

Bayesian multi-state multi-condition modeling of a protein structure from X-ray crystallography

Matthew Hancock^{a,c,1}, James S Fraser^a, Paul D Adams^a, and Andrej Sali^{b,1,2}

This manuscript was compiled on June 26, 2024

A model of a protein structure at atomic resolution is key to rationalizing and predicting its biological function. Many such models are computed from a diffraction pattern from X-ray crystallography. Despite the protein crystal containing billions of protein molecules that independently sample the energy landscape during data collection, most models computed from X-ray data depict a single set of atomic coordinates. A model with multiple sets of atomic coordinates (multi-state) may improve the satisfaction of the X-ray data and is a more accurate, precise, and informative depiction of the protein. However, computing a multi-state model is challenging on account of a low data-to-parameter ratio. X-ray datasets collected for the same system under distinct experimental conditions (eg, temperature) may provide additional observations, thereby improving the data-to-parameter ratio. Here, we develop, benchmark, and illustrate MultiXray: Bayesian multi-state multi-condition modeling for X-ray crystallography. The input information is J X-ray datasets collected under distinct conditions and a molecular mechanics force field. The model representation is N independent atomic models of the protein structure (states) and the weight of each state under each condition. A Bayesian posterior model density quantifies the match of the model with all X-ray datasets and molecular mechanics. A sample of models is drawn from the posterior model density using molecular dynamics simulations. We benchmark MultiXray on simulated X-ray data and analyze the impact of additional states and conditions on the scoring function and model search. We illustrate MultiXray on temperature-dependent X-ray datasets collected for SARS-CoV-2 M^{pro} and compute multi-state multi-condition models that improve the R^{free} relative to the PDB model by up to 0.05. MultiXray is implemented in our open-source *Integrative Modeling Platform*, relying on integration with *Phenix*, thus making it easily applicable to other systems.

X-ray crystallography | Multi-state modeling | Bayesian Inference | Protein structure

A protein crystal is a heterogeneous mix of structural states. X-ray crystallography is an important experimental technique for obtaining structural models of proteins at atomic resolution. In X-ray crystallography, the protein crystal contains between $10^6 - 10^{15}$ copies of the protein (1–3). Due to the high solvent content of the crystal, protein molecules within the crystal can fluctuate nearly independently throughout data collection and adopt distinct structural states based on the energy landscape (4–6). Despite the resulting structural heterogeneity, a majority of models computed from X-ray datasets describe a single structural state with Gaussian isotropic B-factors (7). Fitting a B-factor to an X-ray dataset convolutes all sources of experimental uncertainty, including the heterogeneous mix of structural states, random thermal motion, and long-range crystallographic disorder (8, 9). The inability of a B-factor to describe the heterogeneous mix of structural states found within the protein crystal contributes to the inability of protein models to satisfy an X-ray dataset within its theoretical uncertainty (10, 11). A model that depicts anharmonic conformational substates found in a protein crystal will improve the satisfaction of X-ray data and reveal the structural basis of important biological properties at atomic resolution, such as allosteric networks or hidden cavities for small molecule binding (cryptic pockets) (12, 13).

Approaches to modeling a heterogenous mix of structural states. There are 2 approaches for computing models that depict multiple structural states from X-ray data. First, conformational substates can be captured by computing single-state models that independently satisfy the X-ray dataset, as is seen in the modeling of Nuclear Magnetic Resonance spectra (14). For example, phenix.ensemble_refinement computes an ensemble from snapshots of a molecular simulation restrained by a time-averaged X-ray target function (15). Such approaches may not find weakly occupied states in a reasonable amount of computation time, as they are dependent on

Significance Statement

Authors must submit a 120-word maximum statement about the significance of their research paper written at a level understandable to an undergraduate educated scientist outside their field of specialty. The primary goal of the significance statement is to explain the relevance of the work in broad context to a broad readership. The significance statement appears in the paper itself and is required for all research papers.

Author affiliations: ^aAffiliation One; ^bAffiliation Two; ^cAffiliation Three

Please provide details of author contributions here.

Please declare any competing interests here.

²To whom correspondence should be addressed. E-mail: salis@ilab.org

overcoming potentially large barriers in the scoring function. Alternatively, a model can depict conformational substates by introducing additional structural variables (3). All variables are then collectively fit against the X-ray data. The extent and detail of the additional degrees of freedom depend on the available computational power and data quality (16). For example, qFit-3 avoids introducing excessive structural parameters by representing side chains as ensembles of one or more rotameric states (17). Methods that refine multiple fully parameterized atomic models have been limited to systems that diffract to ultra-high resolution (6, 18–21).

Multi-condition crystallography. Often, multiple X-ray datasets are collected for the same system under distinct experimental conditions. One example is multi-temperature crystallography, where data collection is performed at temperatures from cryogenic to near-physiological (22). Data collection at higher temperatures has been shown to dramatically modify protein dynamics within the protein crystal (23–25). More recently, models computed from higher temperature datasets have revealed a fuller set of structural states within the protein’s energy landscape at atomic resolution (26–28). Comparisons of models of the same system computed at distinct temperatures show similar structural states but shifted thermodynamic equilibrium (26, 29, 30). Therefore, multiple X-ray datasets under distinct conditions containing mutual structural information could increase the data-to-parameter ratio and inform a more accurate, precise, and complete multi-state model.

Computing a multi-state multi-condition model. Here, we seek to compute a multi-state model from multiple X-ray diffraction patterns collected under distinct conditions (1a). We can fit a multi-state model using multiple X-ray datasets without needing ultra-high-resolution X-ray data. We accommodate for the thermodynamic shift of distinct experimental conditions by computing each state’s weight under each condition. In other words, we assume the sets of structural states are consistent across all experimental conditions with varied weights, allowing a single multi-state model to be informed by all X-ray datasets and massively boosting the data-to-parameter ratio. The assumption of a consistent set of structural states under all conditions is not limiting because the structural states may include sparsely populated or completely unpopulated states. We formulate a Bayesian posterior model density for the multi-state multi-condition model. We draw a sample from the Bayesian posterior model density using a molecular dynamics algorithm where all structural states are jointly restrained by the satisfaction of all X-ray datasets in addition to molecular mechanics. We benchmarked the method using synthetic X-ray datasets simulated from multi-state SARS-CoV-2 M^{PTO} structural models and showed that multiple structural states are needed to satisfy the data and that by including multiple X-ray datasets, all datasets are better satisfied individually. We illustrate our method to compute a multi-state model of the SARS-CoV-2 viral target M^{PTO} from multi-temperature X-ray under 6 conditions and show that all datasets are satisfied better by increasing the number of states and X-ray datasets.

Methods

Overview of modeling method. A model is a depiction of our knowledge about a system or process. We wish to inform the model based on the input information, which generally includes experimental data and prior models (*eg*, physical theories and statistical preferences). A model can then rationalize past observations and predict future ones. Modeling is the search for a set of models consistent with the input information. We aim to find all models that satisfy the input information, reflecting the uncertainty of the input information and the modeling process. It is convenient to divide modeling into three steps: (i) specifying all model variables (representation), (ii) ranking alternative models by their agreement with the input information (scoring), and (iii) generating a sample of good-scoring models (sampling). A model should be validated before being interpreted, often by examining the satisfaction of input information withheld from the modeling process. Multiple iterations of gathering input information, modeling, and validation are often necessary to compute a sufficiently precise model (31, 32).

Input information. The input information may be used to inform any step of modeling: representation, scoring, and sampling. Here, we use the following input information: 1) X-ray datasets D_1, \dots, D_J collected under distinct experimental conditions (*eg*, at distinct temperatures) and corresponding PDB models. Each X-ray dataset is a set of observed structure factor amplitudes $\{|F_{\vec{s}}^O|\}_{\vec{s} \in S}$ indexed by a scattering vector \vec{s} . The X-ray datasets are used in scoring. The PDB model is used in scoring and sampling. 2) The CHARMM19 force field parameters (33). The force field parameters are used in scoring to evaluate the prior.

Representation. The representation defines the degrees of freedom whose values are to be determined by modeling. The multi-state multi-condition model M includes the following variables: (i) N states M_i (N is selected before modeling). Each state is an independently parameterized model of a protein structure. The number and type of atoms (composition) of each state are based on the PDB models (Input Information). Only non-hydrogen atoms from the protein are included, however, the representation may be generalized to include additional atoms/molecules (*eg*, hydrogen, solvent, ion, and ligand). As we are computing a model containing a set of discrete structural states, all atomic B-factors are set to 15 and atomic occupancies to 1. (ii) Weight matrix $W_{N \times J}$ containing the weight of M_i under condition j , w_{ij} . The weight matrix is subject to the constraint that the sum of the weights of all states under each condition (*ie*, columns) is 1. (iii) Nuisance variables, taken from established modeling methods, to improve the fit of the X-ray datasets by the model (34, 35). For clarity, we do not include nuisance parameters in the notation.

Scoring. A scoring function assesses the match of the model to the input information. Bayesian inference is one approach to assess this fit. In Bayesian inference, the posterior model density $p(M|D, I)$ describes the relationship between model M , prior information I , and data D . The posterior model density may be factored into a prior $p(M|I)$ that quantifies the state of knowledge prior to the observation of data and a

a

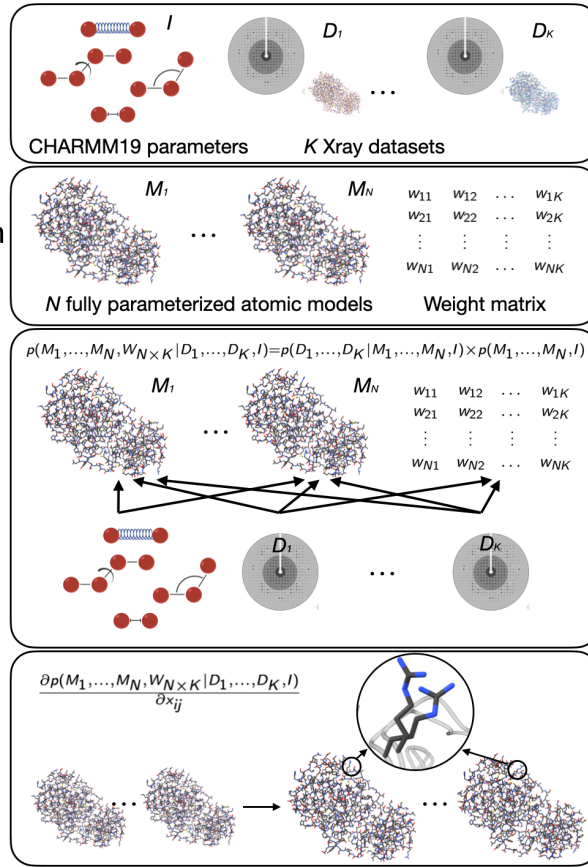
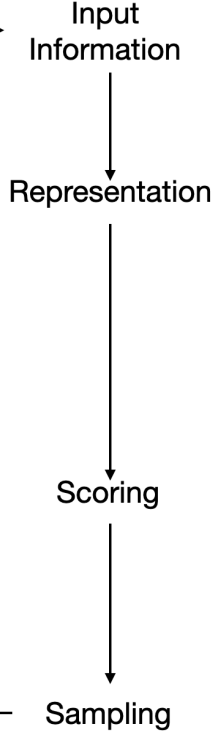


Fig. 1. modeling can be framed as a model search given some input information. Here, the input information is the CHARMM19 force field parameters and J X-ray datasets collected for the same system under distinct experimental conditions (eg, temperature). The representation is the N atomic states containing all the heavy atoms of the system along with the weight matrix that parameterizes the weight of each state under each condition. All states and the weight matrix are scored collectively against each X-ray dataset. Each state is individually scored against the molecular mechanics force field. A sample is drawn from the posterior model density using molecular dynamics. All states are initialized by a starting structure and the force on the atoms is computed from the satisfaction of all X-ray datasets along with the molecular mechanics, and the weights are stochastically sampled.

likelihood $p(D|M, I)$ that quantifies what is learned from the observation of data (36):

$$p(M|D, I) = p(D|M, I) \times p(M|I) \quad [1]$$

It is often helpful to decompose the likelihood into a forward model $f(M)$ that simulates a noiseless data observation and a noise model $N(f(M); D, \sigma)$ that quantifies the difference between the observed and simulated data based on a model of the experiment's noise parameterized by σ (31).

Joint-likelihood. The likelihood of observing all X-ray diffraction patterns from a multi-state multi-condition model is $p(D_1, \dots, D_J | M_1, \dots, M_N, W_{N \times J})$. The likelihood may be factored into the likelihood of observing each condition's dataset:

$$\begin{aligned} p(D_1, \dots, D_J | M_1, \dots, M_N, W_{N \times J}) \\ = \prod_{j=1}^J p(D_j | M_1, \dots, M_N, W_{N \times j}) \end{aligned} \quad [2]$$

The likelihood for each condition's dataset depends only on the set of states and the weights representing the states under the corresponding condition (matrix column):

$$\begin{aligned} p(D_j | M_1, \dots, M_N, W_{N \times j}) \\ = p(D_j | M_1, \dots, M_N, w_{1j}, \dots, w_{Nj}) \end{aligned} \quad [3]$$

The likelihood of an X-ray dataset is factored into the likelihood of observing each structure factor amplitude $F_{\vec{s}}^O$ (35):

$$\begin{aligned} p(D_j | M_1, \dots, M_N, w_{1j}, \dots, w_{Nj}) = \\ \prod_{\vec{s} \in S} p(F_{\vec{s}}^O | M_1, \dots, M_N, w_{1j}, \dots, w_{Nj}) \end{aligned} \quad [4]$$

The likelihood of a structure factor amplitude given a multi-state multi-condition model is a noise model that quantifies the difference between the observed structure factor amplitude $F_{\vec{s}}^O$ and the model structure factor amplitude $F_{\vec{s}}^M$ simulated by the forward model from the multi-condition multi-state model (35):

$$p(F_{\vec{s}}^O | M_1, \dots, M_N, w_{1j}, \dots, w_{Nj}, I) = p(F_{\vec{s}}^O | F_{\vec{s}}^M) \quad [5]$$

The noise model is based on the assumption that the real and imaginary parts of a complex structure factor \mathbf{F} are sampled from a two-dimensional Gaussian with variance $\epsilon\beta$ and scale parameter $\alpha_{\vec{s}}$ (35):

$$\begin{aligned} p(\text{Re}\mathbf{F}, \text{Im}\mathbf{F}) = \frac{1}{\pi\epsilon\beta_{\vec{s}}} \exp(-(\text{Re}\mathbf{F} - \alpha_{\vec{s}}F^M \cos\phi^M)^2 \\ + (\text{Im}\mathbf{F} - \alpha_{\vec{s}}F^M \sin\phi^M)^2) / \epsilon\beta_{\vec{s}} \end{aligned} \quad [6]$$

The likelihood for a single structure factor amplitude is (35):

$$p(F_{\vec{s}}^O | F_{\vec{s}}^M, \alpha_{\vec{s}}, \beta_{\vec{s}}) = \begin{cases} \frac{2F_{\vec{s}}^O}{\epsilon\beta_{\vec{s}}} \exp(-\frac{(F_{\vec{s}}^O)^2 + \alpha_{\vec{s}}^2 (F_{\vec{s}}^M)^2}{\epsilon\beta_{\vec{s}}}) \times \\ I_0(\frac{2\alpha_{\vec{s}} F_{\vec{s}}^O F_{\vec{s}}^M}{2\epsilon\beta_{\vec{s}}}) \\ \dots \text{if } \vec{s} \text{ acentric} \\ (\frac{2}{\epsilon\pi\beta_{\vec{s}}})^{1/2} \exp(-\frac{(F_{\vec{s}}^O)^2 + \alpha_{\vec{s}}^2 (F_{\vec{s}}^M)^2}{2\epsilon\beta_{\vec{s}}}) \times \\ \cosh(\frac{2\alpha_{\vec{s}} F_{\vec{s}}^O F_{\vec{s}}^M}{2\epsilon\beta_{\vec{s}}}) \\ \dots \text{if } \vec{s} \text{ centric} \end{cases} \quad [7]$$

where I_0 is the hyperbolic Bessel function of the first kind ($\alpha = 0$) and \cosh is the hyperbolic cosine function.

Forward model. The forward model $F_{\vec{s}}^M$ is the magnitude of the complex model structure factor $\mathbf{F}_{\vec{s}}^M$:

$$F_{\vec{s}}^M = |\mathbf{F}_{\vec{s}}^M| \quad [8]$$

The model structure factor is the sum of the protein $\mathbf{F}_{\vec{s}}^C$ and bulk solvent structure factor $\mathbf{F}_{\vec{s}}^B$ with an overall scale factor k_{total} and bulk solvent scaling factor k_{mask} (37).

$$\mathbf{F}_{\vec{s}}^M = k_{\text{total}}(\mathbf{F}_{\vec{s}}^C + k_{\text{mask}}\mathbf{F}_{\vec{s}}^B) \quad [9]$$

It is well-established how to compute the scattering from a model of a protein structure assuming a perfect crystal (38). Here, we compute the scattering of a model of a protein structure with multiple weighted structural states. It is important to keep in mind that different structures and positions of molecules in a crystal can result in the same X-ray diffraction data, even when the data are noiseless, complete, and collected instantaneously. We show that the scattering in the 5 cases is equivalent (2ab).

The degeneracy between (i) and (ii) relates the scattering of a multi-state crystal to a set of single-state crystals. The structure factor for a multi-state model may be computed from the structure factor $\mathbf{F}_{i\vec{s}}$:

$$\mathbf{F}_{\vec{s}} = \sum_{i=1}^{\text{states}} \mathbf{F}_{i\vec{s}} \quad [10]$$

The degeneracy of (i), (iv), and (v) complicates the interpretation of a match between a multi-state model and an X-ray dataset. For example, imagine we compute an N -state model with N rotamers for residues A and B each. Without exhaustive sampling of all possible multi-state models, we do not know if the rotamer pair always occurs in the same molecule or is one of many possible combinations of the rotameric states in multiple molecules. In other words, we cannot rule out any combination of the rotameric states of A and B in one molecule. As an aside, the degeneracy of the X-ray data might be broken by using additional information for modeling; for example, a potential energy function may rule out combinations of rotameric states that result in overlapping atoms.

Prior. The prior for the multi-state multi-condition model is the product of a prior for each state:

$$p(M|I) = \prod_i^N p(M_i|I) \quad [11]$$

The prior for a state is the Boltzmann distribution corresponding to the potential energy of the state, including terms for bond lengths b , bond angles θ , dihedral angles ϕ , improper dihedral angles w , and non-bonded interactions r_{ij} :

$$p(M_i|I) = -\frac{1}{Q} \left(\exp \sum_{\text{bnd}} k_b (b - b_0)^2 + \sum_{\text{ang}} k_\theta (\theta - \theta_0)^2 \right. \quad [12] \\ \left. + \sum_{\text{dih}} k_\phi [1 + \cos(n\phi - \delta)] + \sum_{\text{imp}} k_w (\omega - \omega_0)^2 \right. \\ \left. + \sum_{\text{nb}} [\epsilon \left(\left(\frac{R_{\text{min}_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{\text{min}_{ij}}}{r_{ij}} \right)^6 \right)] \right)$$

where the parameters k_b , b_0 , k_θ , θ_0 , k_ϕ , δ , k_w , w_0 , ϵ , and $R_{\text{min}_{ij}}$ are obtained from the CHARMM19 molecular mechanics force field (33) and Q is the partition function, which is ignored in practice.

Scoring function. A Bayesian scoring function for ranking alternative solutions based on the available information is:

$$S = -\log p(D|M, I) - \log p(M|I) \quad [13]$$

Sampling. The purpose of sampling is to find all models consistent with the input information. In Bayesian modeling, this is achieved by computing a sufficiently converged estimate of $p(D|M, I)$.

Atomic coordinates. The atomic positions for each state are sampled via molecular dynamics (MD). The atomic positions of a state are initialized to one of the input PDB models. The velocity of each atomic coordinate is sampled from a Boltzmann distribution with $T = 300\text{K}$. Because of the likelihood, the force on the atoms is non-conservative, and a thermostat is used to maintain a simulation temperature of 300K (15). We do not include explicit solvent molecules in the simulations, but the likelihood restrains each state to well-folded conformations. To account for discrepancies in the X-ray datasets, we reset the center of the mass to the center of mass of the corresponding PDB model before scoring against an X-ray dataset. The force computed on each atom is the partial derivative of the scoring function with respect to the atomic position of atom \vec{x} in state i , \vec{x}_i :

$$\frac{\partial S}{\partial \vec{x}_i} = - \left(\frac{\partial \log p(D_1|M_1, \dots, M_N, w_{11}, \dots, w_{N1}, I)}{\partial \vec{x}_i} + \dots \right. \quad [14] \\ \left. + \frac{\partial \log p(D_J|M_1, \dots, M_N, w_{1J}, \dots, w_{NJ}, I)}{\partial \vec{x}_i} \right. \\ \left. - \frac{\partial \log p(M_i|I)}{\partial \vec{x}_i} \right)$$

To calculate $\frac{\partial \log p(D_j|M_1, \dots, M_N, w_{1j}, \dots, w_{Nj}, I)}{\partial \vec{x}_i}$ more efficiently, a quadratic approximation of $p(F_{\vec{s}}^O | F_{\vec{s}}^M)$ is used (35). By computing the partial derivative of all atomic positions for a multi-state model, the atoms are restrained to satisfy a heterogeneous X-ray dataset collectively. Furthermore, all atoms of all states are restrained to satisfy all J X-ray datasets collectively.

Because the magnitudes of the derivative of the likelihood and prior can vary significantly, we introduce 2 scaling parameters w_{xray} and w_{auto} :

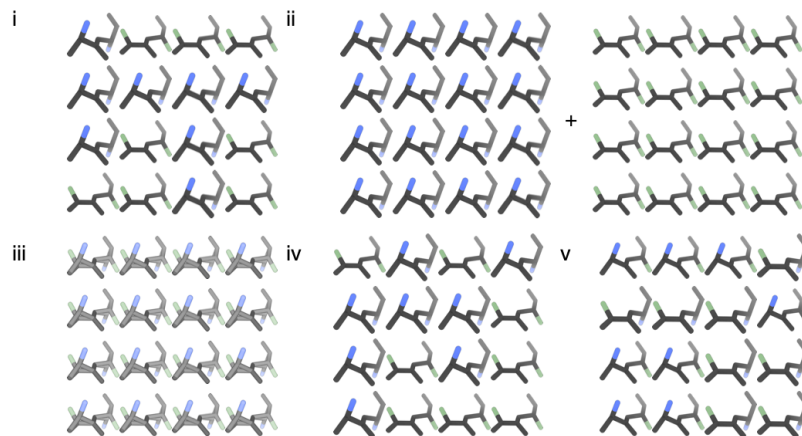
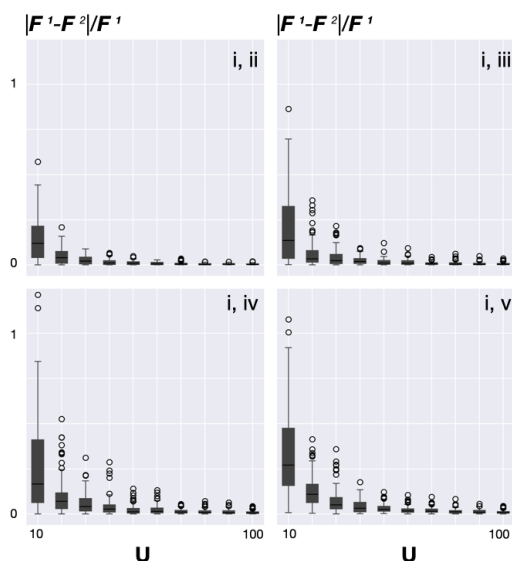
a**b**

Fig. 2. Five cases where the scattering is degenerate based on a multi-state model containing 2 states M_1 and M_2 with w_1 and w_2 . (i) Scattering from a heterogeneous crystal containing unit cells occupied by either M_1 or M_2 with probability w_1 and w_2 , respectively. (ii) A weighted combination of the scattering from homogeneous crystals containing identical unit cells with a single copy of M_1 and M_2 respectively. (iii) Scattering from a homogeneous crystal that contains the occupancy-weighted average of M_1 and M_2 (ie, the atomic occupancy of an atom from M_1 and M_2 is w_1 and w_2 , respectively). (iv) Scattering from a heterogeneous crystal the same as i where the states are re-indexed (ie, M_1 becomes M_2 and w_1 becomes w_2 and vice versa). (v) If $w_1 = w_2$, the scattering from a heterogeneous crystal except the states contain an intermixing of the atomic positions of M_1 and M_2 . **b**, as the number of unit cells goes to infinity, the reciprocal lattice points simulated from a crystal in ii, iii, iv, and v converge to those for the crystal in i. We construct 100 random reference crystals with U^3 unit cells where each unit cell is randomly populated by M_1 or M_2 with probability w_1 , w_2 respectively. M_1 , M_2 each contain 2 scatterers with random atomic position and scattering type. We compute the reciprocal lattice points up to a maximum miller index of $hkl=(3,3,3)$ via a discrete Fourier Transform (DFT). For (ii), (iii), (iv), and (v), we construct 100 random crystals as described and compute the same reciprocal lattice points. For each reciprocal lattice point, we compute the Euclidean distance for each structure factor to the reference structure factor U from 10 to 100. As U grows larger, the average Euclidean distance between the structure factors and reference structure factors converges to 0.

$$\frac{\partial S}{\partial \vec{x}_i} = -w_{\text{xray}}(w_{\text{auto}}^0 \frac{\partial \log p(D_1|M_1, \dots, M_N, w_{11}, \dots, w_{N1}, I)}{\partial \vec{x}_i} + \dots + w_{\text{auto}}^J \frac{\partial \log p(D_J|M_1, \dots, M_N, w_{1J}, \dots, w_{NJ}, I)}{\partial \vec{x}_i}) - \frac{\partial \log p(M_i|I)}{\partial \vec{x}_i} \quad [15]$$

w_{auto}^j is computed automatically to ensure the average force on the atoms from the likelihood of D_j is approximately equal to that from the prior (21). To account for the presence of multiple X-ray datasets, w_{xray} is selected before modeling. We select $w_{\text{xray}} \in \{2^{-i}; -5 \leq i \leq 5\}$ by computing a trial sample using each w_{xray} and identifying the w_{xray} that results in a sample that best satisfies the data (R^{free}) after filtering out w_{xray} that produce non-physiological conformations (3c).

We perform a short refinement minimization of $p(M_i|I)$ for each state of each model.

Weight matrix. For every 100 steps of sampling of the atomic positions, we propose 10 sets of weights per condition. A proposal weight is accepted for a condition if it improves the likelihood for the corresponding X-ray dataset. Each proposal is sampled from a Normal distribution with a mean equal to the current weight and a standard deviation of 0.05 and normalized.

Nuisance parameters. The nuisance variables are not stochastically sampled to enhance computational efficiency but optimized by minimizing the least squares residual between the model and the observed X-ray diffraction data (34).

Software availability. The software, input files, and output files are freely available at as part of our open-source *Integrative Modeling Platform* (<https://integrativemodeling.org/2.20.0/doc/manual/>) (39). The software relies on integration with *Phenix* (<https://phenix-online.org/>) (37), as described previously (40).

Results

Synthetic benchmark.

Framework for analyzing a modeling method. If the ground truth (native) is known, a modeling method may be evaluated by its ability to find the native as the best-scoring model. Assuming all models are enumerated with sufficient precision, the only relevant feature of the scoring function is the native being the global minimum of the scoring function (global minimum accuracy) (3a1, a2). However, in practice, all models cannot be enumerated with sufficient precision due to the high dimensionality of the search space. Therefore, we use stochastic search methods to find the global minimum. The efficacy of most search methods depends strongly on the scoring function having a funnel shape around the global minimum. The radius of convergence of the funnel is the area where meaningful gradients can be computed to assist convergence to the global minimum (the width of the funnel) (3a3). The smoothness of the funnel is defined as the correlation between the score and model coordinates (3a4). Most search methods benefit from a smoother funnel with a larger radius of convergence.

We benchmark our modeling method by comparing the 1-state 1-condition, 1-state 2-condition, 2-state 1-condition, and 2-state 2-condition scoring functions. Because the prior has been evaluated previously, we focus on the negative log-likelihood and refer to the negative log-likelihood as the scoring function. First, we estimate the global minimum of the scoring function. Second, we quantify the thoroughness of the search process by plotting the best score found as a function of the amount of sampling (convergence curve). Third, we interpret the differences between these plots for different scoring functions by considering their radius of convergences and smoothness. If the differences in the convergence curve cannot be explained by the differences in the radius of convergence and smoothness of the scoring function, then they must be a result of higher-dimensional characteristics of the landscape.

Natives. There is no large set of multi-state multi-condition models with corresponding experimentally measured X-ray datasets. Thus, we use a synthetic benchmark, where a reflection dataset is computed, rather than measured, from a native. The benchmark consists of 10 2-state 2-condition models constructed, in part, from a previously determined SARS-CoV-2 M^{Pro} structure (29) (3a).

Native reflection datasets. For a native, X-ray datasets D_1 and D_2 , corresponding to condition 1 and condition 2, are simulated via the forward model at 2.0 Å resolution based on the experimentally determined experimental unit cell dimensions and space group (29) (3b). Gaussian noise is applied to the simulated structure factor amplitudes. Reflections within each resolution shell are randomly withheld for model assessment.

Model error. We define the model error of a N_A -state J -condition model by its deviation from a N_B -state J -condition model where all states are compositionally identical as follows:

$$\text{error} = \frac{1}{J} \sum_{j=0}^J \text{RMSD}(\text{avg}(M_1^A, \dots, M_{N_A}^A, w_{1j}, \dots, w_{Nj}), \text{avg}(M_1^B, \dots, M_{N_B}^B, w_{1j}, \dots, w_{Nj})) \quad [16]$$

The inside sum is the RMSD between the weighted average of M^A and the weighted average of M^B using the j column of the respective weight matrices. In the weighted average, the atomic position vector k is the weighted sum of the respective atomic position vectors across all states. Computing the model error in this fashion is useful for the following reasons. First, we can compute the model error when $N_A \neq N_B$. Second, if $N_A = N_B = 1$, then the model error is the average of the RMSD function across all conditions. Third, the model error equals 0, if and only if the atomic position vector of the model average of M^A and the model average of M^B are equal for all atoms for all conditions. It is possible, however, for distinct models to have the same weighted average structure and thus an error of 0. Fourth, the model error shares the symmetry properties with the multi-state X-ray forward model for a given condition. The model error and multi-state X-ray forward model are a function of the set of atomic positions for a given atom k (along with their weights) and, therefore, independent of the assignment of an atom to a

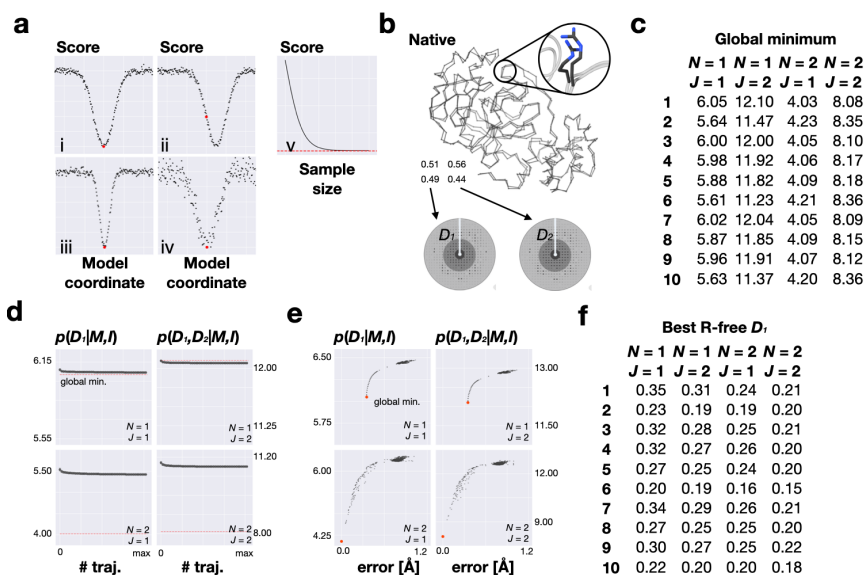


Fig. 3. i-iv, depictions of a scoring function relating the model error to the score. The native is shown in red. ii) a scoring function where the native is not the scoring function global minimum (poor global minimum accuracy). iii) a scoring function with a small funnel (poor radius of convergence). iv) a scoring function where the correlation between the model error and score (poor smoothness). v) a depiction of the thoroughness of the search relating the best score found to the size of the sample. The search process finds the global minimum with a sufficiently large sample (red line). **b**, one of the native models used in the synthetic benchmark. The native model contains 2 states under 2 conditions. A native dataset is simulated from the native model using the forward model for each set of weights (D_1 and D_2). **c**, the thoroughness of the search process for an example native showing the best score found as a function of the amount of the sampling for all 4 scoring functions. The score is either the likelihood of D_1 (if $J = 1$) or the joint likelihood of D_1 and D_2 (if $J = 2$). The amount of sampling is quantified by the number of trajectories included in the sample from 1 to 250 (max). **d**, the multi-state multi-condition scoring function for N in 1, 2 and J in 1, 2 for an example native. The error is the multi-state multi-condition error in \AA relative to the native. The score is either the likelihood of D_1 (if $J = 1$) or the joint likelihood of D_1 and D_2 (if $J = 2$). Each decoy structure is a point and the global minimum of the scoring function is in red. **f**, The best R^{free} found in the sample for each pair of N, J for all natives.

specific state. For example, the model error would not change, if the states in M^A or M^B are re-indexed. More generally, the model error would not change, if the positions of atoms k in any 2 states are swapped (state mixing).

Accuracy of the global minimum. We estimate the value of the global minimum of the 4 scoring functions for each native (3c). For a 2-state scoring function, we know that the global minimum error is 0 for all natives. We estimate the value of the global minimum of the 1-state scoring functions by computing the score of a large set of decoy models. The 1-state scoring functions do not have a global minimum near the native, because the 1-state model representation does not provide a good match to the 2-state model from which the data was simulated. Using a representation that better matches the data-generating process increases the accuracy of the global minimum of the scoring function.

Sampling convergence. To quantify the thoroughness of the search process, we plot the sampling convergence of the 4 scoring functions for each native (3d).

The model search with a 1-state scoring function converges to the best score near the estimated global minimum for the respective scoring functions. In other words, the model search for satisfying the 2-state datasets finds a nearly best-scoring 1-state model. The model search for the 2-state scoring functions converges to a score considerably better than the 1-state samples. However, as expected for a stochastic search, it does not find the global minima of the scoring functions, instead becoming stuck in local minima. The convergence curves suggest that further sampling would not significantly improve the best score found.

For both the 1-state and 2-state scoring functions, the consideration of an additional X-ray dataset does not significantly improve the best score found. The mean improvement in the 1-state 1-condition likelihood of D_1 vs the 1-state 2-condition marginal likelihood of D_1 is 0.050 ± 0.070 . The mean improvement in 2-state 1-condition likelihood of D_1 vs the

2-state 2-condition marginal likelihood of D_1 is 0.054 ± 0.112 . However, we see a significant improvement in R^{free} of D_1 by including a second X-ray dataset (3f). The mean improvement in R^{free} from 1-state 1-condition vs 1-state 2-condition is 0.054 ± 0.026 . The mean improvement in R^{free} from 2-state 1-condition vs 2-state 2-condition is 0.030 ± 0.021 .

The shape of the scoring function. The convergence of the model search to the global minima of the 1-state scoring functions suggests that the scoring functions are smooth and have a large enough radius of convergence to facilitate the model search by molecular dynamics simulations, given the initial conditions.

The score and model error are correlated with each other more weakly for the 2-state vs 1-state scoring functions (3e). This observation is consistent with the 2-state scoring function depending on twice the number of atomic variables and state weights. The inability of the model search to find the global minimum of the 2-state scoring function can likely be attributed to this decrease in the smoothness of the scoring function. Despite the randomized initial conditions, the trajectories tend to be trapped in similarly scoring, but distinct, local minima, as evidenced by the low variance in the convergence curves.

The smoothness and radius of convergence of the scoring function do not appear to be sensitive to considering additional conditions. Thus, we attribute the improvement in R^{free} from adding an additional condition to characteristics of the scoring function beyond our definitions of its smoothness and radius of convergence.

Multi-state multi-condition modeling of SARS-CoV-2 M^{pro}. We illustrate our method by computing multi-state multi-condition models of the SARS-CoV-2 main protease (M^{pro}), a therapeutic target (41). Data was collected and modeled previously at 100K, 240K, 277K, 298K, 298K* (high humidity), and 310K (29).

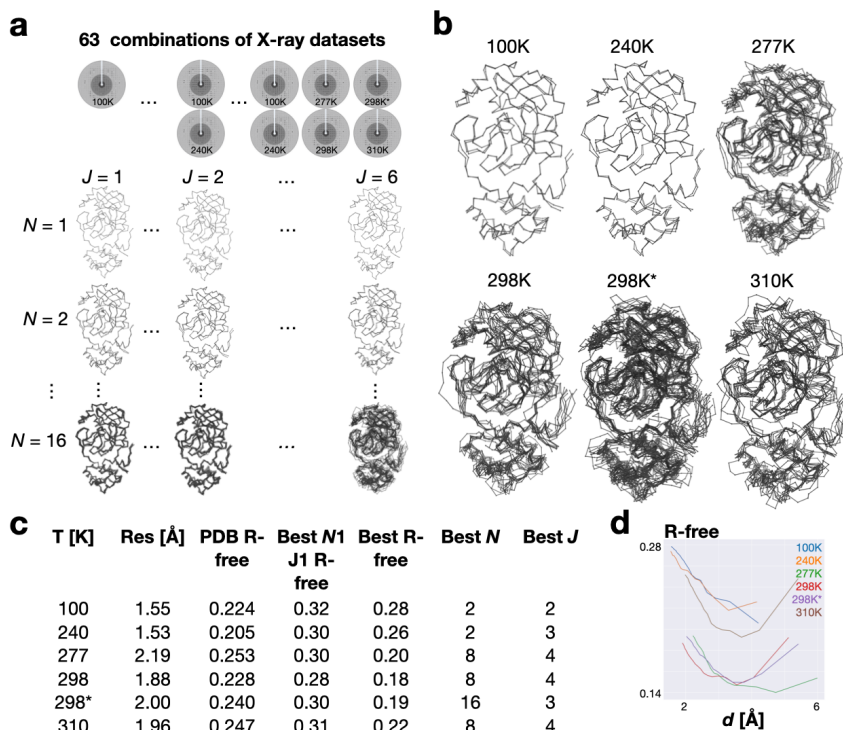


Fig. 4. Overview of all samples built for SARS-Cov-2 M^{pro}. A sample of good-scoring models was computed from the model posterior density for N in 1, 2, 4, 8, 16, and all combinations of X-ray datasets (J from 1 to 6). There are 63 non-empty combinations of X-ray datasets. Each sample contains 25 trajectories (7,500 models total). The super-sample contains 7,500 × 63 × 5 = 2,362,500 models. **b**, The best-free model from the super-sample for each of the 6 temperature-dependent X-ray datasets. The columns include the temperature the dataset was collected under, the resolution of the dataset, the R^{free} of the PDB model, the R^{free} of the best-free 1-state 1-condition model in the supersample for the dataset, the R^{free} of the best-free model in the supersample for the dataset, the number of states (N) in the best-free model, and the number of conditions used to inform the best-free model (J). **d**, The R^{free} of the best-free model for each dataset as a function of the resolution. The observed and model structure factors computed from the best-free model were divided into 20 resolution bins. The y-axis is the R^{free} of the resolution bucket and the X-axis is the minimum resolution (Å) of the resolution bucket.

Using our method as described, we compute a sample of models for each N in 1, 2, 4, 8, and 16 based on each of the 63 combinations of datasets, containing 1 to 6 datasets (J) (4a). We construct a supersample containing all models for all N and J . For an X-ray dataset, the best R^{free} N -state J -condition model is defined as the supersample model with N states and J conditions with the minimum R^{free} ; similarly, the best R^{free} model is defined as the supersample model with the minimum R^{free} (4b). For each dataset, we compare the R^{free} values of the PDB model, best R^{free} model, and best R^{free} 1-state 1-condition model (4c), as follows.

For all 6 datasets, the best R^{free} 1-state 1-condition model has a worse R^{free} than the PDB model. In the 1-state 1-condition case, the accuracy of the global minimum of our scoring function is less accurate, presumably because the PDB model representation includes hydrogen, ion, ligand, and solvent atoms, thus depicting the reality more completely; in addition, another reason may be decreased search thoroughness of our method due to the lack of manual adjustments between rounds of refinement.

For all 6 datasets, the best R^{free} model has a better R^{free} than the best R^{free} 1-state 1-condition model. The best R^{free} model contains 2-16 states (N) computed based on 2-4 datasets (J). As shown previously, the model's fit of the free reflections likely improves for 2 reasons: (i) increased global minimum accuracy from adding additional states to the model representation and (ii) better R^{free} from adding additional datasets. For 4 of the 6 datasets ($\geq 277\text{K}$), the best-free model has $N \geq 8$ states and fits the dataset better than the PDB model. For higher temperature datasets, the improvement in R^{free} from including additional states and datasets is greater than the loss in fit on account of the representation. This is not the case for the 2 lower-temperature datasets. For all 6 datasets, the fit of the free reflections by the best-free model worsens as a function of

resolution, also contributing to the relatively poor fit of the free reflections of the highest resolution datasets (100K, 240K) (4d). The fit of the free reflections of the 277K, 298K, and 298K* is considerably better than the fit of corresponding reflections in 100K and 240K datasets, suggesting the better fit of the higher resolution datasets is at least in part from the including structural heterogeneity rather than the dataset being lower resolution and therefore more ambiguous.

We report the improvement in the R^{free} (ΔR^{free}) for each dataset by the best-free N -state J -condition model relative to the best-free 1-state 1-condition model for each N and J (3a). Generally ΔR^{free} exhibits unimodal behavior with a maximum improvement at $J = 3.4$ and 2.4 for $\geq 277\text{K}$ and $<277\text{K}$ datasets respectively. For 277K, 298K*, and 310K, the minimum ΔR^{free} (biggest improvement) occurs when all 4 of the $\geq 277\text{K}$ datasets were used together. For 298K, the minimum improvement occurs using 3 of the 4 $\geq 277\text{K}$ datasets. The lower temperature datasets are similarly grouped with 240K minimum ΔR^{free} based on the 100K, 240K, and 310K datasets and the 100K minimum ΔR^{free} based on the 100K and 240K datasets.

The partitioning of X-ray datasets by temperature condition is likely a result of the structural heterogeneity in the crystal being similar at similar temperatures. Datasets observed under temperature conditions with similar structural heterogeneity provide additional independent observations that prevent the multi-state model from overfitting to the noise of a single dataset. Computing a sample of N -state 1-condition models with w_{xray} as a multiple (1-6x) of the optimized w_{xray} value causes the best R^{free} to worsen in most cases, demonstrating that adding additional datasets does not satisfy the data better simply because of an increase in the contribution from X-ray datasets to the overall score (5b).

We report the improvement in the fit of the force field (ΔFF) by the best-free N -state J -condition model relative

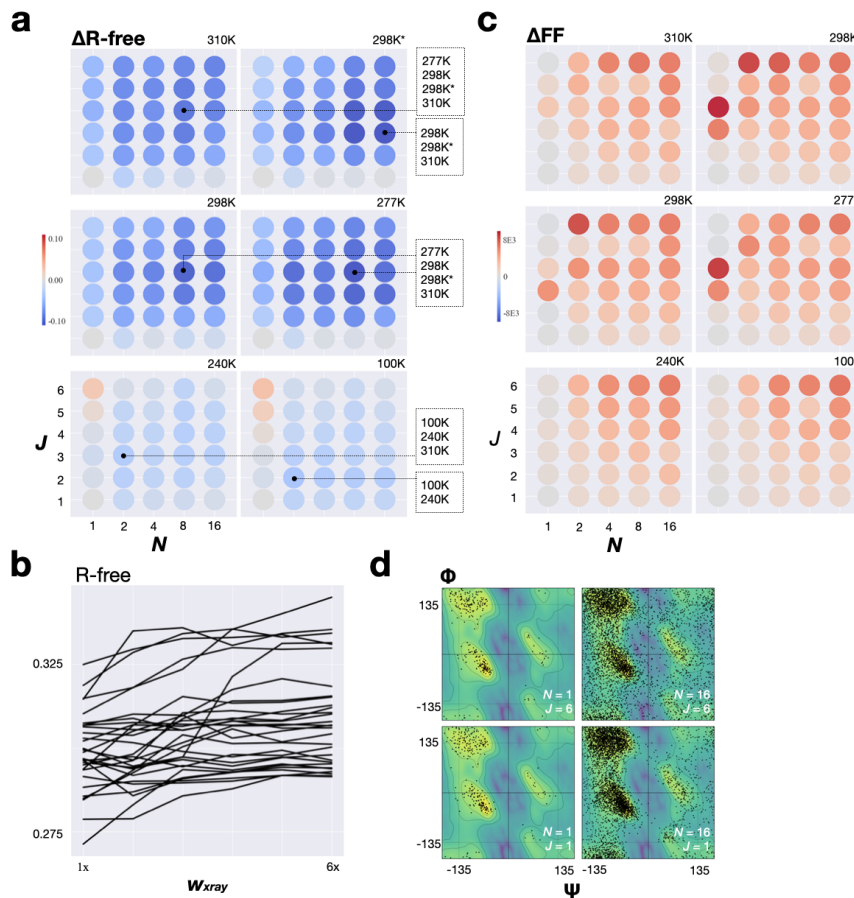


Fig. 5. The difference between the R_{free} of the 1-state 1-condition best-free model and the best-free N -state J -condition model (ΔR_{free}) for all N and all J for all 6 datasets. A negative ΔR_{free} indicates greater improvement in satisfaction of the X-ray dataset. For each dataset, the ΔR_{free} corresponding to the best-free model is identified along with the corresponding input datasets. **b**, The difference between the force field energy (divided by the number of states) of the 1-state 1-condition best-free model and the best-free N -state J -condition model (ΔFF) for all N and all J for all 6 datasets. A negative ΔFF indicates greater improvement in satisfaction of the molecular mechanics force field. **c**, The R_{free} of the best-free N -state 1-condition model as a function of increased w_{xray} , up to $6 \times w_{\text{xray}}$. As the w_{xray} for each sample was picked based on the w_{xray} that produced a best-free model with the minimum R_{free} out of a set of trial w_{xray} values, increasing w_{xray} generally does not improve the R_{free} of the best-free model. **d**, The Ramachandran plot for the best-free 1-state 1-condition, 1-state 6-condition, 16-state 1-condition, and 16-state 6-condition models for the 277K dataset.

to the best-free 1-state 1-condition model for each N and J for each dataset (5c). The satisfaction of the force field decreases as a function of N and J . The dissatisfaction of the force field as a function of N is likely a result of the model being able to use a larger number of parameters to satisfy density in the Fo-Fc map that would not be able to be satisfied using a lesser number of states. For example, the model may violate the force field to satisfy the electron density that represents an ordered solvent molecule. The dissatisfaction of the prior as a function of J is likely a result of the attempt to satisfy distinct datasets with a single set of states. To further explore why the impact of the number of states and conditions on the force field, we see that the number of Ramachandran outliers is a synergistic function of both N and J (5d). Integrative modeling aims to compute a model that best explains all input information, even when the input information represents varied experimental conditions and prior models.

Discussion

Here, we developed, benchmarked, and illustrated a Bayesian method to compute a model depicting multiple structural states from multiple X-ray diffraction patterns representing distinct experimental conditions.

First, we discuss the impact of introducing multiple states into the scoring function for a single condition (X-ray dataset). A multi-state scoring function evaluates the collective satisfaction of an X-ray dataset by all states. The scoring function can be generalized for scoring a model depicting multiple states by exploiting the degeneracy of the X-ray forward model for a weighted set of structural states and the weighted sum of the forward model for a single structural state (2b). In contrast, the degeneracy of the X-ray forward model of assigning an atom to a particular state complicates the interpretation of the multi-state model. To resolve this degeneracy, additional information, such as a molecular mechanics force field, is required.

We show that by introducing additional states, the global minimum of the scoring function is closer to the native (3c). For experimental data, where the true number of structural states far exceeds the number of states reasonably included in a model, adding additional states will continue improving the accuracy of the global minimum of the scoring function. The accuracy of the global minimum will also be further improved by including additional variables to represent the atomic content of a protein crystal, such as water molecules, ions, and ligands. When the optimal representation is unknown before modeling, the number and type of model parameters may be sampled alongside the values of the remaining model parameters. In such cases, it may be beneficial to consider model selection criteria such as Bayesian or Akaike Information Criterion to explain the data with as few model parameters as possible (42).

Adding additional states decreases the smoothness of the scoring function (3e), which may contribute to the inability of the model search to find the global minimum of the multi-state multi-condition scoring function. The decreased smoothness of the scoring function results from the increased number of model parameters. It is conceivable that removing some nuisance parameters from the likelihood, such as the scaling

factors α and β , could increase the smoothness of the scoring function.

Next, we discuss the impact of introducing multiple conditions into the scoring function. A multi-state multi-condition scoring function evaluates the collective satisfaction of all X-ray datasets by all states. The scoring function can be generalized to multiple conditions by computing a joint likelihood from the multiplicative product of the likelihood of each independent X-ray dataset. For a gradient-based sampling algorithm, such as molecular dynamics simulations, the force on an atom is the sum of the partial derivatives of each log-likelihood, thereby pushing the atom to a position that better satisfies all X-ray datasets. It was previously found advantageous to balance the first derivatives of the likelihood and the force field prior (w_{auto}) (21). Correspondingly, we also found by trial and error the best scaling factor for the partial derivatives for all log-likelihoods (w_{xray}).

Considering additional conditions significantly improves the satisfaction of the synthetic and experimental free reflections (3f, 5a), while it does not appear to impact the properties of the scoring function nor the satisfaction of the synthetic or experimental work reflections. Thus, consideration of additional conditions must impact more complex properties of the scoring function than its smoothness and radius of convergence. In other words, the geometry of a scoring function with multiple conditions improves the ability of the model search to find a minimum reflecting the actual native rather than the noise of the work reflections.

The model search may be further improved using advanced sampling algorithms, such as simulated annealing instead of the constant-temperature molecular dynamics simulations used here. Such methods would expand the explored model space for a given amount of CPU time (43). Additionally, future sampling may benefit from including additional information about stereochemistry (*eg*, sampling side chains using known rotameric states).

Several applications may benefit from Bayesian multi-state multi-condition modeling in X-ray crystallography. First, in serial crystallography, a series of small and often imperfect crystals are exposed to the X-ray beam, generating partial diffraction patterns at a series of time points (44). This trajectory could be represented as differently weighted mixtures of the same states along the sampled time points. Using our multi-state multi-condition scoring function would inform a model of each state at each time point based on the totality of the X-ray data, presumably improving the accuracy of the models as demonstrated here. Second, in a fragment screen, the protein is crystallized in the presence of one fragment at a time, corresponding to a single condition (45). Fragments often have low affinity, and therefore, only a fraction of the protein molecules in the crystal will be bound. The multi-state multi-condition representation can conceivably combine data from multiple experiments in several ways. For example, a multi-state multi-condition model may be computed where the states correspond to each fragment's apo state and the dominant holo state. As a result, X-ray diffraction patterns from each fragment will inform all states. Third, non-Bragg scattering is often ignored in modeling based on X-ray crystallography data. For example, diffuse scattering contains information on the correlated motions of atoms in the crystal (46). Our Bayesian framework

1241	may facilitate including another likelihood term based on	32. A Sali, From integrative structural biology to cell biology. <i>J. Biol. Chem.</i> 296 , 100743 (2021).	1303
1242	diffuse scattering data in computing a multi-state model.	33. BR Brooks, et al., CHARMM: the biomolecular simulation program. <i>J. Comput. Chem.</i> 30 , 1545–1614 (2009).	1304
1243	To facilitate the application of our method by scientists to	34. PV Afonine, RW Grosse-Kunstleve, PD Adams, A Urzhumtsev, Bulk-solvent and overall	1305
1244	many other systems, we implemented it as a module of our	scaling revisited: faster calculations, improved results. <i>Acta Crystallogr. D Biol. Crystallogr.</i>	1306
1245	open-source <i>Integrative Modeling Platform</i> (IMP) (39), freely	69 , 625–634 (2013).	
1246	available at https://integrativemodeling.org/2.20.2/doc/manual/ .	35. VY Lunin, PV Afonine, AG Urzhumtsev, Likelihood-based refinement. i. irremovable model	1307
1247		errors. <i>Acta Crystallogr. A</i> 58 , 270–282 (2002).	1308
1248	ACKNOWLEDGMENTS. Please include your acknowledgments	36. R Moelreath, Statistical rethinking: A Bayesian course with examples in R and stan. (2015).	1309
1249	here, set in a single paragraph. Please do not include any	37. D Liebschner, et al., Macromolecular structure determination using x-rays, neutrons and	1310
1250	acknowledgments in the Supporting Information, or anywhere	electrons: recent developments in phenix. <i>Acta Crystallogr D Struct Biol</i> 75 , 861–877	1311
1251	else in the manuscript.	(2019).	
1252	1. PA Reijo, ST Freer, Protein conformational substates from x-ray crystallography. <i>Prog.</i>	38. A Guinier, X-ray diffraction in crystals. <i>Imperfect Crystals, Amorph. Bodies. Dorer</i> (1963).	1312
1253	<i>Biophys. Mol. Biol.</i> 66 , 167–196 (1996).	39. D Russel, et al., Putting the pieces together: integrative modeling platform software for	1313
1254	2. CA Smith, et al., Population shuffling of protein conformations. <i>Angew. Chem. Int. Ed Engl.</i>	structure determination of macromolecular assemblies. <i>PLoS Biol.</i> 10 , e1001244 (2012).	1314
1255	54 , 207–210 (2015).	40. M Hancock, et al., Integration of software tools for integrative modeling of biomolecular	1315
1256	3. RA Woldeyes, DA Sivak, JS Fraser, E pluribus unum, no more: from one crystal, many	systems. <i>J. Struct. Biol.</i> 214 , 107841 (2022).	1316
1257	conformations. <i>Curr. Opin. Struct. Biol.</i> 28 , 56–62 (2014).	41. Z Jin, et al., Structure of mpro from SARS-CoV-2 and discovery of its inhibitors. <i>Nature</i> 582 ,	1317
1258	4. MA DePristo, PIW de Bakker, TL Blundell, Heterogeneity and inaccuracy in protein	289–293 (2020).	1318
1259	structures solved by x-ray crystallography. <i>Structure</i> 12 , 831–838 (2004).	42. P Stoica, Y Selen, Model-order selection: a review of information criterion rules. <i>IEEE</i>	1319
1260	5. LH Jensen, Refinement and reliability of macromolecular models based on x-ray diffraction	<i>Signal Process. Mag.</i> 21 , 36–47 (2004).	1320
1261	data. <i>Methods Enzym.</i> 277 , 353–366 (1997).	43. S Kirkpatrick, CD Gelatt, Jr, MP Vecchi, Optimization by simulated annealing. <i>Science</i> 220 ,	1321
1262	6. D Ringe, GA Petsko, Study of protein dynamics by x-ray diffraction. <i>Methods Enzym.</i> 131 ,	671–680 (1983).	1322
1263	389–433 (1986).	44. TRM Barends, B Stauch, V Cherezov, I Schlichting, Serial femtosecond crystallography. <i>Nat</i>	1323
1264	7. Z Sun, Q Liu, G Qu, Y Feng, MT Reetz, Utility of B-Factors in protein science: Interpreting	<i>Rev Methods Primers</i> 2 (2022).	1324
1265	rigidity, flexibility, and internal motion and engineering thermostability. <i>Chem. Rev.</i> 119 ,	45. T Krojer, JS Fraser, F von Delft, Discovery of allosteric binding sites by crystallographic	1325
1266	1626–1665 (2019).	fragment screening. <i>Curr. Opin. Struct. Biol.</i> 65 , 209–216 (2020).	1326
1267	8. A Kuzmanic, NS Pannu, B Zagrovic, X-ray refinement significantly underestimates the level	46. X Pei, N Bhatt, H Wang, N Ando, SP Meisburger, Introduction to diffuse scattering and data	1327
1268	of microscopic heterogeneity in biomolecular crystals. <i>Nat. Commun.</i> 5 , 3220 (2014).	collection. <i>Methods Enzym.</i> 688 , 1–42 (2023).	1328
1269	9. J Kuriyan, GA Petsko, RM Levy, M Karplus, Effect of anisotropy and anharmonicity on		1329
1270	protein crystallographic refinement. an evaluation by molecular dynamics. <i>J. Mol. Biol.</i> 190 ,		1330
1271	227–254 (1986).		1331
1272	10. JM Holton, S Classen, KA Frankel, JA Tainer, The r-factor gap in macromolecular		1332
1273	crystallography: an untapped potential for insights on accurate structures. <i>FEBS J.</i> 281 ,		1333
1274	4046–4060 (2014).		1334
1275	11. D Vitkup, D Ringe, M Karplus, GA Petsko, Why protein r-factors are so large: a		1335
1276	self-consistent analysis. <i>Proteins</i> 46 , 345–354 (2002).		1336
1277	12. H van den Bedem, JS Fraser, Integrative, dynamic structural biology at atomic		1337
1278	resolution—it's about time. <i>Nat. Methods</i> 12 , 307–318 (2015).		1338
1279	13. K Henzler-Wildman, D Kern, Dynamic personalities of proteins. <i>Nature</i> 450 , 964–972		1339
1280	(2007).		1340
1281	14. CA Schiffer, Time-averaging crystallographic refinement in <i>Computer Simulation of</i>		1341
1282	<i>Biomolecular Systems: Theoretical and Experimental Applications</i> , eds. WF van Gunsteren,		1342
1283	PK Weiner, AJ Wilkinson. (Springer Netherlands, Dordrecht), pp. 265–269 (1997).		1343
1284	15. BT Burnley, PV Afonine, PD Adams, P Gros, Modelling dynamics in protein crystal		1344
1285	structures by ensemble refinement. <i>Elife</i> 1 , e00311 (2012).		1345
1286	16. SA Wankowicz, et al., Uncovering protein ensembles: Automated multiconformer model		1346
1287	building for x-ray crystallography and Cryo-EM. <i>bioRxiv</i> (2024).		1347
1288	17. BT Riley, et al., qfit 3: Protein and ligand multiconformer modeling for x-ray crystallographic		1348
1289	and single-particle cryo-EM density maps. <i>Protein Sci.</i> 30 , 270–285 (2021).		1349
1290	18. MA Wilson, AT Brunger, The 1.0 Å crystal structure of Ca ²⁺ -bound calmodulin: an analysis		1350
1291	of disorder and implications for functionally relevant plasticity. <i>J. Mol. Biol.</i> 301 , 1237–1256		1351
1292	(2000).		1352
1293	19. EJ Levin, DA Kondrashov, GE Wesenberg, GN Phillips, Jr, Ensemble refinement of protein		1353
1294	crystal structures: validation and application. <i>Structure</i> 15 , 1040–1052 (2007).		1354
1295	20. J Kuriyan, et al., Exploration of disorder in protein structures by x-ray restrained molecular		1355
1296	dynamics. <i>Proteins</i> 10 , 340–358 (1991).		1356
1297	21. FT Burling, AT Brünger, Thermal motion and conformational disorder in protein crystal		1357
1298	structures: Comparison of multi-conformer and time-averaging models. <i>Isr. J. Chem.</i> 34 ,		1358
1299	165–175 (1994).		1359
1300	22. MC Thompson, Combining temperature perturbations with x-ray crystallography to study		1360
1301	dynamic macromolecules: A thorough discussion of experimental methods. <i>Methods</i>		1361
1302	<i>Enzym.</i> 688 , 255–305 (2023).		1362
	23. H Frauenfelder, GA Petsko, D Tsernoglou, Temperature-dependent x-ray diffraction as a		1363
	probe of protein structural dynamics. <i>Nature</i> 280 , 558–563 (1979).		1364
	24. H Frauenfelder, SG Sligar, PG Wolynes, The energy landscapes and motions of proteins. <i>Science</i> 254 , 1598–1603 (1991).		
	25. RF Tilton, Jr, JC Dewan, GA Petsko, Effects of temperature on protein structure and		
	dynamics: X-ray crystallographic studies of the protein ribonuclease-a at nine different		
	temperatures from 98 to 320 K. <i>Biochemistry</i> 31 , 2469–2481 (1992).		
	26. JS Fraser, et al., Hidden alternative structures of proline isomerase essential for catalysis. <i>Nature</i> 462 , 669–673 (2009).		
	27. B Halle, Biomolecular cryocrystallography: structural changes during flash-cooling. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 101 , 4793–4798 (2004).		
	28. DA Keedy, et al., Crystal cryocooling distorts conformational heterogeneity in a model		
	Michaelis complex of DHFR. <i>Structure</i> 22 , 899–910 (2014).		
	29. A Ebrahim, et al., The temperature-dependent conformational ensemble of SARS-CoV-2		
	main protease (mpro). <i>bioRxiv</i> (2021).		
	30. S Du, et al., Refinement of multiconformer ensemble models from multi-temperature x-ray		
	diffraction data. <i>bioRxiv</i> (2023).		
	31. MP Rout, A Sali, Principles for integrative structural biology studies. <i>Cell</i> 177 , 1384–1403		
	(2019).		