STRUCTURAL BIOLOGY

Sensitive detection of structural dynamics using a statistical framework for comparative crystallography

Doeke R. Hekstra[1,2]*†, Harrison K. Wang[1,3]†, Margaret A. Klureza[1,4]‡, Jack B. Greisman[1]§, Kevin M. Dalton[1,5,6]*

Copyright © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

Chemical and conformational changes are crucial to protein function and its pharmacological control. X-ray crystallography can reveal these changes in atomic detail, but standard analysis methods, which refine separate datasets, often overlook differences that are subtle or arise in only a subset of molecules. Direct comparison of crystallographic datasets is, in principle, more powerful, but systematic errors ("scales") often mask changes in the crystallographic observables ("structure factors"). Machine learning algorithms that jointly estimate scales and structure factors can address this limitation. Here, we augment this approach with multivariate, structured priors derived from crystallographic theory, implemented in the variational deep learning framework Careless. Doing so strongly improves the detection of protein dynamics, element-specific anomalous signals, and the binding of drug candidates, offering a robust approach to comparative crystallography and, potentially, to detection of protein dynamics by other structure determination methods.

INTRODUCTION

Proteins and their assemblies are dynamic molecular machines that mediate chemical catalysis, molecular transport, signal transduction, and the allosteric control of these processes. To understand the function of these machines, one typically needs to monitor their functional and structural response to perturbations. We illustrate this paradigm with several examples in Fig. 1A. For example, foundational to our understanding of DNA polymerases, the motion of key residues and the substrate of DNA polymerase were revealed by soaking magnesium into crystals of a DNA polymerase (*1*–*3*) with observation by x-ray crystallography. Physical perturbations (e.g., light, temperature jump, and electric field) have also been used to activate the functional dynamics of proteins involved in signal transduction (*4*, *5*), oxygen transport (*6*–*8*), ion transport across membranes (*9*–*11*), and DNA repair (*12*) in crystallographic experiments visualizing key interactions and intermediates with high spatial and temporal resolution. Last, the discovery of drug precursors targeting key enzymes involved in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) replication (*13*–*15*) and pathological signal transduction (*16*) has been addressed by adding libraries of thousands of potential ligands to protein crystals. In addition to identifying medically relevant compounds, these ligand-screening efforts contribute insight into the fundamental biophysical mechanisms of proteins (*17*).

X-ray crystallography makes it possible to measure the effect of these perturbations with high spatial and temporal resolution, over a wide range of timescales, and with high throughput. Recent methodological advances have expanded this capability markedly, because of new x-ray sources such as fourth-generation synchrotrons (*18*) and x-ray free-electron lasers (XFELs) (*19*, *20*), new time-resolved methods such as mix-and-inject studies of enzyme catalysis (*21*–*25*), and high-throughput methods permitting screening of thousands of molecules' interactions with proteins of interest (*13*–*16*, *26*, *27*). High-resolution data collected from these methods can enable the tracking of enzyme catalysis in atomic detail, mapping of protein dynamics (*16*, *28*–*30*), and the identification of drug building blocks ("fragments") binding to their targets. These experiments provide useful insights for enzyme engineering and drug design and complement methods such as cryo–electron microscopy (cryo-EM) and solution scattering. Prospectively, structural data provided by these comparative crystallographic experiments also provide valuable training data for generative models of protein dynamics (*31*). Although crystal lattice contacts can affect large-scale motions, the weak interactions shaping most protein crystals often allow enzymes to remain active in crystalline form (*1*, *3*, *32*–*39*) and functional dynamics to be triggered by light or rapid mixing with substrate, enabling direct observation by time-resolved x-ray crystallography of biological processes out of equilibrium (*4*, *6*, *9*, *10*, *12*, *21*, *40*–*47*). These observations complement observations of equilibrium dynamics by nuclear magnetic resonance (NMR) spectroscopy as well as biochemical and cell-based functional assays.

Often, conformational changes in a protein are below the detection threshold of conventional analysis methods. We consider a paradigmatic example: two models of photoactive yellow protein (PYP) refined against time-resolved crystallography datasets taken before and 2 ms after exposing the crystal to blue light (henceforth dark and 2-ms datasets, respectively). The light-induced *trans*-to-*cis* isomerization of the active site chromophore (Fig. 2, A and B) in PYP is a conformational change that has been studied extensively (*48*–*50*). Refinement of single models against each of the dark and 2-ms datasets fails to detect this conformational change, revealing no structural differences at all (Fig. 2, C and D). This conclusion is underscored by the $F_o - F_c$ map, which crystallographers use to assess the agreement between a structure and the experimental data.

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. [2]School of Engineering and Applied Sciences, Harvard University, Boston, MA 02134, USA. [3]Graduate Program in Biophysics, Harvard University, Boston, MA 02115, USA. [4]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [5]Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [6]Department of Biology, New York University, New York, NY 10003, USA.
*Corresponding author. Email: doeke_hekstra@harvard.edu (D.R.H.); kmdalton@slac.stanford.edu (K.M.D.)
†These authors contributed equally to this work.
‡Present address: Department of Physics, University of Hamburg, Hamburg 22761, Germany.
§Present address: D. E. Shaw Research, New York, NY 10036, USA.

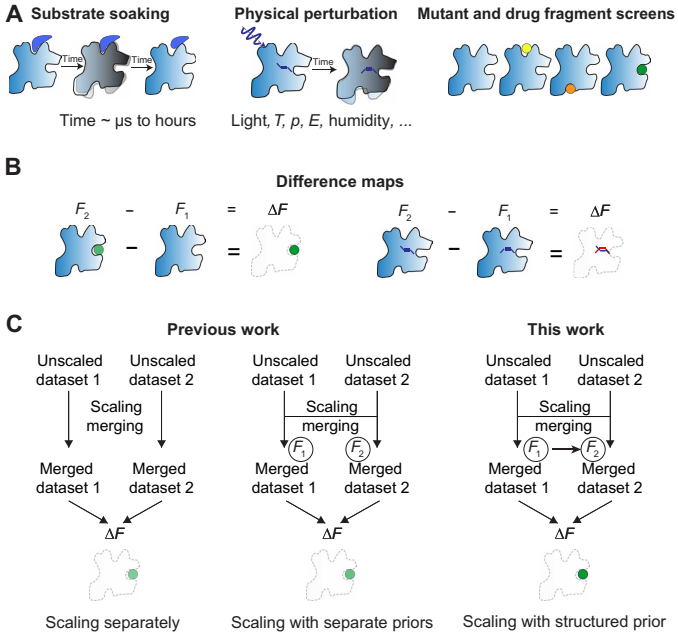

Fig. 1. Schematic overview of comparative crystallography and this work. (**A**) Schematic overview of comparative crystallography experiments. (**B**) Subtracting structure factor amplitudes (top) from isomorphous datasets results in spatial "difference maps" displaying structural differences as changes in electron density. (**C**) Schematic overview of scaling comparative crystallography data. Previous work involved scaling datasets separately (left) or jointly with separate priors (center), while this work involved scaling datasets jointly and with a structured prior (right), which can produce stronger signal in difference maps.

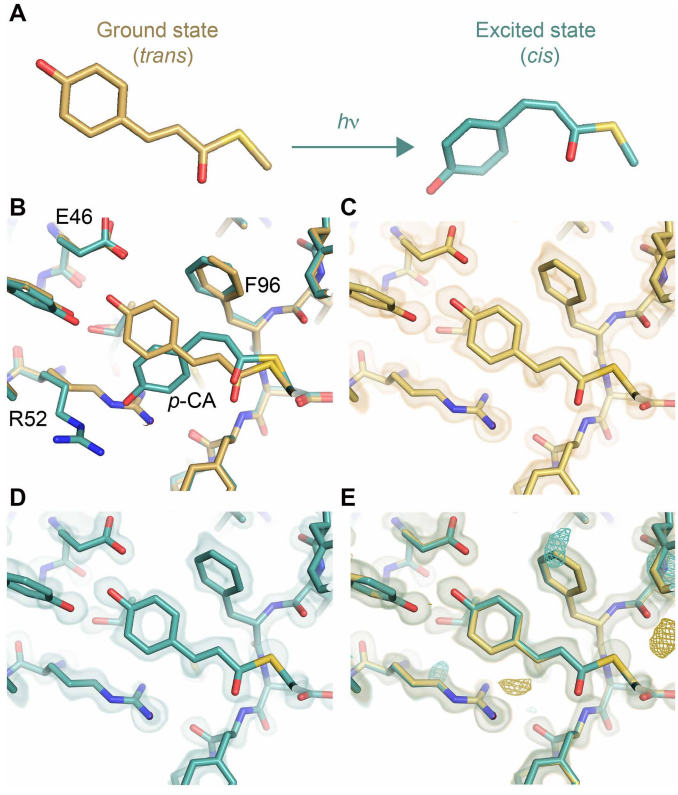

Fig. 2. Schematic overview of comparative crystallography and this work. (**A**) The *p*-coumaric acid (*p*-CA) chromophore of PYP isomerizes when exposed to blue light [adapted from previous work (*54*)]. (**B**) Structural context of the PYP chromophore from Protein Data Bank (PDB) ID 1TS0, in both conformations, with nearby residues indicated. (**C** and **D**) Model of PYP refined against the dark (C) and 2-ms (D) observations before and after exposure to blue light, respectively. The electron density map in each panel is the $2mF_o - DF_c$ map from merging with Careless and refining starting from the ground-state model PDB ID 2PHY. These maps have been contoured to 2σ and carved to 1.5 Å of the model. (**E**) Overlay of models and maps from (C) and (D), along with the $F_o - F_c$ electron density of the 2-ms map and model, a teal and yellow mesh contoured to +3σ and −3σ, respectively.

No difference density appears in the vicinity of the chromophore, indicating that the model appears to agree with the data within experimental error (Fig. 2E and fig. S1). In the end, to resolve the isomerization of the chromophore, the original study did not compare refined structures but visualized changes in electron density using an isomorphous difference map (Fig. 1B) and then refined against differences between the datasets, in a method called extrapolated structure factor (ESF) refinement (*51*, *52*). ESF refinement is popular for inferring perturbed state models in experiments where only a small portion of the crystal has been perturbed (*53*), as is typically the case in nonequilibrium time-resolved experiments and drug fragment screening wherein the fragments typically have low affinity for the protein target. Although ESF refinement has been successful, it relies on merged structure factors that can obscure upstream inaccuracies in data processing. We asked whether we could obtain better sensitivity by considering the comparative nature of crystallography earlier in data reduction.

Here, we introduce a framework that enables statistically efficient comparison of related datasets (Fig. 1C). This method operates at the time of "scaling," during which systematic errors in the unmerged observations are estimated and corrected. We first observe that the structures of proteins at different time points or under different chemical or physical conditions are often very similar and therefore provide nearly redundant measurements. We then formalize this observation into a family of multivariate priors that accurately captures correlations between related crystallographic datasets and implement these structured priors in a software package, Careless, that can scale and merge diffraction data from all major experimental modalities at synchrotrons and XFELs (*54*), as well as microcrystal electron diffraction data (*55*). Careless is based on approximate Bayesian inference combined with a deep neural network that learns scale factors from experimental metadata. By comparing related datasets, we attain large improvements in the detection of signals for key applications—detection of protein dynamics from polychromatic, time-resolved experiments, detection of specific elements in enzymes from serial synchrotron and XFEL experiments, and the detection of bound drug precursors in a large drug fragment screen. In concurrent work (*43*, *56*), we further show that this approach enables direct observation of enzyme catalysis in full spatial and temporal detail. We anticipate that structured priors will broadly improve the detection of structural differences arising from general perturbations across a variety of comparative crystallography experiments.

RESULTS

A statistical framework for comparative crystallography

We first briefly outline our framework for comparing crystallographic datasets, which is explained in mathematical detail in Supplementary

Text. To begin, we note that the complex structure factors that comprise crystallographic datasets are the sum of contributions from each constituent atom, demonstrated as a nine-atom simulation in Fig. 3A. In this situation where the statistical observable is a sum of many random variables, it is reasonable to invoke the central limit theorem, which states that the sum of these variables will follow a normal distribution. Here, each pair of structure factors will follow a bivariate normal distribution in the complex plane. The structure factor amplitudes, which correspond to the distance from the origin in Fig. 3A (dashed circle), follow the well-known Wilson distribution (*57*). Small changes in the positions of constituent atoms, resulting from a perturbation, result in a small change in the total structure factor. In Supplementary Text, we show that two related datasets will closely follow a multivariate normal distribution characterized by a correlation parameter, henceforth the double-Wilson r (*58*, *59*), that quantifies the correlation between the real components of the structure factors. We refer to this statistical model as the double-Wilson model.

To illustrate with a pair of related datasets, Fig. 3B compares structure factor amplitudes for crystals of human protein tyrosine phosphatase 1B (PTP-1B) in the presence and absence (apo.) of the inhibitor TCS-401 (*60*). A fair degree of correlation is evident. A fit to the double-Wilson model (see Supplementary text S2.3) shows that $r \approx 0.85$, with the implied Pearson correlation coefficient between structure factor amplitudes $\approx r^2$ (Supplementary text S5) (*61*). Histograms of structure factor amplitudes for the inhibitor-bound data (Fig. 3, C to E, and fig. S2), given the binned amplitudes of the apo. data (blue shading in Fig. 3B), are better fit by this model (red curves in Fig. 3, C to E) than by assuming statistical independence (blue curves). We likewise find that the double-Wilson model describes other comparative crystallography datasets well, including time-resolved x-ray crystallography experiments (fig. S3), replicate datasets across laboratories (fig. S4), and across temperatures (fig. S5), providing confidence that the formalism generally applies to related datasets.

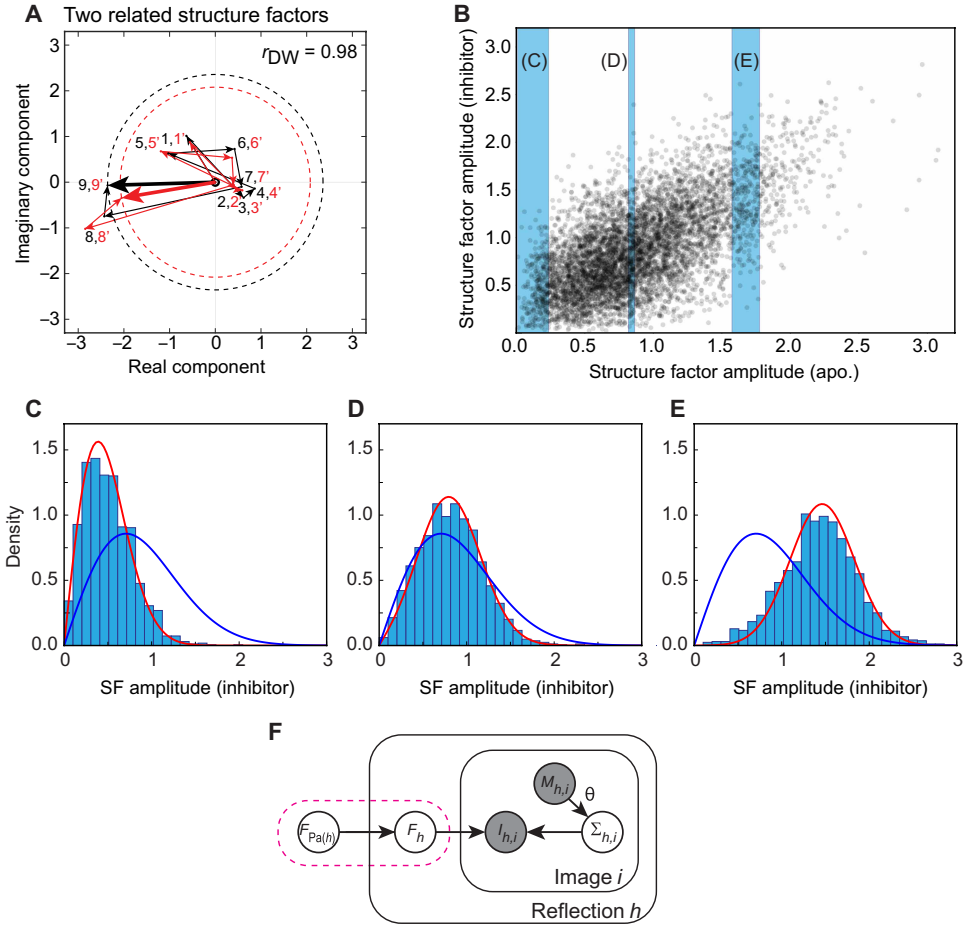

Fig. 3. The double-Wilson model and its implementation in Careless. (**A**) A simulation of complex structure factors of related structures (thick black and red arrows), built up from nine pairs of contributions from individual atoms (numbered black and red arrows). In each atom pair, the contribution from each atom is random but correlated with correlation r_{DW} between the real components and likewise between the imaginary components. The experimentally observed amplitudes of the structure factors are described by the radii of the dashed circles. (**B**) Scatter plot for structure factor amplitudes for isomorphous PTP-1B datasets for apo. and inhibitor-bound forms (for a random subset of Miller indices). Blue slices indicate data points for which histograms are shown in (C) to (E). (**C** to **E**) Histograms for slices through (B) are better approximated by the Rice distributions (red) parametrized by a global r_{DW} (here, 0.85) than by the unconditional, univariate Wilson distribution (blue). (**F**) A graphical model illustrating the probabilistic model underlying Careless: Structure factor amplitudes F_h and scaling corrections Σ_h are inferred from observed diffraction intensities $I_{h,i}$ and reflection- and image-specific metadata $M_{h,i}$. Pink dashed box indicates accounting for expected correlations with a similar structure [the "parent" Pa(h) of reflection h] improves inference.

Scaling of related datasets

Crystallographic datasets inevitably contain systematic errors resulting from numerous physical processes, which confounds detection of dataset differences. These systematic errors are challenging to fully remove (see Supplementary Text S1). We previously introduced a Bayesian statistical framework that simultaneously learns structure factor amplitudes F_h and scales $\Sigma_{h,i}$ from the observed intensities $I_{h,i}$. This framework includes variational approximations of the posterior distributions of the structure factor amplitudes and a multilayer perceptron (with parameters θ) to predict scales from reflection metadata (such as Miller indices, image number, position on the detector, and x-ray wavelength) (*54*). The basic inference challenge is illustrated in Fig. 3F using a probabilistic graphical model (*62*). As in most conventional approaches, we had treated related structure factor amplitudes as statistically independent. However, for comparative applications, we hypothesize that we can better constrain inference using the additional "pseudoredundancy" afforded by simultaneous inference of similar structure factor amplitudes, which could strongly constrain inference [shown in the pink box in Fig. 3F as conditioning of F_h on a "parent amplitude" $F_{\mathrm{Pa}(h)}$]. To test this idea, we implemented the multivariate Wilson distribution as a prior in Careless. We provide guidelines for use of Careless in Supplementary Text and in (*56*, *63*).

Laue anomalous diffraction

Although many x-ray sources are intrinsically polychromatic, most of their photons are typically discarded by use of monochromators to simplify data analysis. For studies of dynamics, however, it can be essential to use the much brighter polychromatic pulses, giving rise to Laue diffraction. To assess whether the use of multivariate priors can improve the processing of Laue data, we first addressed the extraction of anomalous signal, which can be recovered from small differences between Bijvoet pairs—pairs of structure factors that are identical in the absence of element-specific electronic resonances (i.e., anomalous effects) but that differ when these effects are present. Observation of anomalous signal on specific atoms depends critically on stringent removal of systematic errors.

In this example, the data were collected on a single crystal of hen egg white lysozyme soaked with sodium iodide using polychromatic x-ray pulses (about 5% spectral bandwidth, with wavelength from 1.02 to 1.20 Å). As these data were obtained from a single crystal, one might expect little inconsistency across the data and minimal benefit from imposing a bivariate prior on Bijvoet pairs. Nevertheless, we compared scaling and merging in Careless with a univariate Bayesian prior (i.e., the standard Wilson distribution) against a bivariate prior that assumes varying levels of correlation, r, between the Bijvoet mates. For consistency, we phased all sets of structure factor amplitudes using a reference model refined against monochromatic data collected on the same day. In the resulting anomalous difference map, we find strong gains in peak heights of both iodide ions and sulfur atoms when using a bivariate prior (Fig. 4A and Fig. S6A), although the anomalous scattering strength of sulfur is equivalent to only about one-fourth of an electron at the dominant wavelength of the incident x-ray spectrum ($f'' = 0.26$ at 1.04 Å). The signal strength is maximal at an intermediate r of 0.999, with an increase in anomalous peak height of about 70% over the use of a univariate prior (Fig. 4B and fig. S6B). As a guide to choosing r, we recommend scanning a range of possible values. A good estimate can typically be obtained within minutes by scaling and merging a small batch of data (see fig. S7, where we processed the first 12.5% of data). Alternatively, Careless also supports automatic optimization of r (see Supplementary Text).

Cross-validation measures of data processing quality—$CC_{1/2}$ (*64*), CC_{pred} (*63*), and CC_{anom} (*65*)—are also improved by the introduction of the bivariate prior (Fig. 4, C and D). Careless with a bivariate prior also strongly improves over the long-standing state of the art for scaling polychromatic diffraction data, Epinorm (Renz Research Inc.), increasing anomalous signal by 73% for iodine atoms and 118% for sulfur atoms. Notably, the anomalous signal increases substantially after the double-Wilson r exceeds about 0.96 (Fig. 4B). The same trend was observed when the images were processed with the open-source package Laue-DIALS (*66*) (figs. S8 and S9).

Time-resolved Laue diffraction

Laue diffraction at synchrotron sources provides access to the dynamics of proteins on timescales of 100 ps and longer. To determine whether a bivariate prior can improve scaling of time-resolved signal, we processed a time-resolved Laue dataset collected on a single crystal of PYP at BioCARS 14-ID at the Advanced Photon Source. Our dataset, the same as in Fig. 1, contains 20 images of PYP without laser exposure and 20 images 2 ms after exposure to a blue laser pulse, at which point PYP is expected to adopt an intermediate state (*50*). Using Careless, we jointly scaled the dark and 2-ms datasets while imposing a bivariate prior with varying levels of the correlation parameter, r. Merging statistics improve when scaling with a bivariate prior (Fig. 5A).

The extensive literature on PYP enables us to quantify the quality of the resulting electron density difference maps: On the basis of models of the ground-state and excited-state conformations from previous studies, we can calculate correlation coefficients between observed and predicted difference maps. At the best value of r (about 0.9995), this real-space correlation coefficient is more than 0.50 within 10 Å of the chromophore. Without weighting (*67*), this correlation coefficient is about twice as strong as for an observed difference map obtained by either scaling with a univariate prior or scaling with Epinorm, the state of the art for scaling Laue diffraction data (Fig. 4B). Consistent with this quantitative assessment, the difference map near the chromophore is more readily visually interpretable after scaling with a bivariate prior (Fig. 5C and fig. S10) than without. A scan of difference map quality on a small subset of the data again suffices to select r (fig. S9A). We note that alternative approaches that adjust scales after merging can sometimes improve visual interpretability (fig. S11) but are not, in this case, effective in improving correlation with the calculated maps ("SCALEIT" and "local scaling"; Fig. 5B and see Supplementary Text for further discussion). That is, our approach improves detection of dynamics from time-resolved x-ray data over state-of-the-art methods.

Last, it is instructive to compare our scaling approach with commonly used weights (*8*, *29*, *53*, *54*) that are heuristic modifications of a weighting scheme proposed in (*67*), inspired by Bayesian statistics. These weights reduce the contributions of estimated structure factor amplitude differences (ΔF) that are large or have large estimated errors [$\sigma(\Delta F)$]. A coefficient α, used in calculating the weights, tunes the contribution of large ΔF to the difference map (see Materials and Methods). In the case of the PYP data (Fig. 5C), we find that suppression of large $|\Delta F|$ (by choosing $\alpha > 0$) is necessary to achieve a similar effect as the bivariate prior. That is, the bivariate prior

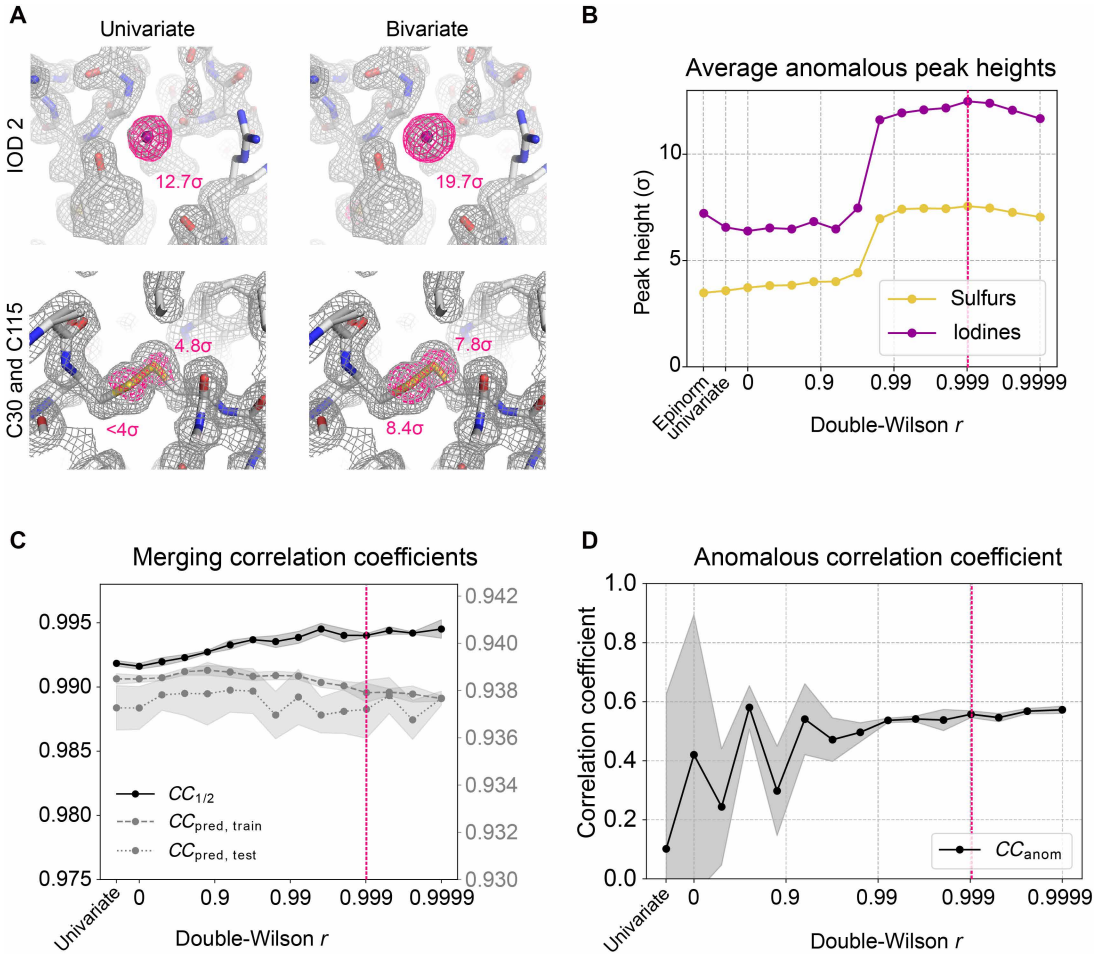

Fig. 4. Use of a bivariate prior improves anomalous signal in a Laue diffraction experiment. (**A**) Comparisons between anomalous omit peaks after scaling and merging with univariate or bivariate priors [using $r = 0.999$ for the bivariate prior, marked in (B) to (D) with a dashed magenta line]. The observed electron density map ($2mF_o - DF_c$) in gray is contoured at 1.5σ, and the anomalous difference omit map in magenta is contoured at 6σ for iodine (IOD) 2 and 4σ for C30 and C115. (**B**) Dependence of average iodine and sulfur anomalous peak height on *r*. Additional peak heights are plotted for merging with the univariate prior or with Epinorm. (**C**) Merging correlation coefficients of the lysozyme dataset across double-Wilson *r* values. The *y*-axis labels for the $CC_{1/2}$ are on the left, and the *y*-axis labels for the CC_{pred} are on the right. A test set of 10% of observations were held out during scaling and merging to evaluate performance of the scaling model, yielding $CC_{pred, test}$ for the test set and $CC_{pred, train}$ for the 90% of data used during scaling. The shaded confidence interval of the $CC_{1/2}$ curve represents the SD over three random half-dataset repeats, and the shaded confidence interval for the CC_{pred} curves represent the SD over three Careless runs. (**D**) Anomalous correlation coefficient, CC_{pred}, of the lysozyme dataset across *r*. The shaded confidence interval represents the SD over three random half-dataset partitions.

suppresses spuriously large contributions to the difference map without the need for heuristic weighting schemes.

Serial femtosecond crystallography

Serial femtosecond x-ray crystallography (SFX) of microcrystals probed by intense XFEL pulses has markedly expanded the range of systems and timescales accessible to time-resolved studies. To determine whether multivariate priors could also improve the accuracy of SFX data processing, we sought to extract anomalous signal from an SFX dataset collected for thermolysin, an enzyme containing a catalytic zinc (Zn) ion in its active site. Specifically, we processed 3160 images from a much larger dataset [Coherent X-ray Imaging Data Bank (CXIDB) 81] (*68*, *69*). To account for the serial monochromatic data collection strategy, we included an Ewald offset and per-image layers in our scaling model, as before (*54*). We split Bijvoet pairs into two half datasets before relating these half datasets by a joint prior during scaling in Careless. The anomalous difference map peak height of the catalytic zinc ion (ZN 317) increases from 15σ without a bivariate prior to 28σ for the optimal *r* value, surpassing state-of-the-art XFEL-specific software [Computational Crystallography Toolbox (CCTBX); Fig. 6A] (*54*, *70*). We also find that two estimation methods can get us close to the best *r* value found by our scan: use of a small subset of data (fig. S12A) or direct optimization of *r* in Careless (fig. S13).

As *r* is varied, we can detect more anomalous sites. First, more associated calcium ions become detectable. Using a univariate prior, only two of four calcium ions appear above 3σ (an approximate noise threshold). After scaling with a multivariate prior, all four calcium ions are detectable (Fig. 6, A and B). Second, in prior analyses of a thermolysin dataset with four times as many diffraction images, anomalous difference signal attributable to a low-occupancy zinc ion was detected at a site adjacent to catalytic zinc ion ZN 317 (*68*, *71*).

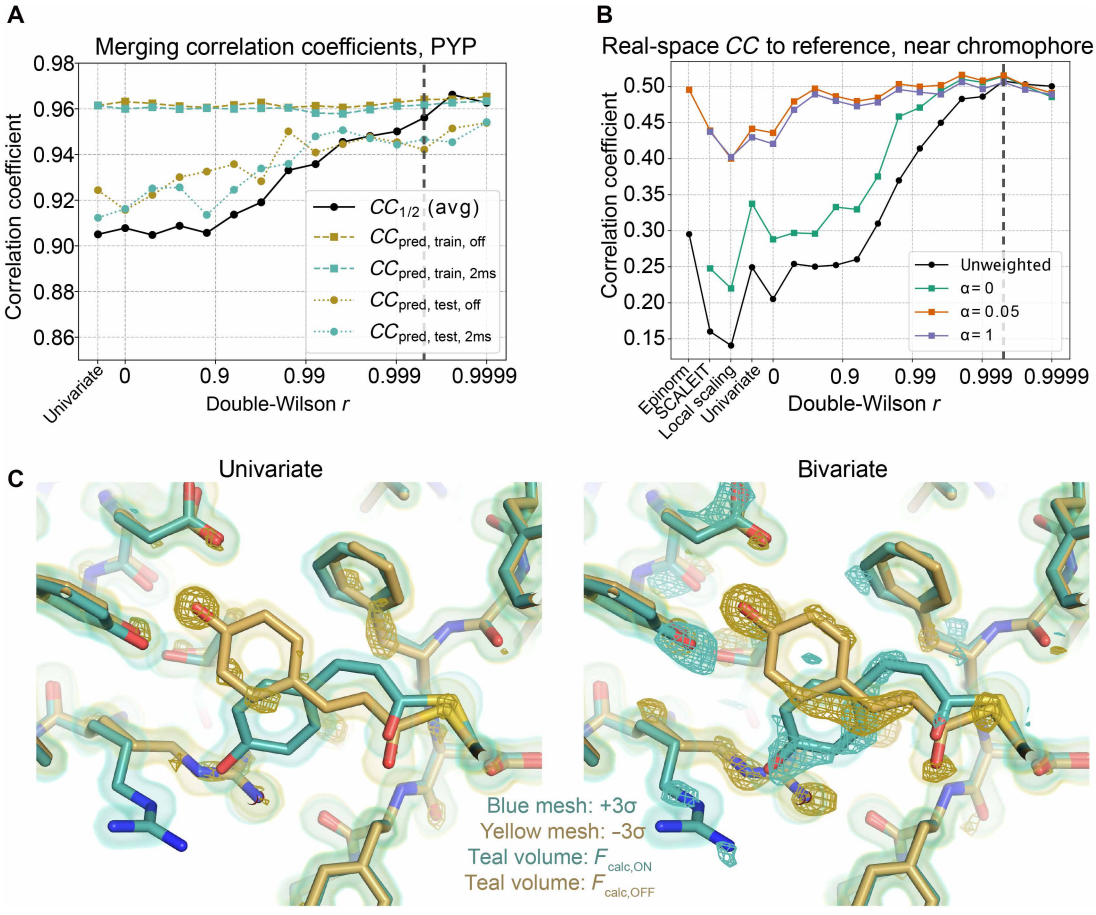

Fig. 5. Use of a bivariate prior improves time-resolved difference signal. (**A**) Merging statistics of the 2-ms and dark structure factor amplitudes as a function of *r*. (**B**) Correlation coefficient of the weighted (α = 0, green; α = 0.05, red; α = 1, blue) and unweighted (black) observed difference maps with the 2 ms − dark calculated difference map ($F_{calc,2ms} - F_{calc,OFF}$), in a region 10 Å around the chromophore. α weights the contributions of large structure factor differences (see Materials and Methods). The correlation coefficient is plotted as a function of the double-Wilson parameter *r*. Benchmarks included are as follows: (i) Epinorm, (ii) Careless with univariate prior, and Careless with the univariate prior followed by (iii) local scaling (*103*) or (iv) SCALEIT (*104*). Vertical dashed lines in (A) and (B) indicate the *r* value, 0.9995, that parametrizes the bivariate prior for merged data displayed in (C). (**C**) Left: Observed difference map (teal and yellow mesh, contoured to 3σ, unweighted), from merging with a univariate prior. The map is overlaid with models (2-ms and dark models colored with teal and yellow sticks, respectively) and the electron density at 2 ms after excitation and in the dark state (teal and yellow surfaces, respectively; contoured to 1.5σ), based on calculated structure factors F_{2ms} and F_{dark} from PDB entry 1TSO. Right: Isomorphous 2-ms − dark difference map (yellow and teal mesh, contoured to 3σ, unweighted), from merging with a bivariate prior, *r* = 0.9995. F_{2ms} map and F_{dark} map are shown in the left panel.

When we previously scaled our smaller dataset with a univariate prior, this site (ZN 317.1) was not detectable (*54*). Upon scaling with a multivariate prior, however, this site can be clearly identified (Fig. 6B, black arrow), emerging above the noise with as few as ~1000 diffraction images (Fig. 6C). We note that these anomalous difference maps can also be improved by weighting, but this is less effective than the use of a multivariate prior during scaling (fig. S12C). In summary, a bivariate prior strongly improves the detection of anomalous signal from SFX data.

Fragment screening

Last, crystallographic drug fragment screens now enable the high-throughput discovery of pharmacological lead compounds against proteins of interest, e.g., SARS-CoV-2 proteins (*13*–*15*, *26*). In these studies, thousands of crystals of a protein of interest are each soaked with a "drug fragment"—a small-molecule building block for larger drug molecules. To determine whether joint scaling can improve the detection of bound fragments, we analyzed data from a fragment screening experiment totaling 1 apo. dataset and 16 holo datasets identified by PanDDA (Pan-Dataset Density Analysis) (*72*), a state-of-the-art algorithm based on real-space electron density map manipulations. This dataset contains structures of SARS-CoV-2 nonstructural protein 3 macromolecular domain 1 (Nsp3 Mac1) with ligands from a fragment screening library bound to the Mac1 adenosine-binding site.

The statistical model generalizes to more than two datasets. In this case, each dataset is represented by a node in a Bayesian network or graphical model (Fig. 7A, left) (*62*). Edges (arrows) represent conditional dependence between datasets. Whenever this graph is acyclic, the resulting joint probability of structure factor amplitudes can be calculated analytically (see Supplementary Text). We will refer to the resulting probability distribution as the multivariate Wilson distribution. The expected correlations in many comparative crystallography experiments can be modeled by these networks (Fig. 7A, right, and fig. S14). Our implementation of the multivariate Wilson

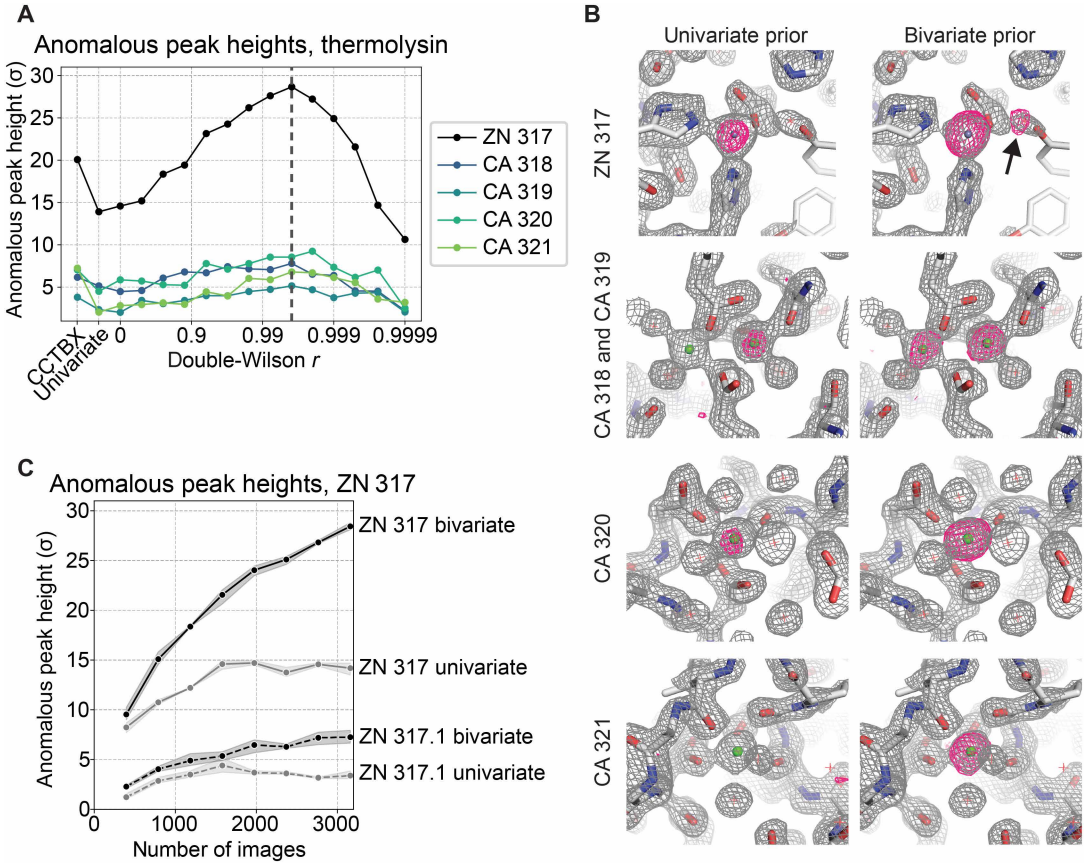

Fig. 6. Scaling and merging with a bivariate prior improves the anomalous signal in a serial XFEL experiment. (**A**) Anomalous difference map peak heights in thermolysin, shown for zinc and four calcium sites, as a function of the double-Wilson parameter *r*. Additional peak heights are shown for merging with the univariate Bayesian prior and merging with CCTBX (*54*, *70*). Vertical dashed black line in (A) indicates the *r* value, 0.9961, that parametrizes the bivariate prior for merged data displayed in the following panel. (**B**) First row: Region near the thermolysin-bound zinc atom, with observed electron density map ($2mF_o - DF_c$) in gray, contoured at 2σ, and anomalous difference omit map in magenta, contoured at 5σ. Arrow indicates the location of an alternative zinc binding site revealed by merging with the bivariate prior. Next three rows: Anomalous difference omit maps of thermolysin calcium sites, carved to 1.5 Å of the model and contoured to 4σ (magenta). (**C**) The anomalous peak height of ZN 317 and its alternate site, ZN 317.1, as a function of the number of frames, at $r = 0.9951$. Shaded band represents 95% confidence interval over three independent scaling repeats.

model expects users to supply the topology of conditional dependencies between datasets and an estimate of the double-Wilson parameter *r*. For practical guidelines on this approach, see Materials and Methods, Supplementary Text, and (*56*).

Following this paradigm, we ran Careless on unmerged intensities, adding several metadata keys, including one-hot encodings of each dataset (see Materials and Methods). By one-hot encoding, we allow the model to express differences in scale between fragment-bound datasets. In addition, we model each holo dataset as dependent on the apo. dataset, while assuming conditional independence between holo datasets (Fig. 7A, right). We find that difference density maps visually improve upon scaling with a multivariate prior (Fig. 7B and fig. S14). We tracked the largest ligand peaks in the difference map for a given dataset, while we varied *r*. We find that in all datasets, this signal is strongest when scaling with a multivariate prior (Fig. 7C). The *r* value for the tallest peaks differs between holo datasets. Without special modifications to scaling and merging, we obtain peak heights on par with those produced by PanDDA (Fig. 7D and table S2). We note that PanDDA analysis outperforms difference map analysis after scaling with a bivariate prior when measured by real-space correlation with a reference difference map (fig. S16A) calculated from bound and apo. models generated by PanDDA. This real-space correlation diminishes more strongly for PanDDA when the ligand mask is expanded (fig. S16B). This may be because PanDDA warps electron density maps using a structure-guided map alignment algorithm, which may suppress the detection of protein conformational changes. Since PanDDA analysis is performed after scaling and merging, further gains may be possible by combining our scaling method with PanDDA.

To determine whether the correlation between apo. and holo datasets, measured by *r*, relates to ligand occupancy, we examined the relationship between optimal *r* and the PanDDA background density correction factor (BDC; a parameter that scales inversely with ligand occupancy and accounts for crystal-to-crystal variation in soaking and data quality) (*27*). We find a moderate correlation with double-Wilson *r* (Pearson $r = 0.505$, two-tailed *t* test: $P < 0.05$, $N = 16$) (fig. S16D). Merging statistics also improve when scaling with a multivariate prior (fig. S16E): in particular, the gap between the training and test CC_{pred} is reduced for higher *r*, indicating that the multivariate prior reduces overfitting of the scales.

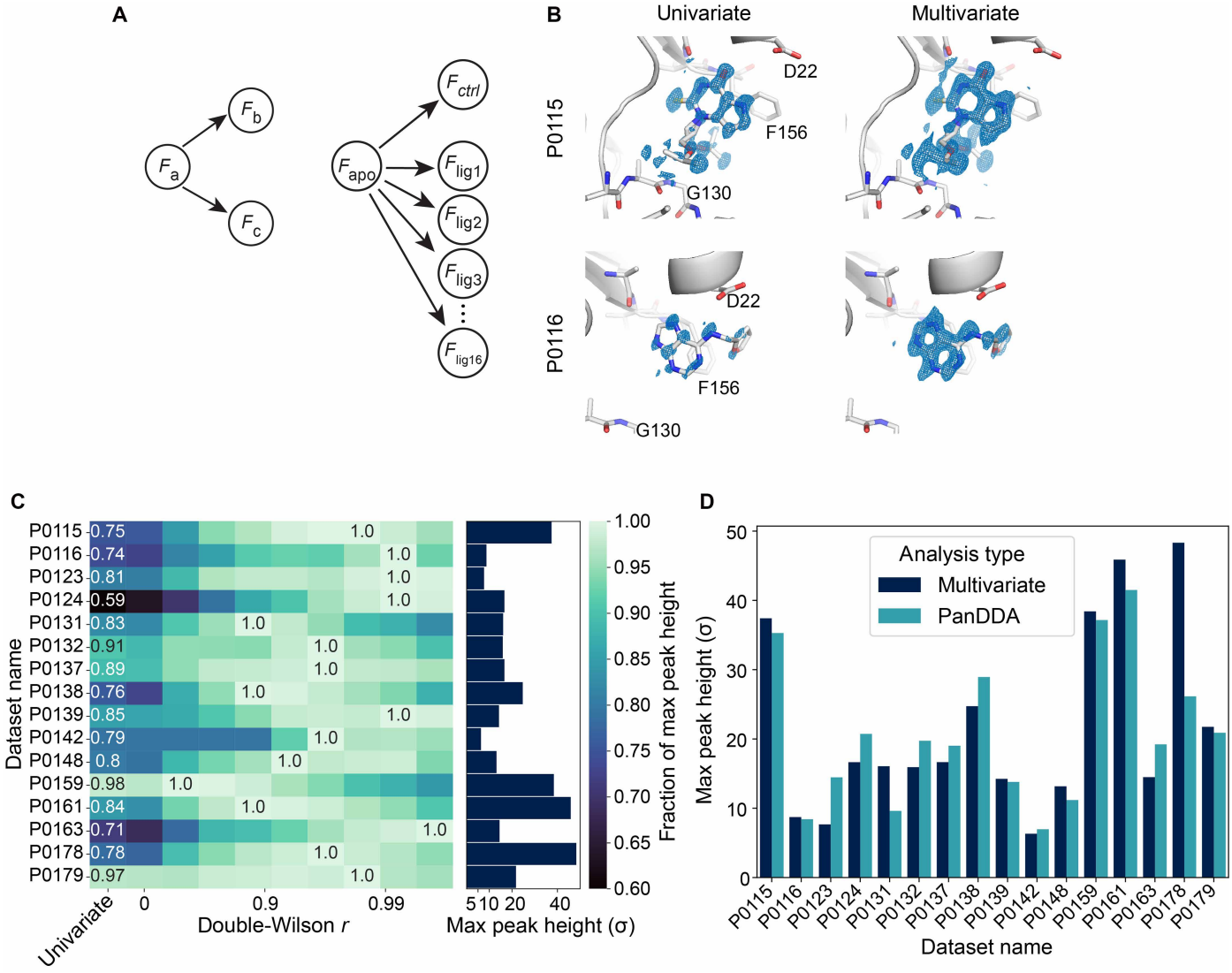

Fig. 7. Use of a multivariate prior improves bound fragment signal in a drug fragment screen. (**A**) A graphical model (or Bayesian network) illustrating the relationship between three datasets (or, more generally, sets of structure factor amplitudes). Arrows indicate conditional dependence; the lack of an arrow between datasets F_b and F_c indicates conditional independence. (**B**) holo − apo. isomorphous difference maps from fragment screening of Mac1, scaled with a univariate or multivariate prior. All maps carved to 1.5 Å from the ligand and contoured to +3σ. (**C**) Tallest positive peak in each holo − apo. difference map as a function of the double-Wilson *r* parameter. For a given holo dataset, the tallest positive peak is measured. The maximum peak height of that peak, across all double-Wilson *r* values, is normalized to 1.0. The maximum peak height is labeled and reported in a bar chart (right). Also labeled is the fraction of the maximum peak height for the tallest peak in the univariate datasets (left). The *r* value, for each dataset's maximum peak height, parametrizes the bivariate prior for merging the electron density displayed in (B). (**D**) Comparison of each dataset's ligand peak heights as processed with PanDDA or Careless with bivariate prior.

DISCUSSION

These results establish that a Bayesian formalism using multivariate priors enables machine learning methods to sensitively detect protein dynamics, chemical signals, and ligand binding from comparative crystallographic data by exploiting correlations among related datasets. In concurrent work, we show that the formalism also strongly aids identification of intermediate states of an enzyme during catalysis (*43*). Naturally, in each case, the biological significance of observed motions is best assessed in combination with other types of data, such as functional readouts of mutations suggested by the crystallographic observations. An intense focus by the community on the prediction and observation of protein dynamics, combined with rapid advances in XFEL and synchrotron instrumentation, suggests that our work will find wide application in protein engineering, drug discovery, and the study of protein dynamics.

We provide further use examples of Careless in its online documentation (*73*), explain use of the available options in (*63*), and describe an ablation study identifying factors affecting Careless' performance in (*56*). We discuss best practices for use of the multivariate prior in Supplementary Text and in extended online descriptions for the cases presented here (*74*). The statistical framework introduced here is not specific to Careless and can be implemented in other software. We encourage adoption of this framework by other crystallographic data processing software. Furthermore, the current memory limitations of

Careless will be addressed in a forthcoming software package (*75*) that also builds on the presented statistical framework.

Our formalism has been inspired by successful Bayesian approaches to the crystallographic phase problem (*59*) and refinement (*76*). Although the developed formalism proves very effective in the form presented here, we note several promising extensions. First, high-quality reference datasets are commonly measured for serial time-resolved crystallography experiments, but these data are not commonly used to improve scaling. These datasets can be added to the Bayesian networks shown in fig. S14 to better condition scaling. Second, we currently treat the correlation parameter r as a global hyperparameter, yet correlations may vary between pairs of datasets (e.g., as a function of ligand occupancy) and vary with resolution. These correlations could be estimated from the data. Third, as not every node in a Bayesian network needs to be observed, our approach allows for interpolation of missing observations, enabling alternative data collection strategies that trade temporal coverage and completeness per time point (*77*). Fourth, the developed framework implies a joint distribution of the unobserved crystallographic phases across related datasets (see Supplementary Text and fig. S17). Future algorithms for model building and structure refinement algorithms could take advantage of such information. Last, cryo-EM and NMR can also measure heterogeneity in samples. These approaches are currently low throughput. Cryo-EM is still limited to large domain motions at slower timescales, while NMR is difficult to apply to large proteins or complexes. Future advances in cryo-EM and NMR could enable similar comparative techniques to be developed, growing the scope of perturbative methods to study protein dynamics. Specifically, the presented formalism may be readily extended to other methods for which the Wilson distribution provides a natural prior such as single-particle cryo-EM (*78*, *79*) and coherent x-ray imaging (*80*), as well as measurements of disordered crystals (*81*).

To take a broader perspective, our framework demonstrates that comparing dataset differences is superior to comparing individually processed structures when it comes to detecting subtle differences. Small structural changes are fundamental to our understanding of key biophysical phenomena such as drug binding, structural dynamics, and the effects of mutations. However, the status quo for structural comparison is to refine individual models against individual datasets and then compare models. In this status quo paradigm, subtle differences between conformations are readily obscured by inconsistent corrections for artifacts. We advocate for using structured priors as early as possible in the analysis pipeline. We expect this conclusion to be applicable in many domains within structural biology and beyond.

MATERIALS AND METHODS

Implementation details

The multivariate Wilson distribution we present here is implemented as a prior distribution in the Careless software package. Careless uses gradient-based optimization to maximize a Bayesian objective function (*82*, *83*), which is the sum of the expected log likelihood of the observed data plus the Kullback-Leibler divergence between the structure factors and the prior distribution. The likelihood of the data describes the probability of the observed diffraction intensities, $I_{h,i}$, conditional on scale $\Sigma_{h,i}$ and structure factor amplitudes F_h. The likelihood also accounts for the estimated errors $\sigma_{I_{h,i}}$ in the observed intensities. The objective function is computed as described in equations 21 and 22 of Dalton *et al.* (*54*) using reparametrized samples (*84*) from a truncated normal surrogate posterior (variational distribution). For clarity, we reproduce this objective function

$$\mathrm{ELBO}(\mu_{qF}, \sigma_{qF}, \theta) \approx \sum_{s=1}^{S}\left[\sum_i \log p\left(I_{h,i}|F_{h,s}^2\Sigma_{h,i,s}, \sigma_{I_{h,i}}\right) - \sum_h\left(\log q_{F_h}(F_{h,s}) - \log p(F_{h,s})\right)\right]$$

The evidence lower bound (ELBO) depends on the parameters of the surrogate posterior distributions, q_F and q_Σ. q_Σ is parameterized by a neural network with parameters θ, whose architecture is described in previous work (*54*) q_F is parameterized by independent truncated normal distributions for each reflection with location and scale parameters $\{\mu_{q_{F_h}}, \sigma_{q_{F_h}}\}$ supported on $[0,\infty)$ for centric reflections and $(0,\infty)$ for acentrics. $I_{h,i}$ refers to the intensity of a particular observation i for reflection h. s indexes reparameterized samples from the surrogate posteriors

$$F_{h,s} \sim q_{F_h}$$
$$\Sigma_{h,i,s} \sim q_{\Sigma_{h,i}}$$

Reparameterization is implemented in the TensorFlow Probability library (*85*). The prior distribution enters the objective function through the ultimate term, $\log p(F_{h,s})$. Previously (*54*), we exclusively used a univariate Wilson distribution as a prior for this term

$$p(F_{h,s}) = \begin{cases} \sqrt{\frac{2}{\pi\varepsilon_h}}\exp\left(-\frac{F_{h,s}^2}{2\varepsilon_h}\right) & h \text{ is centric} \\ \frac{2}{\varepsilon}F_{h,s}\exp\left(\frac{F_{h,s}^2}{\varepsilon_h}\right) & h \text{ is acentric} \end{cases}$$

where ε_h is the multiplicity of the reflection h, an integer value determined by the crystal's symmetry. Note that because the scale function Σ_h is parametrized separately within Careless, it does not appear in expressions for $p(F_{h,s})$.

One of the features of variational inference is its flexibility. In this work, we leverage the inherent flexibility of reparameterization-based variational inference to extend the Careless package with a multivariate prior distribution. To do so, the Careless command line interface requires users to supply separate unmerged datasets for each node they wish to model (as in Fig. 7A and fig. S14). Careless asks the user to specify the parent of each node and the expected double-Wilson parameter r. Internally, this changes how the prior is calculated as follows

$$p(F_{h,s}) = \begin{cases} \mathrm{FoldedNormal}\left[F_{h,s}\middle|rF_{\mathrm{Pa}(h),s}, \sqrt{\varepsilon_h(1-r^2)}\right] & h \text{ is centric} \\ \mathrm{Rice}\left[F_{h,s}\middle|rF_{\mathrm{Pa}(h),s}, \sqrt{\frac{\varepsilon_h}{2}(1-r^2)}\right] & h \text{ is acentric} \end{cases}$$

where $\mathrm{Pa}(h)$ denotes the parent reflection of h. Where a particular reflection has no parent because it is the root node, the prior defaults to the univariate Wilson model. When the parent observation is systematically absent (0) because of crystal symmetry but the child observation is not constrained to be 0, we use the above expression with $F_{\mathrm{Pa}(h),s} = 0$. The folded normal distribution's probability density function is

$$\begin{aligned}&\text{FoldedNormal}(x|\mu,\sigma)\\&=\text{Normal}(x|\mu,\sigma)+\text{Normal}(x|-\mu,\sigma)\\&=\frac{1}{\sigma\sqrt{2\pi}}\left\{\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]+\exp\left[-\frac{1}{2}\left(\frac{x+\mu}{\sigma}\right)^2\right]\right\}\end{aligned}$$

while that of the Rice distribution is

$$\text{Rice}(x|v,\sigma)=\frac{x}{\sigma^2}\exp\left[-\frac{(x^2+v^2)}{2\sigma^2}\right]I_0\left(\frac{xv}{\sigma^2}\right)$$

where I_0 denotes the modified Bessel function of the first kind with order zero. For numerical stability, both probability densities are computed directly in log space.

Photoactive yellow protein

Collection of the Laue diffraction data for PYP was described previously (*86*). We performed indexing, geometry refinement, wavelength assignment, and integration using Precognition version 5.2.2 (Renz Research Inc.). Precognition-integrated intensities were written to mtz files and then scaled and merged with Careless version 0.4.1, with use of a bivariate prior on the dark and 2-ms time point mtz files. Fifteen different double-Wilson r values were used, with r varied for each Careless run as 1 to $0.5^{\text{run number}}$. Epinorm (Renz Research Inc., as bundled with Precognition) was also used for scaling and merging these data. Correlation coefficients were computed using careless.cchalf, careless.ccpred, and custom scripts based on GEMMI version 0.6.2 (*87*) and reciprocalspaceship version 1.0.1 (*88*). Isomorphous difference maps were first phased with a dark model refined in PHENIX version 1.20.1 (*89*) from Protein Data Bank (PDB) ID 2PHY (*90*) and then weighted using reciprocalspaceship according to

$$w_{hkl}=\left(1+\frac{\sigma^2(\Delta F)}{\langle\sigma^2(\Delta F)\rangle}+\alpha\frac{|\Delta F|^2}{\langle|\Delta F|^2\rangle}\right)^{-1}$$

for structure factor amplitude differences ΔF (short for ΔF_{hkl}) and their errors $\sigma_{\Delta F}$, with angled brackets indicating averages over Miller indices. Using the same refinement procedure, dark and 2-ms models were refined for Fig. 2. Local scaling was performed using a custom script available in the Zenodo deposition and SCALEIT (CCP4 software suite version 7.1) was run from rs-booster version 0.1.2 (https://github.com/rs-station/rs-booster) as rs.scaleit (see Supplementary text S7).

Difference maps were visualized in PyMOL version 2.5.2 (*91*). Data processing scripts can be found in the Zenodo deposition.

Lysozyme data collection

NaI-soaked lysozyme crystals were prepared as described previously (*92*). Monochromatic data were collected on the Northeastern Collaborative Access Team (NE-CAT) beamline 24-ID-C (Advanced Photon Source, Argonne National Laboratory) on 12 August 2020. Diffraction data were collected at ambient temperature (about 295 K). Three 1440-image passes (720°) were collected from a single crystal of lysozyme with an exposure time of 0.1 s and an oscillation angle of 0.5°. The incident x-ray intensity, at an energy of 11.95 keV, was attenuated to 0.5% transmission, which corresponds to an estimated flux of 2.4×10^{10} photons s^{-1}. Data were collected using helical acquisition, so that dose was evenly distributed along the crystal. The PILATUS 6M-F detector (Dectris) was positioned at the minimal distance of 150 mm.

Laue data were collected on the same day at BioCARS (Advanced Photon Source, Argonne National Laboratory), similar to data collection described in previous work (*29*), except that Laue stills were collected in 1° steps. Data were collected at ambient temperature (about 295 K).

Thermolysin and lysozyme

Careless 0.4.1 was run with a bivariate prior, sweeping over 15 double-Wilson r values on two datasets, one of thermolysin containing 3160 images collected from a serial XFEL experiment, CXIDB 81 (*68*) and one of lysozyme containing the first 720 images from the above data collection. Lysozyme data were indexed and integrated using Precognition (Renz Research Inc.) and Laue-DIALS version 0.4 (*66*) (https://github.com/rs-station/laue-dials). Thermolysin unmerged intensities (in DIALS .pickle format) were obtained from CXIDB 81 (www.cxidb.org/id-81.html). Both datasets were split according to Friedel symmetry and scaled and merged in Careless, applying the bivariate prior to relate anomalous half datasets. In the case of thermolysin, direct optimization of r was done in Careless 0.5.3. Next, the anomalous half datasets were combined into single datasets of merged structure factor amplitudes. Correlation coefficients on these datasets were computed using careless.cchalf, careless.ccpred, and careless.ccanom (*63*). For lysozyme, we used Epinorm (Renz Research Inc.) as a benchmark for scaling and merging. The scaling benchmark for thermolysin (using cctbx.xfel version 2021.11.dev3 + 4.g05389c3054) was described previously (*54*). Phenix version 1.20.1 (*89*) was used to phase the resultant structure factor amplitudes by isomorphous replacement with PDB ID 2TLI in the case of thermolysin and with a high-resolution monochromatic model in the case of lysozyme (table S1). In addition, an anomalous omit map was phased. The phases and omit map were then used for visualization of anomalous omit peaks with PyMOL and measurement of peak heights with reciprocalspaceship. Data processing scripts can be found in the Zenodo deposition.

Fragment screening of Mac1

Data collection and analysis of a crystallographic drug fragment screen of SARS-CoV-2 Nsp3 Mac1 have been previously described (*14*). For the work described here, 1 unmerged apo. dataset and 16 unmerged holo HKL files (previously processed using XDS (X-ray Detector Software) version 31 January 2020) were converted to mtz files and then processed with Careless version 0.4.1, with use of a univariate or multivariate prior to relate the apo. and 16 holo datasets. The Bayesian network was trained following fig. S14, with dependencies drawn between the apo. dataset and each of the 16 holo datasets (see Supplementary Text). In addition, we introduced a control apo. dataset, a duplicate of the reference that also depended on the reference apo. dataset, and difference maps were computed by subtracting each holo dataset from the control apo. dataset. To compensate for the duplication of the reference apo. dataset, we multiply the errors of the intensities by $\sqrt{2}$. The double-Wilson r was set to a uniform value for all 16 apo-holo dependencies and was varied between Careless runs as 1 to $0.5^{\text{run number}}$. Metadata keys included observed Miller indices, detector position and positional encoding, image number, the azimuthal angle PSI (*93*), diffracted beam direction parameters ALF1 and BET1 (*94*), and dataset one-hot encoding. The PSI angle, ALF1, and BET1 metadata keys are from integration in XDS. To perform

one-hot encoding of dataset IDs, one metadata key is created for every dataset. Each reflection receives the value 1 in the corresponding dataset column and elsewhere 0. One-hot encoding was implemented using a reciprocalspaceship script. Correlation coefficients were computed using careless.ccpred. Apo. phases were obtained by refinement of a model against the apo. amplitudes scaled using a univariate prior. Refinement was performed using PHENIX version 1.20.1. The apo. phases were then used for creating isomorphous difference maps of each holo minus apo. dataset. We created difference maps and calculated peak heights and real-space correlation coefficients using rs-booster version 0.1.1 (https://github.com/rs-station/rs-booster). To compute an $F_c - F_c$ map, we used phenix.fmodel to compute F_c values from PDB ID 7KQO as the apo. model and PanDDA models as the holo models. We then scaled the $F_{c,holo}$ to the $F_{c,apo}$. PanDDA models, 1-BDC values, and Z maps were taken from an analysis that included all previously reported datasets (*14*, *72*). Data processing scripts can be found in the Zenodo deposition.

Supplementary Materials

This PDF file includes:

Supplementary text
Figs. S1 to S19
Tables S1 and S2
References

REFERENCES AND NOTES

1. J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* **391**, 304–307 (1998).
2. N. Chim, R. A. Meza, A. M. Trinh, K. Yang, J. C. Chaput, Following replicative DNA synthesis by time-resolved x-ray crystallography. *Nat. Commun.* **12**, 2641 (2021).
3. T. Nakamura, Y. Zhao, Y. Yamagata, Y.-j. Hua, W. Yang, Watching DNA polymerase η make a phosphodiester bond. *Nature* **487**, 196–201 (2012).
4. P. Mehrabi, E. C. Schulz, R. Dsouza, H. M. Müller-Werkmeister, F. Tellkamp, R. J. D. Miller, E. F. Pai, Time-resolved crystallography reveals allosteric communication aligned with molecular breathing. *Science* **365**, 1167–1170 (2019).
5. T. Gruhl, T. Weinert, M. J. Rodrigues, C. J. Milne, G. Ortolani, K. Nass, E. Nango, S. Sen, P. J. M. Johnson, C. Cirelli, A. Furrer, S. Mous, P. Skopintsev, D. James, F. Dworkowski, P. Båth, D. Kekilli, D. Ozerov, R. Tanaka, H. Glover, C. Bacellar, S. Brünle, C. M. Casadei, A. D. Diethelm, D. Gashi, G. Gotthard, R. Guixà-González, Y. Joti, V. Kabanova, G. Knopp, E. Lesca, P. Ma, I. Martiel, J. Mühle, S. Owada, F. Pamula, D. Sarabi, O. Tejero, C. J. Tsai, N. Varma, A. Wach, S. Boutet, K. Tono, P. Nogly, X. Deupi, S. Iwata, R. Neutze, J. Standfuss, G. Schertler, V. Panneels, Ultrafast structural changes direct the first molecular events of vision. *Nature* **615**, 939–944 (2023).
6. T. R. M. Barends, L. Foucar, A. Ardevol, K. Nass, A. Aquila, S. Botha, R. B. Doak, K. Falahati, E. Hartmann, M. Hilpert, M. Heinz, M. C. Hoffmann, J. Köfinger, J. E. Koglin, G. Kovacsova, M. Liang, D. Milathianaki, H. T. Lemke, J. Reinstein, C. M. Roome, R. L. Shoeman, G. J. Williams, I. Burghardt, G. Hummer, S. Boutet, I. Schlichting, Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350**, 445–450 (2015).
7. T. R. M. Barends, A. Gorel, S. Bhattacharyya, G. Schirò, C. Bacellar, C. Cirelli, J. P. Colletier, L. Foucar, M. L. Grünbein, E. Hartmann, M. Hilpert, J. M. Holton, P. J. M. Johnson, M. Kloos, G. Knopp, B. Marekha, K. Nass, G. Nass Kovacs, D. Ozerov, M. Stricker, M. Weik, R. B. Doak, R. L. Shoeman, C. J. Milne, M. Huix-Rotllant, M. Cammarata, I. Schlichting, Influence of pump laser fluence on ultrafast myoglobin structural dynamics. *Nature* **626**, 905–911 (2024).
8. V. Šrajer, Z. Ren, T. Y. Teng, M. Schmidt, T. Ursby, D. Bourgeois, C. Pradervand, W. Schildkamp, M. Wulff, K. Moffat, Protein conformational relaxation and ligand migration in myoglobin: A nanosecond to millisecond molecular movie from time-resolved Laue x-ray diffraction. *Biochemistry* **40**, 13802–13815 (2001).
9. S. Mous, G. Gotthard, D. Ehrenberg, S. Sen, T. Weinert, P. J. M. Johnson, D. James, K. Nass, A. Furrer, D. Kekilli, P. Ma, S. Brünle, C. M. Casadei, I. Martiel, F. Dworkowski, D. Gashi, P. Skopintsev, M. Wranik, G. Knopp, E. Panepucci, V. Panneels, C. Cirelli, D. Ozerov, G. F. X. Schertler, M. Wang, C. Milne, J. Standfuss, I. Schapiro, J. Heberle, P. Nogly, Dynamics and mechanism of a light-driven chloride pump. *Science* **375**, 845–851 (2022).
10. J. H. Yun, X. Li, J. Yue, J. H. Park, Z. Jin, C. Li, H. Hu, Y. Shi, S. Pandey, S. Carbajo, S. Boutet, M. S. Hunter, M. Liang, R. G. Sierra, T. J. Lane, L. Zhou, U. Weierstall, N. A. Zatsepin, M. Ohki, J. R. H. Tame, S. Y. Park, J. C. H. Spence, W. Zhang, M. Schmidt, W. Lee, H. Liu, Early-stage dynamics of chloride ion-pumping rhodopsin revealed by a femtosecond x-ray laser. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020486118 (2021).
11. B. R. Lee, K. I. White, M. Socolich, M. A. Klureza, R. Henning, V. Srajer, R. Ranganathan, D. R. Hekstra, Direct visualization of electric-field-stimulated ion conduction in a potassium channel. *Cell* **188**, 77–88.e15 (2025).
12. N. E. Christou, V. Apostolopoulou, D. V. M. Melo, M. Ruppert, A. Fadini, A. Henkel, J. Sprenger, D. Oberthuer, S. Günther, A. Pateras, A. Rahmani Mashhour, O. M. Yefanov, M. Galchenkova, P. Y. A. Reinke, V. Kremling, T. E. S. Scheer, E. R. Lange, P. Middendorf, R. Schubert, E. de Zitter, K. Lumbao-Conradson, J. Herrmann, S. Rahighi, A. Kunavar, E. V. Beale, J. H. Beale, C. Cirelli, P. J. M. Johnson, F. Dworkowski, D. Ozerov, Q. Bertrand, M. Wranik, C. Bacellar, S. Bajt, S. Wakatsuki, J. A. Sellberg, N. Huse, D. Turk, H. N. Chapman, T. J. Lane, Time-resolved crystallography captures light-driven DNA repair. *Science* **382**, 1015–1020 (2023).
13. S. Günther, P. Y. A. Reinke, Y. Fernández-García, J. Lieske, T. J. Lane, H. M. Ginn, F. H. M. Koua, C. Ehrt, W. Ewert, D. Oberthuer, O. Yefanov, S. Meier, K. Lorenzen, B. Krichel, J. D. Kopicki, L. Gelisio, W. Brehm, I. Dunkel, B. Seychell, H. Gieseler, B. Norton-Baker, B. Escudero-Pérez, M. Domaracky, S. Saouane, A. Tolstikova, T. A. White, A. Hänle, M. Groessler, H. Fleckenstein, F. Trost, M. Galchenkova, Y. Gevorkov, C. Li, S. Awel, A. Peck, M. Barthelmess, F. Schlünzen, P. Lourdu Xavier, N. Werner, H. Andaleeb, N. Ullah, S. Falke, V. Srinivasan, B. A. França, M. Schwinzer, H. Brognaro, C. Rogers, D. Melo, J. J. Zaitseva-Doyle, J. Knoska, G. E. Peña-Murillo, A. R. Mashhour, V. Hennicke, P. Fischer, J. Hakanpää, J. Meyer, P. Gribbon, B. Ellinger, M. Kuzikov, M. Wolf, A. R. Beccari, G. Bourenkov, D. von Stetten, G. Pompidor, I. Bento, S. Panneerselvam, I. Karpics, T. R. Schneider, M. M. Garcia-Alai, S. Niebling, C. Günther, C. Schmidt, R. Schubert, H. Han, J. Boger, D. C. F. Monteiro, L. Zhang, X. Sun, J. Pletzer-Zelgert, J. Wollenhaupt, C. G. Feiler, M. S. Weiss, E. C. Schulz, P. Mehrabi, K. Karničar, A. Usenik, J. Loboda, H. Tidow, A. Chari, R. Hilgenfeld, C. Uetrecht, R. Cox, A. Zaliani, T. Beck, M. Rarey, S. Günther, D. Turk, W. Hinrichs, H. N. Chapman, A. R. Pearson, C. Betzel, A. Meents, X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **372**, 642–646 (2021).
14. M. Schuller, G. J. Correy, S. Gahbauer, D. Fearon, T. Wu, R. E. Díaz, I. D. Young, L. Carvalho Martins, D. H. Smith, U. Schulze-Gahmen, T. W. Owens, I. Deshpande, G. E. Merz, A. C. Thwin, J. T. Biel, J. K. Peters, M. Moritz, N. Herrera, H. T. Kratochvil, QCRG Structural Biology Consortium, A. Aimon, J. M. Bennett, J. Brandao Neto, A. E. Cohen, A. Dias, A. Douangamath, L. Dunnett, O. Fedorov, M. P. Ferla, M. R. Fuchs, T. J. Gorrie-Stone, J. M. Holton, M. G. Johnson, T. Krojer, G. Meigs, A. J. Powell, J. G. M. Rack, V. L. Rangel, S. Russi, R. E. Skyner, C. A. Smith, A. S. Soares, J. L. Wierman, K. Zhu, P. O'Brien, N. Jura, A. Ashworth, J. J. Irwin, M. C. Thompson, J. E. Gestwicki, F. von Delft, B. K. Shoichet, J. S. Fraser, I. Ahel, Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **7**, eabf8711 (2021).
15. J. A. Newman, A. Douangamath, S. Yadzani, Y. Yosaatmadja, A. Aimon, J. Brandão-Neto, L. Dunnett, T. Gorrie-stone, R. Skyner, D. Fearon, M. Schapira, F. von Delft, O. Gileadi, Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nat. Commun.* **12**, 4848 (2021).
16. D. A. Keedy, Z. B. Hill, J. T. Biel, E. Kang, T. J. Rettenmaier, J. Brandão-Neto, N. M. Pearce, F. von Delft, J. A. Wells, J. S. Fraser, An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *eLife* **7**, e36307 (2018).
17. A. Creon, T. Emilie S. Scheer, P. Reinke, A. R. Mashhour, S. Günther, S. Niebling, K. Schamoni-Kast, C. Uetrecht, A. Meents, H. N. Chapman, J. Sprenger, T. J. Lane, Statistical crystallography reveals an allosteric network in SARS-CoV-2 Mpro. bioRxiv 2025.01.28.635305 [Preprint] (2025). https://doi.org/10.1101/2025.01.28.635305.
18. S. Shin, New era of synchrotron radiation: Fourth-generation storage ring. *AAPPS Bull.* **31**, 21 (2021).
19. P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F. J. Decker, Y. Ding, D. Dowell, S. Edstrom, A. Fisher, J. Frisch, S. Gilevich, J. Hastings, G. Hays, P. Hering, Z. Huang, R. Iverson, H. Loos, M. Messerschmidt, A. Miahnahri, S. Moeller, H. D. Nuhn, G. Pile, D. Ratner, J. Rzepiela, D. Schultz, T. Smith, P. Stefan, H. Tompkins, J. Turner, J. Welch, W. White, J. Wu, G. Yocky, J. Galayda, First lasing and operation of an ångstrom-wavelength free-electron laser. *Nat. Photonics* **4**, 641–647 (2010).
20. D. Reschke, D. Hamburg, Recent Progress with Eu-XFEL. *SFR, Canada, USA* **2015**, (2015).
21. J. L. Olmos Jr., S. Pandey, J. M. Martin-Garcia, G. Calvey, A. Katz, J. Knoska, C. Kupitz, M. S. Hunter, M. Liang, D. Oberthuer, O. Yefanov, M. Wiedorn, M. Heyman, M. Holl, K. Pande, A. Barty, M. D. Miller, S. Stern, S. Roy-Chowdhury, J. Coe, N. Nagaratnam, J. Zook, J. Verburgt, T. Norwood, I. Poudyal, D. Xu, J. Koglin, M. H. Seaberg, Y. Zhao, S. Bajt, T. Grant, V. Mariani, G. Nelson, G. Subramanian, E. Bae, R. Fromme, R. Fung, P. Schwander, M. Frank, T. A. White, U. Weierstall, N. Zatsepin, J. Spence, P. Fromme, H. N. Chapman, L. Pollack, L. Tremblay, A. Ourmazd, G. N. Phillips Jr., M. Schmidt, Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography. *BMC Biol.* **16**, 59 (2018).

22. M. Schmidt, Mix and inject: Reaction initiation by diffusion for time-resolved macromolecular crystallography. *Adv. Condens. Matter Phys.* **2013**, doi.org/10.1155/2013/167276 (2013).
23. G. D. Calvey, A. M. Katz, L. Pollack, Microfluidic mixing injector holder enables routine structural enzymology measurements with mix-and-inject serial crystallography using x-ray free electron lasers. *Anal. Chem.* **91**, 7139–7144 (2019).
24. S. Pandey, G. Calvey, A. M. Katz, T. N. Malla, F. H. M. Koua, J. M. Martin-Garcia, I. Poudyal, J. H. Yang, M. Vakili, O. Yefanov, K. A. Zielinski, S. Bajt, S. Awel, K. Doerner, M. Frank, L. Gelisio, R. Jernigan, H. Kirkwood, M. Kloos, J. Koliyadu, V. Mariani, M. D. Miller, G. Mills, G. Nelson, J. L. Olmos Jr., A. Sadri, T. Sato, A. Tolstikova, W. Xu, A. Ourmazd, J. C. H. Spence, P. Schwander, A. Barty, H. N. Chapman, P. Fromme, A. P. Mancuso, G. N. Phillips Jr., R. Bean, L. Pollack, M. Schmidt, Observation of substrate diffusion and ligand binding in enzyme crystals using high-repetition-rate mix-and-inject serial crystallography. *IUCrJ* **8**, 878–895 (2021).
25. J. R. Stagno, Y. Liu, Y. R. Bhandari, C. E. Conrad, S. Panja, M. Swain, L. Fan, G. Nelson, C. Li, D. R. Wendel, T. A. White, J. D. Coe, M. O. Wiedorn, J. Knoska, D. Oberthuer, R. A. Tuckey, P. Yu, M. Dyba, S. G. Tarasov, U. Weierstall, T. D. Grant, C. D. Schwieters, J. Zhang, A. R. Ferré-D'Amaré, P. Fromme, D. E. Draper, M. Liang, M. S. Hunter, S. Boutet, K. Tan, X. Zuo, X. Ji, A. Barty, N. A. Zatsepin, H. N. Chapman, J. C. H. Spence, S. A. Woodson, Y.-X. Wang, Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541**, 242–246 (2017).
26. A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandão-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L. Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keeley, G. M. Keserü, H. F. Klein, M. P. Martin, M. E. M. Noble, P. O'Brien, A. Powell, R. N. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft, M. A. Walsh, Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **11**, 5047 (2020).
27. T. S. Mehlman, J. T. Biel, S. M. Azeem, E. R. Nelson, S. Hossain, L. Dunnett, N. G. Paterson, A. Douangamath, R. Talon, D. Axford, H. Orins, F. von Delft, D. A. Keedy, Room-temperature crystallography reveals altered binding of small-molecule fragments to PTP1B. *eLife* **12**, e84632 (2023).
28. A. M. Wolff, E. Nango, I. D. Young, A. S. Brewster, M. Kubo, T. Nomura, M. Sugahara, S. Owada, B. A. Barad, K. Ito, A. Bhowmick, S. Carbajo, T. Hino, J. M. Holton, D. Im, L. J. O'Riordan, T. Tanaka, R. Tanaka, R. G. Sierra, F. Yumoto, K. Tono, S. Iwata, N. K. Sauter, J. S. Fraser, M. C. Thompson, Mapping protein dynamics at high spatial resolution with temperature-jump x-ray crystallography. *Nat. Chem.* **15**, 1549–1558 (2023).
29. D. R. Hekstra, K. I. White, M. A. Socolich, R. W. Henning, V. Šrajer, R. Ranganathan, Electric-field-stimulated protein mechanics. *Nature* **540**, 400–405 (2016).
30. D. A. Keedy, L. R. Kenner, M. Warkentin, R. A. Woldeyes, J. B. Hopkins, M. C. Thompson, A. S. Brewster, A. H. van Benschoten, E. L. Baxter, M. Uervirojnangkoorn, S. E. McPhillips, J. Song, R. Alonso-Mori, J. M. Holton, W. I. Weis, A. T. Brunger, S. M. Soltis, H. Lemke, A. Gonzalez, N. K. Sauter, A. E. Cohen, H. van den Bedem, R. E. Thorne, J. S. Fraser, Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife* **4**, e07574 (2015).
31. P. Bryant, F. Noé, Structure prediction of alternative protein conformations. *Nat. Commun.* **15**, 7328 (2024).
32. J. B. Sumner, The isolation and crystallization of the enzyme urease: Preliminary paper. *J. Biol. Chem.* **69**, 435–441 (1926).
33. P. J. Kasvinsky, N. B. Madsen, Activity of glycogen phosphorylase in the crystalline state. *J. Biol. Chem.* **251**, 6852–6859 (1976).
34. I. Schlichting, S. C. Almo, G. Rapp, K. Wilson, K. Petratos, A. Lentfer, A. Wittinghofer, W. Kabsch, E. F. Pai, G. A. Petsko, R. S. Goody, Time-resolved x-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature* **345**, 309–315 (1990).
35. A. L. Fink, D. Kar, R. Kotin, Ribonuclease structure and catalysis: Effects of crystalline enzyme, alcohol cryosolvents, low temperatures, and product inhibition. *Biochemistry* **26**, 8571–8579 (1987).
36. S. K. Whittier, A. C. Hengge, J. P. Loria, Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. *Science* **341**, 899–903 (2013).
37. J. F. Acheson, L. J. Bailey, T. C. Brunold, B. G. Fox, In-crystal reaction cycle of a toluene-bound diiron hydroxylase. *Nature* **544**, 191–195 (2017).
38. A. B. Law, E. J. Fuentes, A. L. Lee, Conservation of side-chain dynamics within a protein family. *J. Am. Chem. Soc.* **131**, 6322–6323 (2009).
39. F. A. Quiocho, C. H. Mcmurray, W. N. Lipscomb, Similarities between the conformation of arsanilazotyrosine 248 of carboxypeptidase A α in the crystalline state and in solution. *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2850–2854 (1972).
40. D. Sorigué, K. Hadjidemetriou, S. Blangy, G. Gotthard, A. Bonvalet, N. Coquelle, P. Samire, A. Aleksandrov, L. Antonucci, A. Benachir, S. Boutet, M. Byrdin, M. Cammarata, S. Carbajo, S. Cuiné, R. B. Doak, L. Foucar, A. Gorel, M. Grünbein, E. Hartmann, R. Hienerwadel, M. Hilpert, M. Kloos, T. J. Lane, B. Légeret, P. Legrand, Y. Li-Beisson, S. L. Y. Moulin, D. Nurizzo, G. Peltier, G. Schirò, R. L. Shoeman, M. Sliwa, X. Solinas, B. Zhuang, T. R. M. Barends, J.-P. Colletier, M. Joffre, A. Royant, C. Berthomieu, M. Weik, T. Domratcheva, K. Brettel, M. H. Vos, I. Schlichting, P. Arnoux, P. Müller, F. Beisson, Mechanism and dynamics of fatty acid photodecarboxylase. *Science* **372**, eabd5687 (2021).
41. C. Kupitz, J. L. Olmos Jr., M. Holl, L. Tremblay, K. Pande, S. Pandey, D. Oberthür, M. Hunter, M. Liang, A. Aquila, J. Tenboer, G. Calvey, A. Katz, Y. Chen, M. O. Wiedorn, J. Knoska, A. Meents, V. Majriani, T. Norwood, I. Poudyal, T. Grant, M. D. Miller, W. Xu, A. Tolstikova, A. Morgan, M. Metz, J. M. Martin-Garcia, J. D. Zook, S. Roy-Chowdhury, J. Coe, N. Nagaratnam, D. Meza, R. Fromme, S. Basu, M. Frank, T. White, A. Barty, S. Bajt, O. Yefanov, H. N. Chapman, N. Zatsepin, G. Nelson, U. Weierstall, J. Spence, P. Schwander, L. Pollack, P. Fromme, A. Ourmazd, G. N. Phillips Jr., M. Schmidt, Structural enzymology using x-ray free electron lasers. *Struct. Dyn.* **4**, 044003 (2017).
42. M. Wilamowski, D. A. Sherrell, Y. Kim, A. Lavens, R. W. Henning, K. Lazarski, A. Shigemoto, M. Endres, N. Maltseva, G. Babnigg, S. C. Burdette, V. Srajer, A. Joachimiak, Time-resolved beta-lactam cleavage by L1 metallo-beta-lactamase. *Nat. Commun.* **13**, 7379 (2022).
43. K. A. Zielinski, C. Dolamore, K. M. Dalton, N. Smith, J. Termini, R. Henning, V. Srajer, D. R. Hekstra, L. Pollack, M. A. Wilson, Resolving DJ-1 glyoxalase catalysis using mix-and-inject serial crystallography at a synchrotron. bioRxiv 2024.07.19.604369 [Preprint] (2024). https://doi.org/10.1101/2024.07.19.604369.
44. P. Nogly, T. Weinert, D. James, S. Carbajo, D. Ozerov, A. Furrer, D. Gashi, V. Borin, P. Skopintsev, K. Jaeger, K. Nass, P. Båth, R. Bosman, J. Koglin, M. Seaberg, T. Lane, D. Kekilli, S. Brünle, T. Tanaka, W. Wu, C. Milne, T. White, A. Barty, U. Weierstall, V. Panneels, E. Nango, S. Iwata, M. Hunter, I. Schapiro, G. Schertler, R. Neutze, J. Standfuss, Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* **361**, eaat0094 (2018).
45. J. Kern, R. Chatterjee, I. D. Young, F. D. Fuller, L. Lassalle, M. Ibrahim, S. Gul, T. Fransson, A. S. Brewster, R. Alonso-Mori, R. Hussein, M. Zhang, L. Douthit, C. de Lichtenberg, M. H. Cheah, D. Shevela, J. Wersig, I. Seuffert, D. Sokaras, E. Pastor, C. Weninger, T. Kroll, R. G. Sierra, P. Aller, A. Butryn, A. M. Orville, M. Liang, A. Batyuk, J. E. Koglin, S. Carbajo, S. Boutet, N. W. Moriarty, J. M. Holton, H. Dobbek, P. D. Adams, U. Bergmann, N. K. Sauter, A. Zouni, J. Messinger, J. Yano, V. K. Yachandra, Structures of the intermediates of Kok's photosynthetic water oxidation clock. *Nature* **563**, 421–425 (2018).
46. A. Bhowmick, R. Hussein, I. Bogacz, P. S. Simon, M. Ibrahim, R. Chatterjee, M. D. Doyle, M. H. Cheah, T. Fransson, P. Chernev, I. S. Kim, H. Makita, M. Dasgupta, C. J. Kaminsky, M. Zhang, J. Gätcke, S. Haupt, I. I. Nangca, S. M. Keable, A. O. Aydin, K. Tono, S. Owada, L. B. Gee, F. D. Fuller, A. Batyuk, R. Alonso-Mori, J. M. Holton, D. W. Paley, N. W. Moriarty, F. Mamedov, P. D. Adams, A. S. Brewster, H. Dobbek, N. K. Sauter, U. Bergmann, A. Zouni, J. Messinger, J. Kern, J. Yano, V. K. Yachandra, Structural evidence for intermediates during O_2 formation in photosystem II. *Nature* **617**, 629–636 (2023).
47. T. Tosha, T. Nomura, T. Nishida, N. Saeki, K. Okubayashi, R. Yamagiwa, M. Sugahara, T. Nakane, K. Yamashita, K. Hirata, G. Ueno, T. Kimura, T. Hisano, K. Muramoto, H. Sawai, H. Takeda, E. Mizohata, A. Yamashita, Y. Kanematsu, Y. Takano, E. Nango, R. Tanaka, O. Nureki, O. Shoji, Y. Ikemoto, H. Murakami, S. Owada, K. Tono, M. Yabashi, M. Yamamoto, H. Ago, S. Iwata, H. Sugimoto, Y. Shiro, M. Kubo, Capturing an initial intermediate during the P450nor enzymatic reaction using time-resolved XFEL crystallography and caged-substrate. *Nat. Commun.* **8**, 1585 (2017).
48. J. Tenboer, S. Basu, N. Zatsepin, K. Pande, D. Milathianaki, M. Frank, M. Hunter, S. Boutet, G. J. Williams, J. E. Koglin, D. Oberthuer, M. Heymann, C. Kupitz, C. Conrad, J. Coe, S. Roy-Chowdhury, U. Weierstall, D. James, D. Wang, T. Grant, A. Barty, O. Yefanov, J. Scales, C. Gati, C. Seuring, V. Srajer, R. Henning, P. Schwander, R. Fromme, A. Ourmazd, K. Moffat, J. J. van Thor, J. C. H. Spence, P. Fromme, H. N. Chapman, M. Schmidt, Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242–1246 (2014).
49. K. Pande, C. D. M. Hutchison, G. Groenhof, A. Aquila, J. S. Robinson, J. Tenboer, S. Basu, S. Boutet, D. P. DePonte, M. Liang, T. A. White, N. A. Zatsepin, O. Yefanov, D. Morozov, D. Oberthuer, C. Gati, G. Subramanian, D. James, Y. Zhao, J. Koralek, J. Brayshaw, C. Kupitz, C. Conrad, S. Roy-Chowdhury, J. D. Coe, M. Metz, P. L. Xavier, T. D. Grant, J. E. Koglin, G. Ketawala, R. Fromme, V. Šrajer, R. Henning, J. C. H. Spence, A. Ourmazd, P. Schwander, U. Weierstall, M. Frank, P. Fromme, A. Barty, H. N. Chapman, K. Moffat, J. J. van Thor, M. Schmidt, Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725–729 (2016).
50. H. Ihee, S. Rajagopal, V. Šrajer, R. Pahl, S. Anderson, M. Schmidt, F. Schotte, P. A. Anfinrud, M. Wulff, K. Moffat, Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7145–7150 (2005).
51. U. K. Genick, G. E. O. Borgstahl, K. Ng, Z. Ren, C. Pradervand, P. M. Burke, V. Šrajer, T. Y. Teng, W. Schildkamp, D. E. McRee, K. Moffat, E. D. Getzoff, Structure of a protein photocycle intermediate by millisecond time-resolved crystallography. *Science* **275**, 1471–1475 (1997).
52. U. K. Genick, Structure-factor extrapolation using the scalar approximation: Theory, applications and limitations. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 1029–1041 (2007).
53. E. De Zitter, N. Coquelle, P. Oeser, T. R. M. Barends, J.-P. Colletier, Xtrapol8 enables automatic elucidation of low-occupancy intermediate-states in crystallographic studies. *Commun. Biol.* **5**, 640 (2022).

54. K. M. Dalton, J. B. Greisman, D. R. Hekstra, A unifying Bayesian framework for merging x-ray diffraction data. *Nat. Commun.* **13**, 7764 (2022).
55. H. Mai, A. Peck, K. M. Dalton, L. S. de Moraes, J. E. Burch, F. Poitevin, H. M. Nelson, Assessing the applicability of Bayesian inference for merging small molecule microED data. ChemRxiv chemrxiv-2024-62bmk [Preprint] (2024). https://doi.org/10.26434/chemrxiv-2024-62bmk.
56. K. A. Zielinski, C. Dolamore, H. K. Wang, R. W. Henning, M. A. Wilson, L. Pollack, V. Srajer, D. R. Hekstra, K. M. Dalton, Scaling and merging time-resolved pink-beam diffraction with variational inference. *Struct. Dyn.* **11**, 064301 (2024).
57. A. J. C. Wilson, The probability distribution of x-ray intensities. *Acta Crystallogr.* **2**, 318–321 (1949).
58. R. J. Read, Structure-factor probabilities for related structures. *Acta Crystallogr. A* **46**, 900–912 (1990).
59. G. Bricogne, in *Methods in Enzymology* (Academic Press, 1997), vol. 276, pp. 361–423.
60. J. B. Greisman, K. M. Dalton, C. J. Sheehan, M. A. Klureza, I. Kurinov, D. R. Hekstra, Native SAD phasing at room temperature. *Acta. Crystallogr. D Struct. Biol.* **78**, 986–996 (2022).
61. H. Hauptman, On integrating the techniques of direct methods and isomorphous replacement. I. The theoretical basis. *Acta Cryst.* **38**, 289–294 (1982).
62. K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, 2012).
63. L. A. Aldama, K. M. Dalton, D. R. Hekstra, Correcting systematic errors in diffraction data with modern scaling algorithms. *Acta Crystallogr. D Struct. Biol.* **79**, 796–805 (2023).
64. P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
65. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
66. R. A. Hewitt, K. M. Dalton, D. A. Mendez, H. K. Wang, M. A. Klureza, D. E. Brookner, J. B. Greisman, D. McDonagh, V. Šrajer, N. K. Sauter, A. S. Brewster, D. R. Hekstra, Laue-DIALS: Open-source software for polychromatic x-ray diffraction data. *Struct. Dyn.* **11**, 054701 (2024).
67. T. Ursby, D. Bourgeois, Improved estimation of structure-factor difference amplitudesfrom poorly accurate data. *Acta Cryst.* **53**, 564–575 (1997).
68. J. Kern, R. Tran, R. Alonso-Mori, S. Koroidov, N. Echols, J. Hattne, M. Ibrahim, S. Gul, H. Laksmono, R. G. Sierra, R. J. Gildea, G. Han, J. Hellmich, B. Lassalle-Kaiser, R. Chatterjee, A. S. Brewster, C. A. Stan, C. Glöckner, A. Lampe, D. DiFiore, D. Milathianaki, A. R. Fry, M. M. Seibert, J. E. Koglin, E. Gallo, J. Uhlig, D. Sokaras, T. C. Weng, P. H. Zwart, D. E. Skinner, M. J. Bogan, M. Messerschmidt, P. Glatzel, G. J. Williams, S. Boutet, P. D. Adams, A. Zouni, J. Messinger, N. K. Sauter, U. Bergmann, J. Yano, V. K. Yachandra, Taking snapshots of photosynthetic water oxidation using femtosecond x-ray diffraction and spectroscopy. *Nat. Commun.* **5**, 4371 (2014).
69. A. S. Brewster, D. G. Waterman, J. M. Parkhurst, R. J. Gildea, I. D. Young, L. J. O'Riordan, J. Yano, G. Winter, G. Evans, N. K. Sauter, Improving signal strength in serial crystallography with DIALS geometry refinement. *Acta. Crystallogr. D Struct. Biol.* **74**, 877–894 (2018).
70. A. S. Brewster, D. G. Waterman, J. M. Parkhurst, R. J. Gildea, T. M. Michels-Clark, I. D. Young, H. J. Bernstein, G. Winter, G. Evans, N. K. Sauter, Processing XFEL data with cctbx. xfel and DIALS. *Comput. Crystallogr. Newsl.* **7**, 32–53 (2016).
71. M. Uervirojnangkoorn, O. B. Zeldin, A. Y. Lyubimov, J. Hattne, A. S. Brewster, N. K. Sauter, A. T. Brunger, W. I. Weis, Enabling x-ray free electron laser crystallography for challenging biological systems from a limited number of crystals. *eLife* **4**, e05421 (2015).
72. N. M. Pearce, T. Krojer, A. R. Bradley, P. Collins, R. P. Nowak, R. Talon, B. D. Marsden, S. Kelm, J. Shi, C. M. Deane, F. von Delft, A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **8**, 15123 (2017).
73. careless-examples; https://github.com/rs-station/careless-examples.
74. dw-examples; https://github.com/Hekstra-Lab/dw-examples/.
75. abismal; https://github.com/rs-station/abismal.
76. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
77. R. J. Read, New ways of looking at experimental phasing. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1891–1902 (2003).
78. M. A. Gilles, A. Singer, A molecular prior distribution for Bayesian inference based on Wilson statistics. *Comput. Methods Programs Biomed.* **221**, 106830 (2022).
79. A. Singer, Wilson statistics: Derivation, generalization and applications to electron cryomicroscopy. *Acta Crystallogr. A Found. Adv.* **77**, 472–479 (2021).
80. G. Huldt, A. Szoke, J. Hajdu, Diffraction imaging of single particles and biomolecules. *J. Struct. Biol.* **144**, 219–227 (2003).
81. H. N. Chapman, O. M. Yefanov, K. Ayyer, T. A. White, A. Barty, A. Morgan, V. Mariani, D. Oberthuer, K. Pande, Continuous diffraction of molecules and disordered molecular crystals. *J. Appl. Cryst.* **50**, 1084–1103 (2017).
82. D. M. Blei, A. Kucukelbir, J. D. McAuliffe, Variational inference: A review for statisticians. *J. Am. Stat. Assoc.* **112**, 859–877 (2017).
83. M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, L. K. Saul, An introduction to variational methods for graphical models. *Mach. Learn.* **37**, 183–233 (1999).
84. D. P. Kingma, M. Welling, Auto-encoding variational bayes. arXiv:1312.6114 [stat.ML] (2013).
85. J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, R. A. Saurous, TensorFlow distributions. arXiv:1711.10604 [cs.LG] (2017).
86. M. Schmidt, V. Srajer, R. Henning, H. Ihee, N. Purwar, J. Tenboer, S. Tripathi, Protein energy landscapes determined by five-dimensional crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 2534–2542 (2013).
87. M. Wojdyr, GEMMI: A library for structural biology. *J. Open Source Softw.* **7**, 4200 (2022).
88. J. B. Greisman, K. M. Dalton, D. R. Hekstra, Reciprocalspaceship: A Python library for crystallographic data analysis. *J. Appl. Cryst.* **54**, 1521–1529 (2021).
89. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
90. G. E. O. Borgstahl, D. R. Williams, E. D. Getzoff, 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: Unusual fold, active site, and chromophore. *Biochemistry* **34**, 6278–6287 (1995).
91. W. L. DeLano, PyMOL: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr* **40**, 82–92 (2002).
92. K. Takeda, H. Miyatake, S. Y. Park, M. Kawamoto, N. Kamiya, K. Miki, Multi-wavelength anomalous diffraction method for I and Xe atoms using ultra-high-energy x-rays from SPring-8. *J. Appl. Cryst.* **37**, 925–933 (2004).
93. D. Schwarzenbach, H. D. Flack, On the definition and practical use of crystal-based azimuthal angles. *J. Appl. Cryst.* **22**, 601–605 (1989).
94. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
95. G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter, G. Evans, DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. D Struct. Biol.* **74**, 85–97 (2018).
96. J. D. Hunter, Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
97. C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with NumPy. *Nature* **585**, 357–362 (2020).
98. W. McKinney, pandas: A foundational Python library for data analysis and statistics. *Python High Perform. Sci. Comput.* **14**, 1–9 (2011).
99. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. V. Plas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
100. M. L. Waskom, Seaborn: Statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
101. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, "TensorFlow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (USENIX Association, 2016), pp. 265–283.
102. H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook, C. Zardecki, The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 899–907 (2002).
103. B. W. Matthews, E. W. Czerwinski, Local scaling: A method to reduce systematic errors in isomorphous replacement and anomalous scattering measurements. *Acta Crystallogr. A* **31**, 480–487 (1975).
104. P. Howell, G. Smith, Identification of heavy-atom derivatives by normal probability methods. *J. Appl. Cryst.* **25**, 81–86 (1992).
105. A. Gelman, J. Hullman, C. Wlezien, G. E. Morris, Information, incentives, and goals in election forecasts. *Judgm. Decis. Mak.* **15**, 863–880 (2020).
106. S. L. Lohr, J. M. Brick, Roosevelt predicted to win: Revisiting the 1936 Literary Digest poll. *Stat. Politics Policy* **8**, 65–84 (2017).
107. C. Kennedy, M. Blumenthal, S. Clement, J. D. Clinton, C. Durand, C. Franklin, K. McGeeney, L. Miringoff, K. Olson, D. Rivers, L. Saad, G. E. Witt, C. Wlezien, An evaluation of the 2016 election polls in the United States. *Public Opin. Q.* **82**, 1–33 (2018).
108. T. C. Terwilliger, J. Berendzen, Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861 (1999).

109. H. Cao, J. Skolnick, Time-resolved x-ray crystallography capture of a slow reaction tetrahydrofolate intermediate. *Struct. Dyn.* **6**, 024701 (2019).
110. B. Rupp, *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology* (Garland Science, ed. 1, 2010), pp. 1–809; https://doi.org/10.1201/9780429258756.
111. H. A. Hauptman, The phase problem of x-ray crystallography. *Rep. Prog. Phys.* **54**, 1427–1454 (1991).
112. R. J. Read, Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **42**, 140–149 (1986).
113. A. B. Yoo, M. A. Jette, M. Grondona, in *Workshop on Job Scheduling Strategies for Parallel Processing* (Springer, 2003), pp. 44–60.
114. N. S. Pannu, A. J. McCoy, R. J. Read, Application of the complex multivariate normal distribution to crystallographic methods with insights into multiple isomorphous replacement phasing. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1801–1808 (2003).
115. Z. Otwinowski, W. Minor, "Denzo and scalepack," in *International Tables for Crystallography* (International Union of Crystallography, 2006), pp. 226–235.
116. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
117. A. N. Popov, G. P. Bourenkov, Choice of data-collection parameters based on statistic modelling. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1145–1153 (2003).
118. J. B. Greisman, K. M. Dalton, D. E. Brookner, M. A. Klureza, C. J. Sheehan, I. S. Kim, R. W. Henning, S. Russi, D. R. Hekstra, Perturbative diffraction methods resolve a conformational switch that facilitates a two-step enzymatic mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2313192121 (2024).
119. I. I. Vorontsov, P. Coppens, On the refinement of time-resolved diffraction data: Comparison of the random-distribution and cluster-formation models and analysis of the light-induced increase in the atomic displacement parameters. *J. Synchrotron Radiat.* **12**, 488–493 (2005).
120. G. Sim, The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal heavy-atom method for non-centrosymmetrical structures. *Acta Crystallogr.* **12**, 813–815 (1959).
121. S. Russi, A. González, L. R. Kenner, D. A. Keedy, J. S. Fraser, H. van den Bedem, Conformational variation of proteins at room temperature is not dominated by radiation damage. *J. Synchrotron Radiat.* **24**, 73–82 (2017).
122. J. A. Arpino, P. J. Rizkallah, D. D. Jones, Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222. *PLOS One* **7**, e47132 (2012).
123. A. Royant, M. Noirclerc-Savoye, Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein. *J. Struct. Biol.* **174**, 385–390 (2011).
124. E. D. Getzoff, K. N. Gutwin, U. K. Genick, Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation. *Nat. Struct. Mol. Biol.* **10**, 663–668 (2003).
125. U. K. Genick, S. M. Soltis, P. Kuhn, I. L. Canestrelli, E. D. Getzoff, Structure at 0.85 Å resolution of an early protein photocycle intermediate. *Nature* **392**, 206–209 (1998).

Acknowledgments: We thank M. Socolich (University of Chicago) for sodium iodide–soaked lysozyme crystals; V. Šrajer (University of Chicago) and M. Schmidt (University of Wisconsin) for providing Laue time-resolved data; G. Correy and J. S. Fraser (UCSF) for unmerged Mac1 drug fragment screening data; and S. Berry, S. Liu, T. J. Lane, V. Šrajer, and R. Read for comments on the manuscript. We thank the staff at the NE-CAT, beamline 24-ID-C of the Advanced Photon Source (APS), particularly I. Kourinov, for assistance with data collection. We thank the staff at the BioCARS facility at beamline 14-ID-B of APS, particularly R. Henning and I.-S. Kim, for assistance with polychromatic data collection. We thank members of the Hekstra laboratory for discussion. We acknowledge the open-source software projects used in this work including DIALS (*95*) GEMMI (*87*), Matplotlib (*96*), NumPy (*97*), Pandas (*98*), reciprocalspaceship (*88*), SciPy (*99*), Seaborn (*100*), TensorFlow (*101*), TensorFlow Probability (*85*), and Careless. This publication's contents are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS, the NIH, or the DOE. **Funding:** NE-CAT beamlines are supported by the National Institute of General Medical Sciences (NIGMS), NIH (P30 GM124165). BioCARS is supported by NIGMS under grant number P41 GM118217. The time-resolved setup at Sector 14 was funded in part through a collaboration with P. Anfinrud (NIH/National Institute of Diabetes and Digestive and Kidney Diseases). APS, a US Department of Energy (DOE) Office of Science User Facility, is operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. This work was also supported by National Institutes of Health (NIH) grant DP2-GM141000 (to D.R.H.), National Science Foundation Graduate Research Fellowship grant DGE2140743 (to H.K.W.), Burroughs Wellcome Fund Career Award at the Scientific Interface (to K.M.D.), and DOE, Office of Science, Basic Energy Sciences under contract no. DEAC02-76SF00515 (to K.M.D.). **Author contributions:** Conceptualization: D.R.H. and K.M.D. Methodology: D.R.H., M.A.K., J.B.G., and K.M.D. Software: D.R.H., H.K.W., J.B.G., and K.M.D. Validation: D.R.H., H.K.W., and K.M.D. Formal analysis: D.R.H., H.K.W., and K.M.D. Investigation: D.R.H., H.K.W., M.A.K., and K.M.D. Resources: D.R.H. and J.B.G. Data curation: D.R.H., H.K.W., and K.M.D. Writing—original draft: D.R.H., H.K.W., and K.M.D. Writing—review and editing: D.R.H., H.K.W., M.A.K., J.B.G., and K.M.D. Visualization: D.R.H., H.K.W., and K.M.D. Supervision: D.R.H. and K.M.D. Project administration: D.R.H. and K.M.D. Funding acquisition: D.R.H. and K.M.D. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Supplementary notebooks, including analyses used for generating Fig. 3 and figs. S2 to S5, S11, S13, and S16 to S18, are available as deposited in Zenodo under accession code 10.5281/zenodo.11099738 at https://zenodo.org/records/11099738, as well as in a GitHub repository at https://github.com/Hekstra-Lab/dw. The monochromatic sodium iodide–soaked lysozyme structure presented here has been deposited in the PDB (*102*) with ID 9B7C. The polychromatic sodium iodide–soaked lysozyme diffraction images have been deposited in SBGrid (https://data.sbgrid.org/dataset/1118/). All unmerged integrated intensities, merged structure factor amplitudes, cross-validation data, and analysis used to generate Figs. 2 and 4 to 7 and figs. S1, S6 to S10, S12, S14, and S15 have been deposited in Zenodo under accession code 10.5281/zenodo.11099738 at https://zenodo.org/records/11099738. These analyses, without merged structure factors, are additionally included in a GitHub repository at https://github.com/Hekstra-Lab/dw-examples. Conda environments for data processing can be reproduced using .yaml files in the respective Zenodo depositions and repositories. The source code used for generating all figures and tables are available as deposited in Zenodo at https://zenodo.org/records/11099738 and in the GitHub repositories https://github.com/Hekstra-Lab/dw and https://github.com/Hekstra-Lab/dw-examples. The multivariate prior scaling model is implemented in the Careless Python package, available in the GitHub repository https://github.com/rs-station/careless and as deposited in Zenodo at https://zenodo.org/records/11099738.

Submitted 10 April 2025
Accepted 31 October 2025
Published 3 December 2025
10.1126/sciadv.adj2921