton, R. F., Dewan, J. C. & Petsko, G. A. (1992). *Biochemistry*, **31**, 2469–2481.
Vitkup, D., Ringe, D., Petsko, G. A. & Karplus, M. (2000). *Nat. Struct. Biol.* **7**, 34–38.
Wang, J., Dauter, M. & Dauter, Z. (2006). *Acta Cryst.* D**62**, 1475–1483.
Warkentin, M., Badeau, R., Hopkins, J. & Thorne, R. E. (2011). *Acta Cryst.* D**67**, 792–803.
Warkentin, M. & Thorne, R. E. (2010). *Acta Cryst.* D**66**, 1092–1100.
Wei, G., Xi, W., Nussinov, R. & Ma, B. (2016). *Chem. Rev.* **116**, 6516–6551.
Yabukarski, F., Doukov, T., Mokhtari, D. A., Du, S. & Herschlag, D. (2022). *Acta Cryst.* D**78**, 945–963.
Yabukarski, F., Doukov, T., Pinney, M. M., Biel, J. T., Fraser, J. S. & Herschlag, D. (2022). *Sci. Adv.* **8**, eabn7738.
Zander, U., Bourenkov, G., Popov, A. N., de Sanctis, D., Svensson, O., McCarthy, A. A., Round, E., Gordeliy, V., Mueller-Dieckmann, C. & Leonard, G. A. (2015). *Acta Cryst.* D**71**, 2328–2343.
Zheng, H., Chordia, M. D., Cooper, D. R., Chruszcz, M., Müller, P., Sheldrick, G. M. & Minor, W. (2014). *Nat. Protoc.* **9**, 156–170.