eLife

Reviewed Preprint
v2 • May 14, 2024
Revised by authors

Reviewed Preprint
v1 • September 5, 2023

Structural Biology and Molecular Biophysics

Uncovering Protein Ensembles: Automated Multiconformer Model Building for X- ray Crystallography and Cryo-EM

Stephanie A. Wankowicz ✉, **Ashraya Ravikumar, Shivani Sharma, Blake T. Riley, Akshay Raju, Jessica Flowers, Daniel Hogan, Henry van den Bedem, Daniel A. Keedy, James S. Fraser**

Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA • Structural Biology Initiative, CUNY Advanced Science Research Center, New York, NY 10031 • Ph.D. Program in Biology, The Graduate Center – City University of New York, New York, NY 10016 • Atomwise, Inc., San Francisco, CA, United States • Department of Chemistry and Biochemistry, City College of New York, New York, NY 10031 • Ph.D. Programs in Biochemistry, Biology, and Chemistry, The Graduate Center – City University of New York, New York, NY 10016

Abstract

In their folded state, biomolecules exchange between multiple conformational states that are crucial for their function. Traditional structural biology methods, such as X-ray crystallography and cryogenic electron microscopy (cryo-EM), produce density maps that are ensemble averages, reflecting molecules in various conformations. Yet, most models derived from these maps explicitly represent only a single conformation, overlooking the complexity of biomolecular structures. To accurately reflect the diversity of biomolecular forms, there is a pressing need to shift towards modeling structural ensembles that mirror the experimental data. However, the challenge of distinguishing signal from noise complicates manual efforts to create these models. In response, we introduce the latest enhancements to qFit, an automated computational strategy designed to incorporate protein conformational heterogeneity into models built into density maps. These algorithmic improvements in qFit are substantiated by superior R_{free} and geometry metrics across a wide range of proteins. Importantly, unlike more complex multicopy ensemble models, the multiconformer models produced by qFit can be manually modified in most major model building software (e.g. Coot) and fit can be further improved by refinement using standard pipelines (e.g. Phenix, Refmac, Buster). By reducing the barrier of creating multiconformer models, qFit can foster the development of new hypotheses about the relationship between macromolecular conformational dynamics and function.

eLife assessment

This work describes **important** updates to qFit, the state-of-the art tool for modeling alternative conformations of protein molecules based on high resolution X-ray diffraction or Cryo-EM data. The authors provide some **convincing** analyses of qFit's performance in selected test cases. This manuscript will be of interest to structural biologists and protein biochemists, since the adoption of qFit in structural refinement may lead to new mechanistic insights into protein function.

https://doi.org/10.7554/eLife.90606.2.sa3

Introduction

Macromolecular X-ray crystallography and single-particle electron microscopy (cryo- EM) can provide valuable information on macromolecular conformational ensembles. These experiments cannot capture all conformations present in solution, as many would disrupt the ability to obtain crystals or align classifiable particles[1]. However, careful modeling from high-resolution X-ray crystallography and cryo-EM data can reveal widespread conformational heterogeneity, particularly for protein side chains and local backbone regions[2,3]. Such discrete, local conformational heterogeneity is significant for many biological functions, including macromolecular binding, catalysis, and allostery[4–6].

While the underlying data from X-ray diffraction and cryo-EM experiments contains information on temporal and spatial averages of tens of thousands to billions of protein copies, conventional structural modeling and refinement procedures fail to capture much of this valuable information. Most depositions in the Protein Data Bank reflect only an averaged, single ground state set of atomic coordinates[7], ignoring weak but potentially biologically rich signals encoding alternative conformations sampled by distinct copies of the protein in the experiment.

Ideally, we would accurately model the complete ensemble of protein conformations reflected in experimental data[8]. The two ways to model the conformational heterogeneity present in the sample are to create ensembles or to use alternative conformations (multiconformers)[9]. The PDB “ensemble” format encodes multiple complete copies of the entire system in different models within a single file. Ensemble refinement approaches are implemented in phenix.ensemble_refinement[10] and Vagabond[11]. In contrast, multiconformers extend the conventional single structure model by encoding each individual conformation using a distinct "alternative location indicator (altloc)” within a single model. Altlocs are assigned distinct letters and can range from single atoms to a large number of connected or non-connected residues. Refinement and validation programs treat atoms sharing the same altloc as having the ability to interact with each other and with atoms lacking an altloc. In contrast, atoms with different altlocs cannot interact. By representing the underlying heterogeneity through discrete conformations with labeled altlocs, multiconformer models encode the distribution of states that contribute to the density map. Multiconformer models are notably easier to modify and more interpretable in software like Coot[12], unlike ensemble methods that generate multiple complete protein copies[10,13,14].

However, many factors make manually creating multiconformer models difficult and time-consuming. Interpreting weak density is complicated by noise arising from many sources, including crystal imperfections, radiation damage, and poor modeling[15–17] in X-ray crystallography, and errors in particle alignment and classification, poor modeling of beam-

induced motion, and imperfect detector Detector Quantum Efficiency (DQE) in high-resolution cryo-EM[18]. These factors make visually distinguishing signals in Coot[12] or other visualization software very difficult, especially when genuine low-occupancy signals overlap. Additionally, in X-ray crystallography, this process is iterative. Each time a new alternative conformation is placed, the resulting improvement in phases can impact the entire electron density map, often requiring adjustments to previously modeled regions. The difficulty of this process can lead to burnout and human bias, where parts of the protein are carefully modeled as multiconformers, whereas other regions remain modeled as single conformers. Despite these complications, multiconformer modeling can be implemented manually or using software such as FLEXR[19] or qFit, as described below.

To enable more routine and impartial multiconformer modeling, we have previously developed qFit[20–22]. This program leverages the ensemble-rich experimental data from density maps that are better than 2.0 angstroms (Å) resolution to automatically generate parsimonious multiconformer models[20,21]. As input, qFit takes a refined single-conformer structure and either a high-resolution X-ray or cryo-EM map as input, and then leverages powerful optimization algorithms to identify alternative protein[20,21] or ligand[23] conformations.

Here, we present updates to qFit including algorithmic changes to protein conformation selection based on Bayesian information criteria (BIC), B-factor sampling, and updated cryo-EM scoring. Collectively, these advances enable the unsupervised generation of multiconformer models that routinely improve R_{free} and model geometry metrics over single-conformer X-ray structures derived from high-resolution data across a diverse test set. We further demonstrate that qFit can identify alternative side-chain conformations in high-resolution cryo-EM datasets. With the improvements in model quality outlined here, qFit can now increasingly be used for finalizing high-resolution models to derive ensemble-function insights.

Results

Overview of qFit protein algorithm

qFit protein is a tool that automatically identifies alternative conformations based on a high-resolution density map (generally better than ~2 Å) and a well-refined single- conformer structure (generally R_{free} below 20%). For X-ray maps, we recommend using a composite omit map as input to minimize model bias[24]. For cryo-EM modeling applications, equivalent metrics of map and model quality are still developing, rendering the use of qFit for cryo-EM more exploratory.

Since our previous paper, we have made several modifications to the code, both algorithmically (e.g. scoring now includes BIC, and sampling of B-factors) and computationally (improving the efficiency and reliability of the code). All code and associated documentation can be found in the qFit GitHub repository (*https://github.com/ExcitedStates/qfit-3.0*). The version of qFit associated with this paper is 2024.2 and is available at SBGrid (*https://sbgrid.org/*)[25].

A - qFit residue

For each residue, qFit samples backbone conformations, side-chain dihedral angles, and B-factors(**Figure 1A**). Using mixed quadratic programming optimization (MIQP) and Bayesian information criterion (BIC), we select a parsimonious multiconformer for each residue . The details of each component of this procedure are outlined below. The sampling and scoring of residues can be run in parallel using Python multiprocessing.

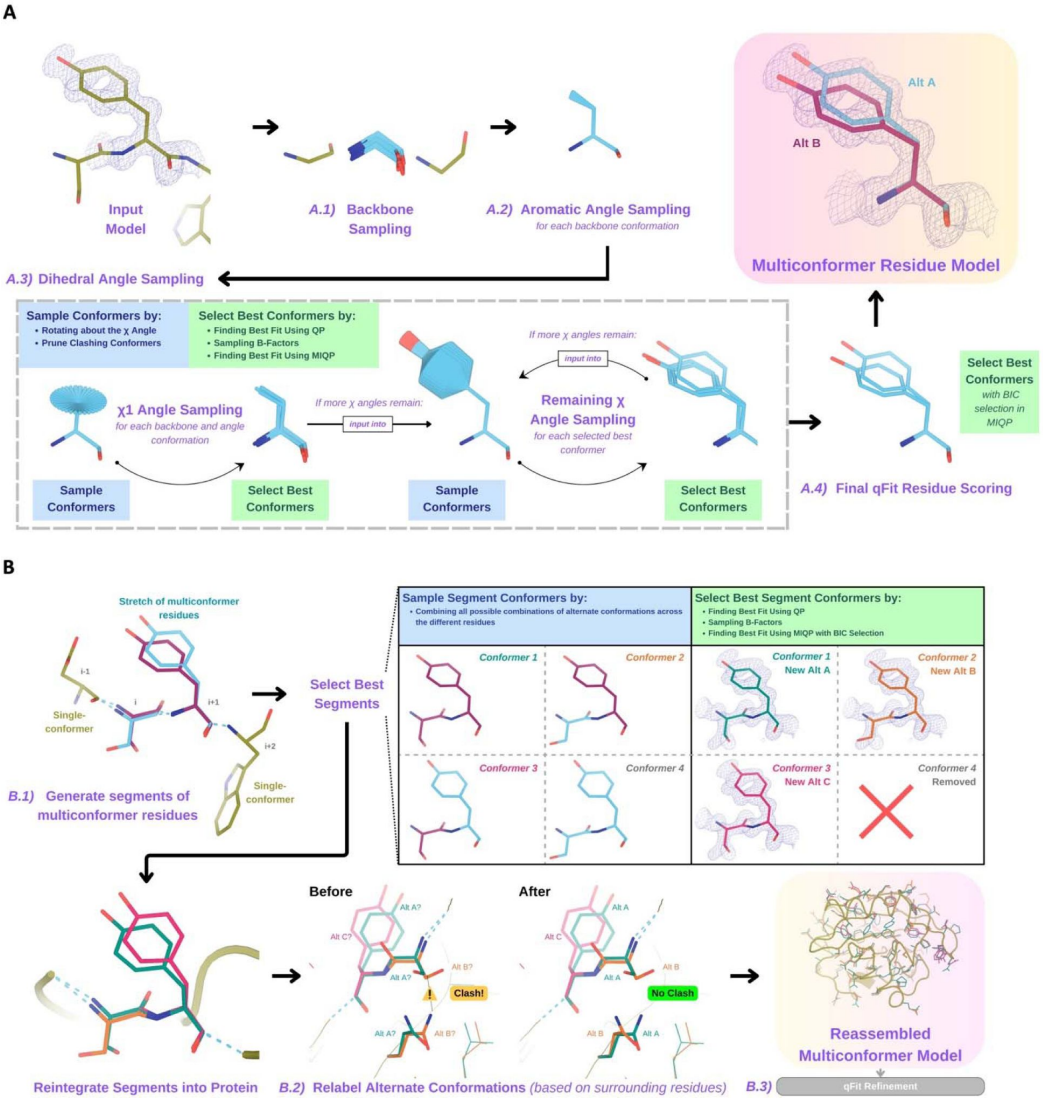

A
Input Model
A.1) Backbone Sampling
A.2) Aromatic Angle Sampling
for each backbone conformation
Alt A
Alt B
Multiconformer Residue Model
A.3) Dihedral Angle Sampling
Sample Conformers by:
• Rotating about the χ Angle
• Prune Clashing Conformers
Select Best Conformers by:
• Finding Best Fit Using QP
• Sampling B-Factors
• Finding Best Fit Using MIQP
χ1 Angle Sampling
for each backbone and angle conformation
If more χ angles remain:
input into
Remaining χ Angle Sampling
for each selected best conformer
Sample Conformers
Select Best Conformers
Sample Conformers
Select Best Conformers
Select Best Conformers
with BIC selection in MIQP
A.4) Final qFit Residue Scoring
B
Stretch of multiconformer residues
Single-conformer
Single-conformer
B.1) Generate segments of multiconformer residues
Select Best Segments
Sample Segment Conformers by:
• Combining all possible combinations of alternate conformations across the different residues
Select Best Segment Conformers by:
• Finding Best Fit Using QP
• Sampling B-Factors
• Finding Best Fit Using MIQP with BIC Selection
Conformer 1
Conformer 2
Conformer 3
Conformer 4
Conformer 1 New Alt A
Conformer 2 New Alt B
Conformer 3 New Alt C
Conformer 4 Removed
Before
After
Clash!
No Clash
Reassembled Multiconformer Model
Reintegrate Segments into Protein
B.2) Relabel Alternate Conformations (based on surrounding residues)
B.3)
qFit Refinement

Figure 1.

Programmatic flow of qFit protein algorithm.

A. qFit residue algorithm, demonstrated by Tyr118 in the E46Q mutant structure of the photoactive yellow protein from *Halorhodospira halophila* (PDB: 1OTA)[29]. The 2mFoCDFc composite omit density map contoured at 1 σ is shown as a blue mesh.

A.1. *Backbone sampling:* For each residue, qFit performs a collective translation of backbone atom (N, C, Cα, O) coordinates.

A.2. *Aromatic angle sampling:* For aromatic residues (His, Tyr, Phe, Trp), qFit takes the conformations from the backbone step and samples the Cα-Cβ-Cγ angle.

A.3. *Dihedral angle sampling:* Since Tyr has two χ angles, qFit starts by taking the output conformers from the aromatic angle sampling step and exhaustively samples the χ1 angle, scoring the best conformations based on QP/B-factor/MIQP scoring. qFit then uses these best conformations as input to sample the remaining χ angles in the Tyr residue. Since the only angle left to be sampled is the χ2 angle, qFit rotates about the terminal ring of the Tyr and then scores the conformations that best fit the density.

A.4. *Final qFit residue scoring:* Once we reach the terminal ring (all sampling steps have occurred), we perform QP and B-factor sampling, followed by MIQP with BIC selection. MIQP with BIC selection removes a redundant overlapping conformation, resulting in two distinct conformations of this Tyr residue. This model is then output as the residue multiconformer.

B. qFit segment algorithm, demonstrated by Tyr118 in PDB: 1OTA. After identifying all optimal conformations for each individual residue, qFit works to connect the protein back together.

B.1 qFit segment: Moving linearly along the protein sequence, qFit identifies 'segments' of residues with multiple backbone conformations. Here, Ser117 (i) and Tyr118 (i+1) have multiple backbone conformations. qFit segment enumerates each possible combination of alternate conformations between these two residues, creating four possible combinations. The optimal combination of conformations is then determined by the QP/MIQP scoring, leading to one combination being culled.

B.2 qFit relabel: qFit uses Monte Carlo optimization with a steric model to assign altloc labels to spatially coupled alternative conformers. In this example, Ser117 and the neighboring Gln32 initially have clashing altloc B conformers. However, relabeling swaps the A and B labels of Gln32 to relieve this clash.

B.3 qFit refinement: We then refine the occupancies, coordinates, and B-factors of the raw qFit output file to produce a final qFit model.

A.1 - Backbone sampling

The qFit process begins with sampling backbone conformations (**Figure 1A.1**). We first strip all hydrogens. For each residue, we perform a collective translation of backbone atom (N, C, Cα, O) coordinates. If the model has anisotropic B-factors, this translation is guided by the anisotropic B-factors of the Cβ. If anisotropic B-factors are absent, the translation of coordinates occurs in the Cα-Cβ, C-N, and (Cβ-Cα × C-N) directions.

Each translation takes place in steps of 0.1 Å along each coordinate axis, extending to 0.3 Å, resulting in 9 (if isotropic) or 81 (if anisotropic) distinct backbone conformations for further analysis. For Gly and Ala, this is the only sampling that occurs.

A.2 - Aromatic angle sampling

For aromatic residues (His, Tyr, Phe, Trp), qFit takes the conformations from the backbone step (above) and builds part of the side chain out to Cγ (start of the aromatic ring) based on the input model coordinates (**Figure 1A.2**). Then, we alter the Cα-Cβ-Cγ angle ("the aromatic angle") in steps of +/- 3.75°, extending to +/- 7.5°, creating 5 partial side-chain conformations per backbone conformation. For non-aromatic residues, there is no sampling of this angle. These conformers provide variability in the placement of the aromatic ring prior to dihedral angle sampling.

A.3 - Dihedral angle sampling

The following steps occur for each L dihedral angle for every residue (**Figure 1A.3**). For the first dihedral angle (L1), the input is the sampled backbone conformations (or for aromatic residues the backbone and "aromatic angle" conformers described above). We sample around the L1 dihedral angle by enumerating a conformation every 6° for for 24° on each side of an idealized rotamer angle. rotamer +/- around each rotamer. For proline, we sample the exo and endo conformations of the pyrrolidine ring, by +/- 24° in steps of 6°. We then eliminate conformations that clash (with other parts of the same sampled conformation of heavy atoms (based on hard spheres) or are redundant (using an all-atom RMSD threshold of 0.01LÅ).

These sampled conformations are then subjected to a quadratic programming (QP) optimization[26], which identifies the set of conformations whose weighted calculated density best fits the experimental electron density. The output of QP typically yields 5 to 15 conformations that best explain the density.

Next, qFit samples the B-factors of the conformers. The input atomic B-factors are multiplied by a factor ranging from 0.5 to 1.5 in increments of 0.2. The resulting 50 to 150 conformation/B-factor combinations are subjected to a mixed-integer quadratic programming (MIQP) optimization. The MIQP algorithm incorporates two additional constraints relative to QP: a cardinality term, which limits the maximum number of conformations to five, and a threshold term, which stipulates that no individual conformation can have an occupancy weight below 0.2. In qFit, MIQP then outputs up to 5 conformations.

For residues with subsequent dihedral angles, the conformations selected by the MIQP procedure at the L(n-1) angle serve as the starting conformers for sampling the L(n) angle. For residues with only one dihedral angle (Ser, Cys, Thr, Val, Pro), we proceed directly to scoring L1.

A.4 - Final qFit residue Scoring

Upon reaching the terminal dihedral angle, we perform the optimization steps outlined above (QP/MIQP), but instead of relying only on the optimization algorithm to decide on the number of conformations to output, we also consider the model complexity (**Figure 1A.4**). qFit runs the MIQP step 5 times with a cardinality term ranging from 1 to 5.

Taking each output, we calculate the Bayesian information criterion (BIC). The BIC provides a numerical value of the tradeoff between the difference between the calculated and experimental density (residual sum of squares) and the number of parameters (k). The number of parameters (k) is defined by the following: number of conformers * number of atoms * 4 (representing the x, y,

z coordinates and B-factor). A heuristic scaling factor of 0.95 accounts for the fact that the coordinate parameters are not independent due to chemical constraints between atoms during sampling.

$$BIC = n * \ln(rss / n) + k * \ln(n) * \text{scaling factor}$$
$$k = \text{number of conformers} * \text{number of atoms} * 4$$
$$rss = \text{residual sum of squares}$$
$$n = \text{number of voxels in density map}$$
$$\text{scaling factor} = 0.95$$

qFit then outputs the set of conformations with the lowest BIC value, concluding the qFit residue routine.

1. B. *Connecting residues together into a multiconformer model*

After the sampling and scoring of each individual residue, qFit considers the entire protein together. First, we use MIQP and BIC to select the best fitting conformations among connected residues, ensuring that neighboring backbone conformations have the same occupancy. Second, we label the alternative conformers while being aware of clashes.

B.1 - qFit segment

After identifying the optimal conformations for each residue in parallel, qFit reconnects the backbone atoms (**Figure 1B.1**). Moving from N- to C-terminus along the protein, we identify 'segments' of residues with multiple backbone conformations, delimited on each end by a residue with a single backbone conformation. The main reason for this step is to find a harmonious set of occupancies for adjacent residues in a segment. Within each segment, qFit creates fragments of three residues, enumerating all possible combinations of conformations in those residues, and selects the final combination of conformations and their relative occupancies using the optimization algorithms outlined above. The BIC is modified for qFit segment such that k equals the number of conformations. qFit then moves along the protein, enumerating and selecting optimal combinations of fragment conformations until reaching the end of the segment.

B.2 - qFit relabel

Next, qFit determines the correct altloc labeling (A, B, C, D, E) of coupled alternative conformers using Monte Carlo optimization with a simple steric model of heavy atoms to prevent spatially adjacent conformers from sterically clashing (**Figure 1B.2**). There is also an option ('qFit segment only') to input a multiconformer model and run only the qFit segment and relabel procedures. This procedure can be especially helpful after manually adding or deleting conformations in Coot[12]. Running 'qFit segment only' will adjust the occupancy of the remaining conformations and correct the labeling of alternative conformations. This labeling step is not parallelized.

B.3 - qFit refinement

The raw output of qFit (a multiconfomer model) should then be refined. We provide scripts for a refinement procedure with Phenix[27], where we iteratively refine the occupancy, coordinates, and B-factors, removing conformations with occupancies under 10%. Once the model is stable (has no conformations with occupancies less than 10%), we perform a final round of refinement which optimizes the placements of ordered water molecules (**Methods**). We then apply a mosaic bulk solvent (*phenix.mosaic*) to the final model, which allows for partial bulk solvent occupancy[28]. This refinement protocol outputs a final 'qFit model'. This model can then be examined and edited in Coot[12] or other visualization software, and further refined using software such as phenix.refine, Refmac, or Buster as the modeler sees fit.

qFit improves overall fit to data relative to deposited structures

To evaluate the impact of qFit algorithmic and code improvements, we collated a dataset of single-chain, unliganded, high-resolution (1.2-1.5 Å) protein X-ray crystallography structures from the PDB[30]. We clustered these structures at a sequence identity threshold of 30% and selected the highest-resolution structure per cluster.

Finally, we ensured that the datasets ran without error through the qFit pipeline, including refinement with Phenix, resulting in 144 diverse structures (**Supplementary Figure 1**).

Each deposited structure was initially re-refined using *phenix.refine* (**Methods**) to eliminate differences from the original refinement protocols. The resulting re-refined model, which we refer to as the 'deposited model', was used as the input for qFit. Next, we ran qFit protein using the default parameters and refinement protocol to produce the 'qFit model'.

To evaluate the crystallographic modeling differences between the *deposited* and *qFit* models, we compared the R_{free} values as an indicator of overall model/data agreement. The qFit model has a lower (improved) R_{free} value for 76% (109/144) of structures (**Figure 2A**; **Supplementary Figure 2A**; Supplementary Table 1). On average, there is an absolute decrease of R_{free} value by 0.6% (median deposited models R_{free}: 18.1%, median qFit models R_{free}: 17.5%), which is in line with theoretical expectations for the increase in model complexity created by qFit[31,32]. R_{free} is a valuable metric for monitoring overfitting, which is an important concern when increasing model parameters as is done in multiconformer modeling. An additional check on overfitting comes from monitoring R-gap, calculated as the difference between R_{work} and R_{free}. qFit models have similar R-gap values compared to deposited models (mean: 3.0% for both models). Collectively, these results indicate that qFit improves the quality of most models without overfitting (**Supplementary Figure 2B**).

Despite this general trend of improved models, 24% of the qFit models have worse R_{free} than the deposited models (n=35). The majority of these structures had a deposited model R_{free} of over 20%. These high R_{free} values are notable because our re-refinement procedure generally improved R_{free} relative to the originally deposited model, particularly for structures with higher starting R_{free} (**Supplementary Figure 2C**). Since qFit builds off of the input structure and the map quality relies on model phases, accurately detecting alternative conformers depends heavily on the agreement between input model and data. This trend reinforced the idea that poor modeling in a deposited model, which serves as input to qFit, will result in poor performance of qFit. It further suggests that qFit is best employed at a late stage of modeling, after the single-structure model is of sufficient quality that it would be deposited in the PDB.

As an example of how qFit can uncover previously unnoticed conformational heterogeneity, we examined differences in conformations in the deposited versus qFit models of the *Pyrococcus horikoshii* fibrillarin pre-rRNA processing protein (PDB: 1G8A)[33]. We focused on the residues

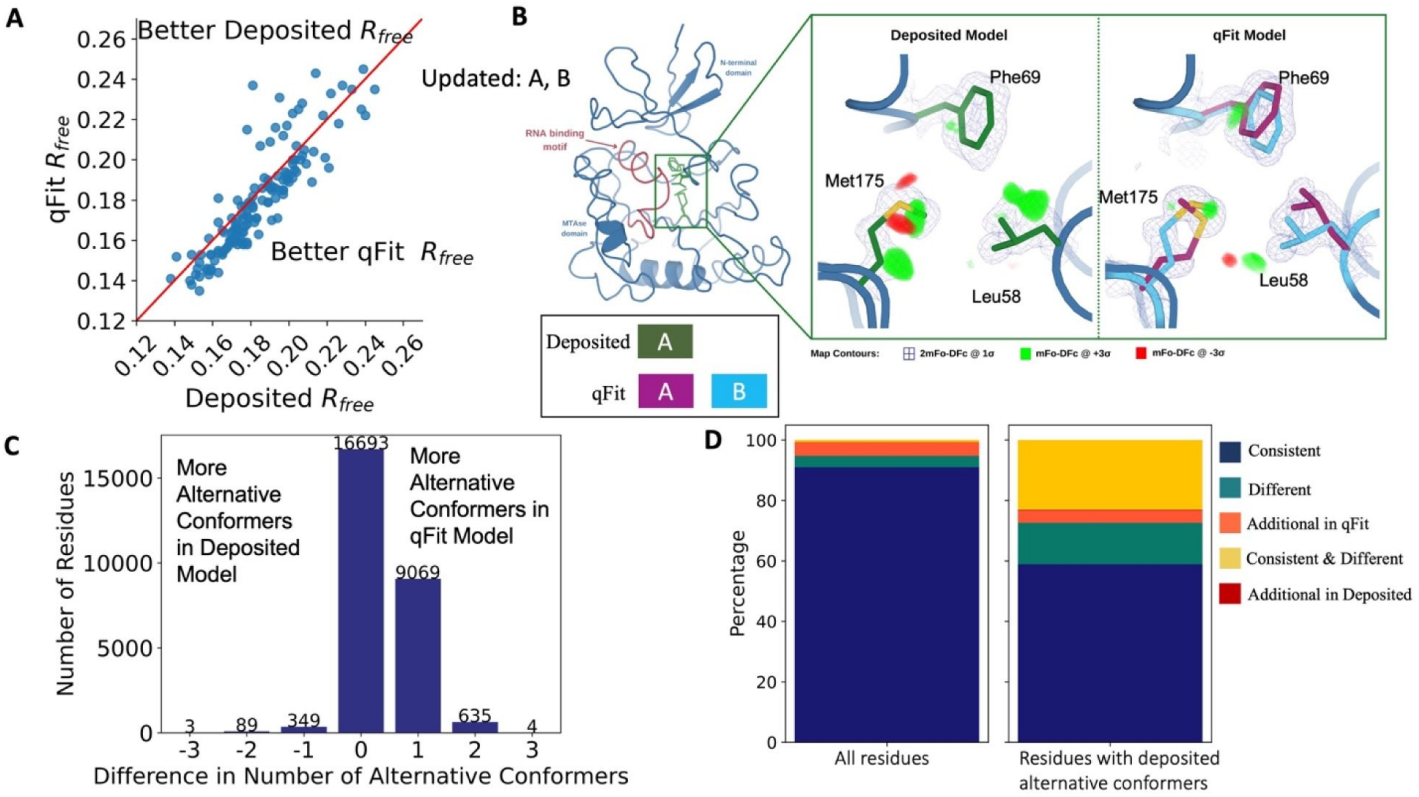


Figure 2.

Multiconformer models created by qFit are better models than deposited single- conformer models.

A. The distribution of R_{free} value in deposited models versus qFit models. The qFit R_{free} values improve in 73% of structures.
B. qFit identifies new alternative conformations adjacent to the RNA binding motif in the *Pyrococcus horikoshii* fibrillarin pre-rRNA processing protein (PDB: 1G8A). **(Left)** qFit multiconformer model with the region in the right panel highlighted in green and the adjacent RNA binding motif highlighted in red. Key domains in the fibrillarin protein are also annotated in blue. **(Right)** Comparison of the deposited versus qFit model in a region with several conformationally heterogeneous residues. qFit identified new rotamers for Leu58 (tp) and Met175 (ttp and mtp)[36] and significantly different alternative conformations within the original rotameric well for Phe69.
C. The differences in the number of alternative conformations per residue in deposited models versus qFit models. qFit adds at least one additional alternative conformation in 31.7% of residues (n=9,998).
D. The distribution of rotamer assignment agreement between the deposited and qFit models for different (sub)sets of residues. **(Left)** All residues (n=42,626). **(Right)** Only residues with alternative conformations in the deposited model (n=970). See main text for definitions of categories.

adjacent to the RNA binding motif. Among these residues, qFit identified well-justified alternative conformations for residues Leu58, Phe69, and Met175, including new rotamers for Leu58 and Met175, that were not present in the deposited model (**Figure 2B**). Beyond detecting alternative conformers in each of these residues, the qFit labeling process identified potential coupled motions between the alternative conformers. For example, when Leu58 is in the 'up' position (altloc A), Phe69 is also in the 'up' position (altloc A). It is possible that this coupled motion plays a role in RNA binding, a hypothesis that may merit further investigation.

qFit recovers alternative conformations of deposited models and discovers new ones

As qFit mainly alters structures by adding alternative conformations, we examined the differences in the number of alternative conformations between the deposited models and qFit models. Only 2.9% of residues in the deposited models were multiconformers (two or more alternative conformations, n=970). In contrast, 40.7% (n=11,049) of residues in the qFit models were multiconformers (**Figure 2C**). The vast majority (92.5%) of multiconformer residues in the qFit models have only two alternative conformations; only 2.4% of residues have more than two alternative conformations.

Alternative conformations come in a few varieties. First and most obvious are alternative conformations that represent drastic changes in coordinates, most commonly in the form of rotameric changes. Most alternative conformations found in deposited models fall into this category. Second are more subtle changes in side-chain and backbone coordinates to represent heterogeneity within a rotameric state. This behavior is exemplified by the Tyr residue in **Figure 1A**. Third is even more subtle changes in coordinates to avoid strain because of the alternative conformations of neighboring residues[34]. This category is essentially imperceptible to visual inspection, as the atom centers are nearly superimposable, but is important to avoid outlier bond geometry because of adjacent residues having larger displacements.

To quantify how often qFit models new rotameric states, we analyzed the qFit models with *phenix.rotalyze*, which outputs the rotamer state for each conformer(**Methods**)[35,36]. We classified the agreement between the deposited and qFit models into 5 categories (**Figure 2D**, **Supplementary Figure 3**). The first category contains residues that have the same rotameric state(s) in both models. This category entails most single-conformer and multiconformer residues with agreement between the two models. Moreover, residues that have multiple conformations in the same rotamer in the qFit model (for the reasons described above) generally populated the same rotamer as found in single- conformer residues in the deposited models. Overall this category, "Consistent", represents 93.7% of residues (n=42,626) in the dataset.

The second and third categories deal with imbalance in alternative conformations that populate distinct rotamers. Since the original premise of qFit was to discover unmodeled alternative conformations, it is unsurprising that many residues in qFit models populate additional rotameric states that are absent in the deposited model. This category, *Additional Rotamer(s) in qFit model*, represents 2.38% of residues (n=1,082). In contrast, only two residues (0.06% of the dataset) are classified in the converse category, *Additional Rotamer(s) in deposited model*.

The final two categories cover disagreements in rotamer assignments. There are many cases where we observe only partial agreement between alternative conformers modeled in both the deposited and qFit models. These multiconformer residues share at least one common rotamer, but also populate alternative rotamers that are distinct between the two models. This behavior generally occurs in longer residues where subtle differences at higher L angles leads to distinct rotameric assignments. This category, *Consistent & Different Rotamers*, represents 0.82% of residues (n=373). The final category, *Different*, covers both multiconformer and single-conformer

residues where there are no shared rotamer states between the two models. One reason this category occurs is for similar reasons as the *Consistent & Different* category: differences in terminal L angles in weak density lead to distinct rotamer assignments.

Another contributor to this category is single conformers, generally in the deposited model, modeled into density that qFit interprets as multiconformer. Often the rotamer modeled by the single conformer fits an "average" rather than the two distinct minima fit by the multiconformer model. *Different* rotamer assignments represent 3.04% of residues (n=1,384). While the analyses above include all residues, focusing on residues that were modeled in as multiconformers in the deposited models (n=970) reveals a large increase in the *Different* and *Consistent & Different Rotamers* categories, to 14.88% (n=144) and 27.68% (n=268) of residues, respectively. This increase highlights the sensitivity of the rotamer assignments and motivates benchmarking qFit on "true positive" synthetic data in addition to deposited multiconformers.

Collectively, these analyses revealed that qFit identifies the majority of deposited alternative conformations and discovers new ones. Discrepancies between manually modeled and qFit alternative conformations predominantly result from weak density at terminal L angles. When considered with the improvements in R_{free}, these results indicate that qFit is detecting more of the true underlying conformational heterogeneity that exists in crystallographic data.

qFit improves multiple side-chain model geometry metrics

Although qFit improves the agreement of model to data by the addition of alternative conformations, we questioned whether this improvement comes at the cost of degrading model geometry. On one hand, the absence of geometric constraints in qFit backbone residue sampling and the connections made during qFit segment may result in worse geometry. On the other hand, placing additional alternative conformers may alleviate strain in the model that can result from fitting a single conformer into density that should be supported by multiple conformers[11,19,37].

To validate geometry, we used MolProbity to evaluate the deposited and qFit models. MolProbity compares input models with idealized values and then provides component scores for various geometric and steric features that are summarized in an overall "MolProbity score"[38]. Component scores that examine all atoms (bond angle/length, clashscore) or side-chain atoms (rotamers) account for all alternative conformers. In contrast, scores that evaluate the backbone (Ramachandran, Cβ deviations) are reported for single-conformer residues or using only altloc A for multiconformer residues. Therefore, the overall MolProbity score includes some of the contributions of alternative conformations, but also misses the potential impact on some other aspects. In the future, we aim to explore updated metrics that consider all alternative conformations.

Compared to deposited models, qFit models had improved MolProbity scores (1.27 median deposited versus 1.09 median qFit, p=0.006 from two-sided t-test; **Figure 3A**), which indicated that overall qFit improves the geometry while also usually improving fit to data. To further understand which parts of the model geometry were different (if any) between the deposited and qFit models, we explored the individual component scores, and observed multiple component scores that improved in the qFit models. This included considerable improvements in bond lengths and angles in the qFit models (RMSD between idealized values for bond lengths: 0.010 Å median deposited vs. 0.007 Å median qFit, p=0.021 from two-sided t-test; RMSD between idealized values for bond angles: 1.30° median deposited vs. 0.91° median qFit, p=3.79e-16 from two-sided t-test; **Figure 3B,C**). We suspect that the primary factor behind this improvement was the incorporation of multiconformers, rather than straining a single conformer, to explain the density. To visualize an example of these differences, we investigated Met189 from PDB: 1V8F. In the deposited model this residue has Sδ-Cε bond lengths of 1.596 Å, which are significantly shorter than the idealized lengths of 1.791 +/- 0.025 Å[38]. qFit adds an additional conformation, both

explaining previously unmodeled density and bringing the Sδ-Cε bond lengths much closer to the expected values: 1.790 Å (alternative conformer A) and 1.794 Å (alternative conformer B) for the two conformations (**Figure 3E**). This multiconformer residue with improved geometry is consistent with the hypothesis that qFit is alleviating strained geometry by modeling multiple conformations.

Additionally, qFit models have improved clashscores (2.50 median deposited, 1.80 median qFit, p=0.0028 from two-sided t-test; **Figure 3D**). We hypothesized that this was due to a mixture of modeling of alternative conformers and improved fit of single- conformer residues which are re-sampled and refined during the qFit procedure. We looked at the qFit modeling differences in a cluster of Met and Leu residues in PDB: 6HEQ, which had one of the largest changes in clashscores between the deposited and qFit models. We observed that qFit fixes the positioning of Met83, preventing the clash with both conformers of Leu81 and improving the local fit to density (**Figure 3F**).

We observed almost equivalent rotamer scores, favored Ramachandran values, and C- beta values (median number of rotamer outliers: 0.94 deposited vs. 0.800 qFit; percentage of Ramachandran favored: 97.7% deposited vs. 97.8% qFit; median value of clashscore: 2.50 deposited vs. 1.78 qFit) (**Supplementary Figure 4**). Overall, the MolProbity scores suggest that qFit improved the model geometry, aligning with improved model/data agreement.

Simulated data demonstrates qFit is appropriate for high-resolution data

In the previous sections, we established that qFit has the potential to improve R_{free} and some geometry metrics relative to deposited structures. However, the vast majority of the residues in these deposited structures are modeled exclusively as single conformers. This homogeneity in single-conformation models limited our ability to assess how well qFit can recapitulate existing alternative conformers across a wide resolution range. To address this question, we generated artificial structure factors using an ultra-high-resolution structure (0.77 _) of the SARS-CoV-2 Nsp3 macrodomain (PDB: 7KR0)[39]. This model had a high proportion of residues (47%) manually modeled as alternative conformations and did not employ qFit during model building or refinement, making it an ideal comparison structure. We refer to this structure as the “ground truth 7KR0 model”, and evaluated how well its alternative conformations were recapitulated by qFit as resolution was artificially worsened across synthetic datasets.

To create the dataset for resolution dependence, we used the ground truth 7KR0 model, including all alternative conformations, and generated artificial structure factors with a high resolution limit ranging from 0.8 to 3.0 _ (in increments of 0.1 _). We then added random noise to the structure factors that increased as resolution worsened (**Methods; Supplementary Figure 5A** and **5B**). To create a single-conformer model appropriate for input to qFit, we removed all alternative conformations from the ground truth model, maintaining all single conformations and altloc A. Next, we refined this single-conformer model against the synthetic datasets. Finally, we used the refined single-conformer model as input for qFit.

We then turned to evaluate the fidelity of qFit in recapitulating the ground truth 7KR0 model. For each residue, we first classified the residue as being a multiconformer or single conformer. Due to many residues in both the ground truth and qFit models having alternative conformations that nearly overlap each other, we categorize residues as multiconformer only if they possess at least two alternative conformers with a side-chain heavy-atom root-mean-square deviation (RMSD) greater than 0.5 _. From this cutoff, 50 out of the 169 residues (30%) in the ground truth model are classified as multiconformers.

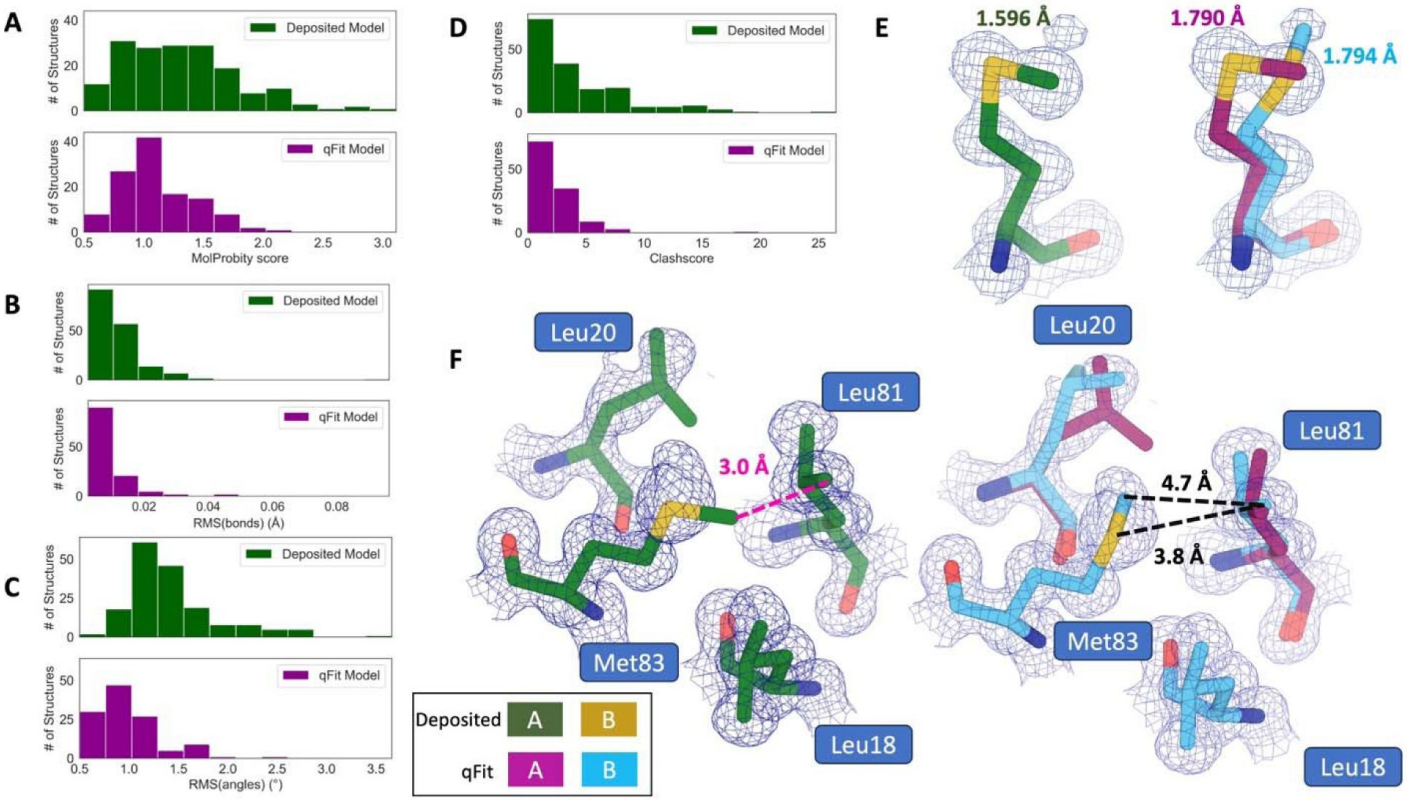


Figure 3.

qFit improves some geometry metrics compared to deposited structures.

A. Model MolProbity score (deposited model: 1.27 (median) [0.94-0.16] (interquartile range), qFit model: 1.09 (median) [0.90-1.30] (interquartile range)), p-value=0.006 from two-sided t-test.
B. Model averaged RMSD (Å) of idealized versus model bond lengths (deposited model: 0.010 [0.0070- 0.015], qFit model: 0.0073 [0.005-0.011]), p-value=0.002 from two-sided t-test.
C. Model averaged RMSD (Å) of idealized versus model bond angles (deposited model: 1.30 [1.14-1.57], qFit model: 0.91 [0.77-1.13]), p-value=3.79e-16 from two-sided t-test.
D. Model clashscore (deposited model: 2.50 [1.30-5.92], qFit model: 1.80 [1.31-3.73]), p-value=0.0028 from two-sided t-test.
E. Example of qFit (right, blue and magenta) fixing bond length by appropriately modeling in a second conformation. Meshes represent 2Fo-Fc density at 1 σ. Met189 from deposited structure (PDB: 1VF8; left, green) has a Sδ-Cε bond length of 1.596 Å (7.8 σ from idealized length of 1.791 Å)[38]. qFit models two alternative conformations, filling in unmodeled density, and fixing the Sδ-Cε bond length (1.790 Å for alternative conformation A and 1.794 Å for alternative conformation B).
F. Example of qFit (right, blue and magenta) fixing a clash between Met83 and Leu81 from deposited structure (PDB: 6HEQ). Meshes represent density at 1 σ. In the deposited model (left, green), Met83 is not correctly fitted into density and is clashing with Leu81 (closest contact: 3.0 Å). qFit corrects this by improving the fit of Met83, leading to the closest contact being 3.8 Å.

Next, we define each residue as having an agreement between the outputted qFit model and the ground truth 7KR0 model. If all qFit modeled conformers are within 0.5 _ of the deposited 7KR0 model, we classify it as a match. If not, we classify it as no match. A “multiconformer match” has agreement between multiconformers across ground truth and qFit models; a “single conformer match” has agreement between single conformers in the ground truth and qFit models. Generally, a “multiconformer no match” has extra or distinct conformations in the qFit model; a “single conformer no match” has at least one alternative conformation in the ground truth model that is not present in the qFit model, or discordant single-conformer conformations.

We observed that qFit is consistently strong at capturing single-conformer residues (single conformer match) across resolutions. We did observe a drop off of detecting alternative conformations (multiconformer match) beyond resolutions of ~1.8-2.0 _ (**Figure 4B**; **Supplementary Figure 5C**). This behavior is exemplified by Glu114, which is multiconformer in the ground truth model (**Figure 4C**). At high resolution (1.0 _), qFit correctly models the alternative conformation and this residue is categorized as a multiconformer match. However, as resolution gets worse, qFit begins to mismodel this residue. At 1.8 _ resolution, qFit still models two alternative conformations and has a good fit to density; however, the secondary conformer has an RMSD greater than 0.5 _ away from the ground truth model; consequently this residue is now categorized as a multiconformer no match. Finally, at 2.8 _ resolution, qFit only models a single conformer, moving the residue to the single conformer no match category.

Simulated multiconformer data illustrate the convergence of qFit

Next, we tested the ability of qFit to detect alternative conformations over a larger, more diverse dataset. We generated artificial structure factors for the qFit models with improved R_{free} values over the deposited values from the previous sections (n=109).

Although this dataset is more diverse, it has a notable weakness relative to the 7KR0 dataset test: the 7KR0 alternative conformations were modeled manually, whereas the larger dataset has alternative conformations modeled by qFit. Therefore, this second synthetic dataset assesses convergence of the qFit models across resolution.

Using these qFit models as ground truth models, we generated structure factors, performed refinement of single-conformer models, and ran qFit over the resolution range of 1.0 to 3.0 _ (**Supplementary Figure 5A**). We observed a similar fall-off of multiconformer match residues around 2.0 _ (**Supplementary Figure 5D**). Importantly, this dataset indicates that qFit still models single conformers well at lower resolutions. We also observe a trend of increased no match multiconformers/single conformers for longer residues that are just outside the 0.5 _ RMSD cutoff (**Supplementary Figure 6**). We did not observe a relationship between input model R_{free} and the number of correctly modeled conformers, but it is difficult to tell whether our synthetic noise procedures properly capture the dependence of qFit performance on input model/data agreement (**Supplementary Figure 7A/B**).

We then assessed the agreement between individual conformers and the map. To do this, we used the Q-score[40], which compares the map profile of an atom with an ideal Gaussian distribution that would be observed if the atom perfectly fits into the density. Across the test dataset, residues that qFit models as single conformers have an almost equivalent Q-score to the ground truth model even at lower resolutions (**Figure 4D**).

The primary alternative conformations in qFit models (occupancy between 0.5 and 1.0) and lower-occupancy alternative conformations (occupancy <0.5) display Q-scores that are very close to the equivalent “ground truth model” alternative conformations until a resolution of about 1.8 _. At lower resolutions there is a dramatic fall-off in model/map agreement for these alternative

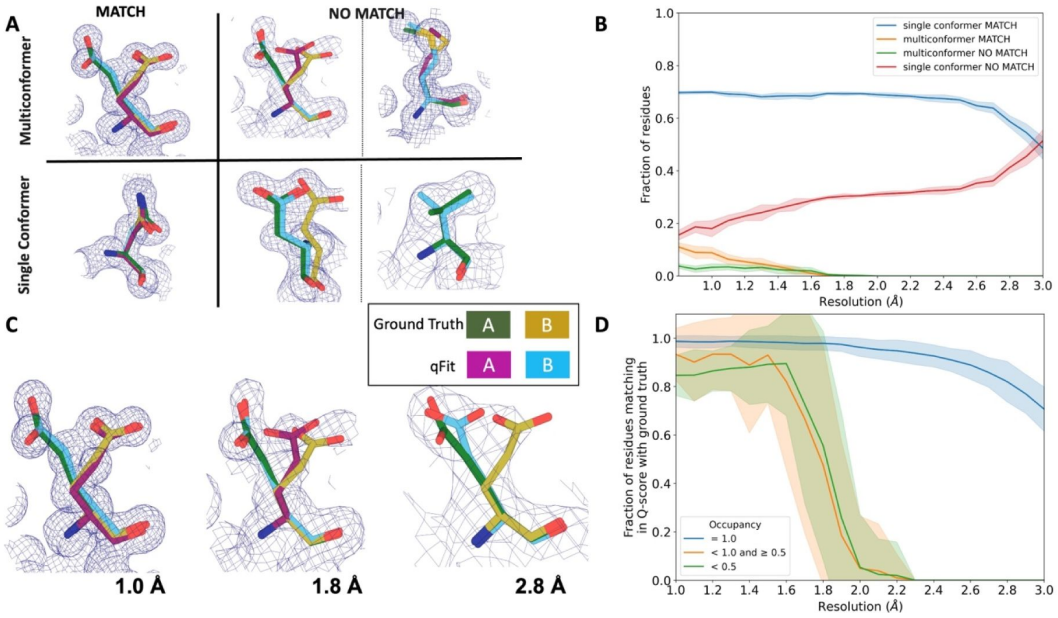


Figure 4

qFit performs best at high resolution of input dataset

A. Ground truth model residues are shown as green and yellow sticks; qFit model residues are shown as magenta, cyan, and gray. Meshes represent density at 1 σ.
Multiconformer match - Residue is multiconformer in qFit model with RMSD < 0.5 Å from ground truth residue. qFit models two distinct alternate conformations which recapitulate the ground truth residue's alternate conformations.
Multiconformer no match - Residue is multiconformer in qFit model with RMSD > 0.5 Å from ground truth residue. The example on the left has two alternate conformations in the ground truth. qFit models only one of them correctly. The example on the right is a single-conformation residue in ground truth but qFit models three alternate conformations.
Single conformer match - Residue is single-conformer in qFit model with RMSD < 0.5 Å from ground truth residue. Both ground truth model and qFit model have one distinct conformation and they align well.
Single conformer no match - Residue is single-conformer in qFit model with RMSD > 0.5 Å from ground truth residue. The example on the left has two alternative conformations in the ground truth residue but only one conformation in the qFit residue. In the example on the right, the single conformer modeled by qFit does not align with the ground truth single conformer.
B. Proportion of all residues in the qFit models of 7KR0 that are modeled as multiconformer match (orange), single conformer match (blue), multiconformer no match(green), and single conformer no match(red) as a function of resolution of input synthetic data from the 7KR0 dataset. The shaded region denotes the 95% confidence interval.
C. Glu114 in the 7KR0 dataset modeled by qFit (cyan and magenta) compared to the ground truth structure (green and yellow) at different synthetic resolutions. Meshes represent density at 1 σ.
D. The fraction of residues in the qFit models of the qFit test dataset with a Q-score within 0.01 to that of the ground truth model as a function of resolution. In multiconformer residues, Q-score for every alternative conformation is calculated separately. Q-scores of residues (or) conformers which have matching occupancy (range) are compared. Occupancies of conformers were binned into three classes – occupancy equal to 1 (blue), 1 > occupancy >= 0.5 (orange) and occupancy < 0.5 (green).

conformers. These trends were also observed with the 7KR0 dataset (**Supplementary Figure 7C**). Overall, these analyses on both the 7KR0 and larger synthetic datasets confirm that qFit will best detect alternative conformations with high-resolution (1.8-2.0 _ or better) data.

qFit models alternative conformers in cryo-EM density maps

As single-particle cryo-EM is increasingly producing high-resolution (better than 2 Å) reconstructions where alternative conformers can be detected[41,42], we wanted to improve and test the ability of qFit to model alternative conformations guided by cryo- EM maps. While a previous version of qFit introduced cryo-EM compatibility[21], we had not optimized the approach to work with cryo-EM maps and models. qFit can now be run in 'EM mode' which uses electron structure factors, improves the treatment of solvent background levels, and reduces the default maximum number of alternative conformations (cardinality) (**Methods**).

To benchmark our ability to model alternative conformations in high-resolution cryo-EM structures, we initially gathered a dataset of 22 structures with a depositor-provided resolution better than 2 Å (Fourier Shell Correlation (FSC) at 0.143). However, only 8 of these structures have a resolution better than 2 Å (FSC at 0.143) when calculated by the Electron Microscopy Data Bank (EMDB)[43]. Some of the original 22 structures did not have FSC curves in EMDB (n=6) due a lack of data, and others had an EMDB calculated resolution worse than 2 Å (n=8) (**Supplementary Table 2**). The absence of standardized maps for determining cryo-EM structure resolution complicated our selection of structures for qFit analysis.

We downloaded the eight models with resolution better than 2 Å from the PDB and their corresponding maps from EMDB. Using the default parameters of *phenix.autosharpen*, we sharpened all maps and re-refined each structure (*phenix.real_space_refine*) against its sharpened map. qFit was run with the 'EM' flag and the output model was refined using the qFit real space refinement script (**Methods**).

Across the first asymmetric unit of the 8 models, 8.21% (n=64) of residues in the deposited model had at least two alternative conformers in the deposited structure, compared with 39.6% (n=266) in the qFit model. To determine if qFit could recapitulate the modeling of alternative conformers from deposited structures, we compared the high-resolution apoferritin deposited model (PDB: 7A4M, resolution: 1.22 Å) with the qFit model using the same criteria outlined in the resolution dependence section above (RMSD within 0.5 Å). qFit correctly models 77% of residues in the first asymmetric unit. This includes Arg22, which has two alternative conformations in the deposited model. qFit was able to recapitulate both alternative conformations (**Figure 5A**), highlighting that qFit can detect manually modeled alternative conformations in cryo-EM maps. In addition, qFit detected several unmodeled alternative conformers that were visually confirmed (**Figure 5B-D**).

As with the X-ray models, we wanted to determine how qFit changes the model geometry. Similar to the X-ray models, we observed that qFit improves bond lengths and angles, and similar Cβ deviations. Unlike the observations in the X-ray dataset, qFit does increase (worsen) the MolProbity score, likely coming from high clashscore of most structures, highlighting a future improvement in the algorithm(**Supplementary Figure 8**).

While we have made significant progress in modeling alternative conformations in cryo- EM data, the lack of consistent map handling, validation, and metrics with cryo-EM structures and maps is a major impediment to further development. Even among this select group of structures, there were varying levels of experimental and computational map details on EMDB and in manuscripts[41,42,44], including information on masking, handling of bulk solvent, and local resolution. Our approach depends on sampling and scoring based on resolution. While there is an accepted formula for calculating resolution (FSC at 0.143), the maps to calculate these are not consistent leading to differences in resolution as we observed between the deposited versus EMDB

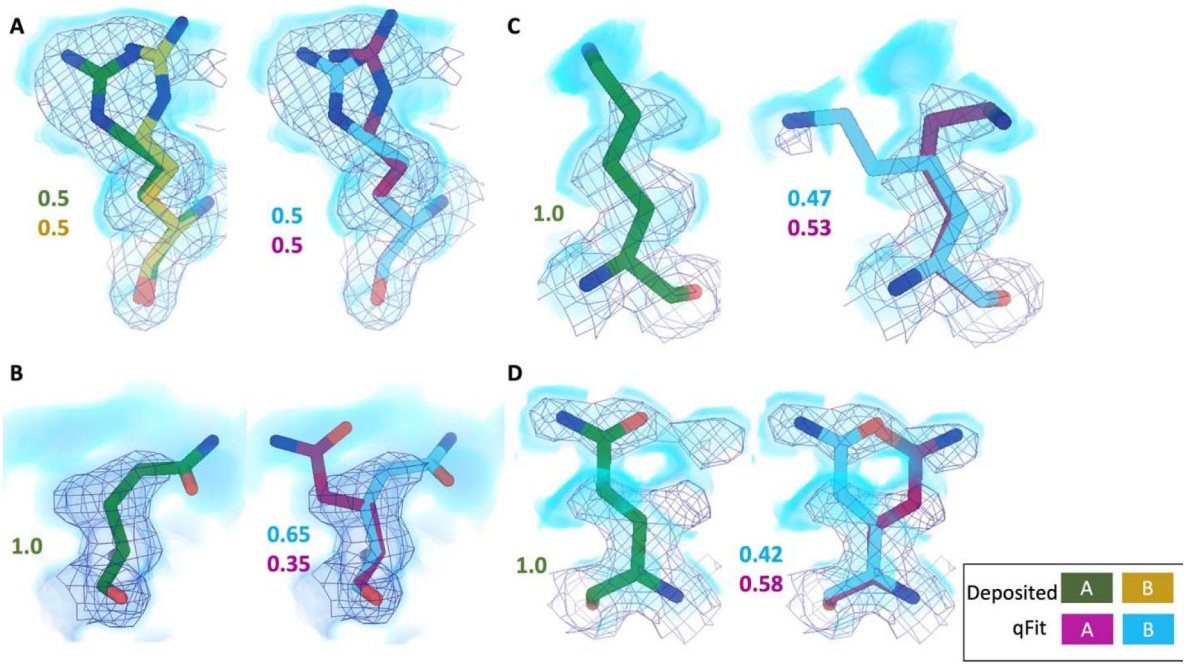


Figure 5.

qFit identifies alternative conformations in high-resolution cryo-EM models.

Meshes represent density at 1 σ, with blue volumes representing density at 0.5 σ. Green and yellow sticks represent deposited conformation(s). Cyan and magenta sticks represent qFit conformations. Occupancy is labeled based on each conformer.

A. qFit recapitulated the deposited alternative conformations of Arg22 (chain A) in apoferritin (PDB: 7A4M, resolution: 1.22 Å).

B. qFit identified a previously unmodeled alternative conformation of Glu14 (chain A) in apoferritin (PDB: 7A4M, resolution: 1.22 Å).

C. qFit identified a previously unmodeled alternative conformation of Lys49 (chain A) in a different structure of apoferritin (PDB: 6Z9E, resolution: 1.55 Å). .

D. qFit identified a previously unmodeled alternative conformation of Gln403 (chain A) in adeno- associated virus (PDB: 7KFR, resolution: 1.56 Å).

calculated resolutions. Further, resolution can vary across a single model, and metrics for such local resolutions are not always widely available. Additionally, the handling of background bulk solvent values varies widely, from masking to flattening these values. New methods for cryo-EM ensemble modeling will benefit from ongoing efforts to standardize the storage of raw, meta, and processed data[45].

Discussion

Structural biology plays a vital role in understanding the complex connection between protein structure and function. However, since proteins exist as ensembles, structural biology modeling approaches need to adapt accordingly. X-ray crystallography and cryo-EM data hold significant information on these ensembles that is often ignored. qFit offers a solution by leveraging powerful optimization algorithms to transform well- modeled single-conformer models into multiconformer models. Here we demonstrate that qFit can uncover widespread conformational heterogeneity that better represents the true underlying conformational ensemble data as demonstrated by lower R_{free} values. Further, we determine that qFit can reliably pick up on alternative conformers that were modeled manually, highlighting that qFit could be used as a tool to significantly speed up modeling of high-resolution structures.

This automation in modeling is needed especially in light of advances in data collection automation and fast detectors. These tools have revolutionized the field of X-ray crystallography, enabling high-temperature datasets, time-resolved experiments, and high-throughput data collection[46–50]. With the ability to capture different conformations, there is a growing demand for methods that can detect protein alternative conformers to extract as much biological information as possible. This is highlighted in massive ligand- soaking campaigns[39,51–53], where there are often hundreds of structures with different ligands to parse. qFit provides a key tool to help extract the most out of these structures by improving the models and providing a better jumping-off point to determine how ligand binding impacts the protein. However, our data here show that not only does qFit need a high-resolution map to be able to detect signal from noise, it also requires a very well-modeled structure as input.

While both throughput and resolution are currently lower for cryo-EM, recent high- resolution maps have observable conformational heterogeneity[41,44]. Current classification approaches do not allow sorting based on signals as small as alternative side-chain conformations[54–56], necessitating approaches like qFit for modeling. We see great potential in combining qFit with classification approaches to understand conformational heterogeneity at different scales. In the future, qFit can likely be applied more widely to EM maps in regions with high local resolution[57]. In addition, we will also incorporate modeling of nucleic acids, with an emphasis on automating refinement of alternative base positions in high-resolution ribosome structures in future work[58–60].

However, we encountered many difficulties in applying qFit to EM data relative to the more established X-ray data. In particular, there are still disparities in how maps are sharpened and how masks are used to exclude noise or lower experimental signals, such as solvent[61,62], making it very challenging to evaluate whether models, especially multiconformer or ensemble models, have improved fit to the data. We suggest strengthening guidelines for reporting computational processing, and improving validation tools to gauge agreement between models and cryo-EM maps[61–63].

We envision many other future improvements that will further enhance the quality and accuracy of multiconformer models for both X-ray crystallography and cryo-EM. Simulations have demonstrated that subpar modeling of the macromolecule(s) and surrounding solvent is a major

potential avenue to further reduce R-factors[31,32]. To accurately account for water molecules in multiconformer models, partially occupied water molecules must be identified and labeled in connection with protein atoms.

Automated detection and refinement of partial-occupancy waters should help improve fit to experimental data[15] and provide additional insights into hydrogen-bond patterns and the influence of solvent on alternative conformations[15].

Additionally, while qFit models have overall improved geometry in some respects relative to single-conformer models, we still have room for improvement for fixing backbone metrics (Ramachandran and Cβ deviations). The geometry improvements are likely mostly due to single-conformer models having strained conformations that fit the "mean" conformation rather than multiple partially overlapping conformations. Further gains in both accuracy and geometry quality will emerge with better sampling of backbone conformations[20]. Such improvements are important because splitting the backbone, where appropriate, can result in detection of biologically important side-chain alternative conformations[64]. Notably, the recently described FLEXR approach, which leverages Ringer and Coot to model alternative side chains into density peaks, illustrates that many gains can be made with side-chain focused modeling alone[19].

However, further improvements to backbone modeling, including larger-scale motions such as alternative loop conformations[65] or coordinated larger-scale shifts of secondary- structural elements[66,67], will likely yield even higher-quality multiconformer models.

Lastly, experimental and computational advancements in structural biology have increased the focus on ensemble-based models[11,21,55,56,68]. But the current data format for structural models (PDB, mmCIF) does not allow for more complex representation of ensembles. qFit is compatible with manual modification and further refinement as long as the subsequent software uses the PDB standard altloc column, as is common in most popular modeling and refinement programs. The models can therefore generally also be deposited in the PDB using the standard deposition and validation process.

However, to even more appropriately capture the many aspects of ensembles, we would ideally like to have multiple nested ensembles representing both larger and local conformational changes, or to be able to show how two different backbone conformations can each be "parents" to different side-chain conformations[69]. Currently, neither the PDB nor CIF format allows for this type of representation[70–72].

In summary, qFit drastically reduces the time and effort required to create multiconformer models from X-ray and cryo-EM data, thereby lowering the barrier to generating new hypotheses about the relationship between conformational ensembles and biological function[4,5,73,74]. Additionally, qFit can provide key data to bridge to the next frontier of structure prediction. While AlphaFold[75] has achieved stunning success in predicting protein structure by training against single-conformation models, future improvements to structure prediction might be gained by more accurately modeling the extent of conformational heterogeneity[76].

Methods

Generating and running the qFit test set

To test the impact of algorithmic changes in qFit, we created a dataset of 144 high- resolution (1.2-1.5 Å) X-ray crystallography structures deposited in the PDB (**Supplementary Table 1**). These were single-chain protein structures (in the asymmetric unit and at the level of biological assembly) and contained no ligands or mutations. The maximum sequence identity between any two structures

was set as 30%. Based on CATH classification[35], the resultant entries represented 72 folds (**Supplementary Table 1**). The structures represented 24 space groups. All these structures were re-refined as described in "Initial refinement protocol". These re-refined models are referred to as deposited models. To create multiconformer models, we input the re-refined structures in qFit protein, followed by the post qFit refinement protocol.

These multiconformer models are referred to as qFit models.

Initial refinement protocol

All structures from the PDB were re-refined using *phenix.refine* with the following parameters:

```
refinement.refine.strategy=*individual_sites *individual_adp *occupancies
refinement.output.serial=5
refinement.main.number_of_macro_cycles=5
refinement.main.nqh_flips=False
refinement.output.write_maps=False
refinement.hydrogens.refine=riding
refinement.main.ordered_solvent=True
refinement.target_weights.optimize_xyz_weight=true
refinement.target_weights.optimize_adp_weight=true
```

The re-refined models were used as the input for subsequent qFit models.

Running qFit

For this analysis qFit was run using the following command from qFit version 2023.1:

X-ray:

qfit_protein composite_omit_map.mtz -l 2FOFCWT,PH2FOFCWT rerefine_pdb.pdb

Cryo-EM:

qfit_protein sharpened_map.ccp4 rerefine_cryo-EM.pdb -r

-em -n 10 -s 5

qFit New Features

Parallelization of large maps

Often, cryo-EM maps are very large and reach memory limits using Python multiprocessing. Multiprocessing is used to model multiple residues independently in parallel. We have now implemented a new scheme to divide the density map into portions centered around each residue of interest and feed those portions of the map into our parallelization.

B-factor sampling

To sample B-factors along with atomic coordinates at each step of qFit residue, we first perform one round of quadratic programming to reduce the number of conformations. For all remaining conformations, the input B-factor of each atom in the residue is multiplied by 0.5-1.5 in increments

of 0.2. All conformations with sampled B-factors and coordinates are inputs for mixed integer quadratic programming.

Bayesian information criterion (BIC)

BIC was implemented in the final selection of residue and segment conformations. BIC is defined as the real space residual correlation coefficient penalized by the number of parameters (k):

$$BIC = n * np.log(rss / n) + k * np.log(n) * 0.95$$
$$rss = residual\ sum\ of\ squares$$
$$n = number\ of\ datapoints$$

In qFit residue, k is defined as:

$$k = 4 * number\ of\ atoms * number\ of\ conformations$$

In qFit segment, k is defined as:

$$k = number\ of\ conformations$$

BIC is calculated for each candidate cardinality (1-5). We then choose the set of conformations with the lowest BIC as the final conformations for the residue or segment under consideration.

Iterative optimization algorithm with non-convex problems

Due to our exhaustive sampling, there are times when the MIQP optimization algorithm fails to find a non-convex solution. To address this limitation, we have implemented a procedure that iteratively removes solutions one-by-one based on the two solutions with the closest root-mean-square deviation (RMSD) until MIQP identifies a solution.

Implementation of open source QP/MIQP algorithms

qFit previously relied on IBM CPLEX to score conformations. While this is free to academics, it is not open source. We have switched to CVXPY, an open source QP and MIQP solver[26,77].

Occupancy constraints

To help refine segments (i.e. sets of residues with alternative conformations flanked by residues with only a single conformation) during X-ray refinement, we now output a restraint file at the end of the qFit protein run for X-ray refinement. This restraint file enables “group occupancy refinement” for residues in a segment with the same alternative conformation. In group occupancy refinement, all residues within the group are refined to the same occupancy, reducing the free parameters to fit.

Finalizing qFit models with iterative refinement

We iteratively run 5 macrocycles of refinement followed by a script that removes any conformations with occupancy less than 0.10. This script also renormalizes the occupancies of any remaining conformations in that segment, ensuring that the occupancy sums to 1. This procedure ends when no conformations have a refined occupancy of less than 0.10 or after 50 total rounds of refinement (whichever comes first). Afterwards, we perform one final refinement where we release the occupancy constraints on the segments, turn on automated solvent picking, and optimize B-factors (specified as ADP parameters in Phenix) and coordinate weights.

Cryo-EM

To improve the detection of alternative conformations in cryo-EM structures, we made some key updates to part of the qFit algorithm. All of these updates to the algorithm will turn on with the *-em* flag. First, we now use electron scattering factors when calculating the modeled electron density. Second, we have removed bulk solvent electron density values (set at 0.3 in X-ray qFit protein). We also restricted the occupancy threshold cardinality to be 0.3 (compared to 0.2 in X-ray qFit protein) to reduce misplaced conformations.

Q-score

We implemented the option for users to use Q-scores to determine if qFit should be run on a residue or not. This option is off by default. To utilize this feature, first generate Q- scores by using the mapq.py script, which is included in the Q-score command-line interface package.(*https://github.com/gregdp/mapq*). qFit takes in a text file of Q-scores by using the *–qscore* option in *qFit_protein*. By default, all residues with a Q-score of less than 0.7 are not modeled as multiconformers, but are considered in qFit segment. Users can also adjust this level by using the *–qscore_cutoff* option in qFit protein.

qFit-segment-only runs

qFit can be used as a tool along with iterative model building and refinement. If a user manually removes or adds additional conformations using Coot[12] or similar software, this can disrupt the occupancy sum of the residue and the connectivity of the backbone. To alleviate such problems, we developed an option (*qfit_protein –only-segment*) to facilitate manual model adjustment after running qFit. This procedure generates connected backbones with consistent occupancies for coupled neighboring conformers.

For example, suppose residue n has four alternative backbone conformations (A, B, C, D) and residue n+1 has two alternative conformations (A, B). In that case, this procedure will create C and D conformers for residue n+1 by duplicating its A and B conformers. This duplication continues until we reach the end of a segment so that all backbones have the same number of alternative conformations (A, B, C, D) and are, therefore, properly connected. Subsequent crystallographic refinement of this model (see “Post-qFit refinement script” above) will cause the duplicated conformations to diverge slightly, and will behave as expected without introducing geometry errors.

Analysis metrics

Scripts for all metrics can be found in the scripts folder in the qFit GitHub repository (*https://github.com/ExcitedStates/qfit-3.0*). Our scripts for running qFit protein on an SGE-based server and all scripts for figures can be found here: *https://github.com/fraser-lab/qFit_biological_testset/tree/main*.

R-values

R-values were obtained after the final round of refinement for the re-refined deposited models (*deposited_rerefine.sh*) and for the qFit models after the iterative refinementscript (*qfit_final_xray_refine.sh*).

B-Factors

For each residue, we calculate an occupancy weighted B-factor (each heavy atom B- factor is weighted by its occupancy). For each heavy atom, we calculate the weighted using the following formula:

$$Occupancy\ Weighted\ B\text{-}factor = Occupancy * (4*pi/B\text{-}factor)^{1.5}$$

Rotamers

The rotamer name for each alternative conformation was determined by *phenix.rotalyze*[38] while manually relaxing the outlier criteria to 0.1%. Rotamers were compared on a residue-by-residue basis. To compare rotamers, we only consider the first two _ dihedral angles. Each residue was classified into four categories: same, additional rotamer in qFit model, additional rotamer in the deposited model, or different.

Generating synthetic data for resolution dependence

To generate artificial electron density data at increasingly poorer resolutions, we first increased the B-factors of all atoms of the ground truth model by 1 $_{}^{2}$ for every 0.1 _ reduction in resolution and placed the models in a P1 box. We randomly shook the coordinates using the *shake* argument in *phenix.pdbtools* with root-mean-square error of shaking given as 0.2 * desired resolution of synthetic data. We generated structure factors (F_{shake}) for each of these shaken models from 0.8 _ to 3.0 _ in increments of 0.1

_ using the *phenix.fmodel* command-line function (with bulk solvent parameters k_sol=0.4, b_sol=45, and 5% R-free flags). We then added noise to the structure factors as follows:

$$F_{noisy} = F_{shake} + (sqrt(F_{shake}) * random\ number\ from\ normal\ distribution * resolution\ of\ model * 0.5)$$

The scaling factors of 0.2 and 0.5 for shake RMSD and noise addition were determined by trying out different values and identifying the values which gave the lowest R_{free} over the resolution range after refining the model against the generated structure factors.

The addition of noise to F_{shake} was done using the *sftools* command in CCP4[78]. Then, the ground truth model with adjusted B-factors was stripped of alternative conformations (if any) at every residue position. The resulting single-conformer model was refined with the F_{noisy} structure factors (**Supplementary Figure 5A**).

The final refined model was given as input to qFit and the composite omit map was obtained for the F_{noisy} structure factors. The multiconformer model given by qFit was refined with *phenix.refine* as explained in the post-qFit refinement script section. Since there is some randomness involved in simulating noise in the synthetic datasets, at each resolution, we generate ten synthetic datasets, and apply the qFit protocol to each one. The same steps of data synthesis were followed for the larger qFit test dataset containing 103 models, except that one set of structure factors was generated for each model at each resolution instead of ten as in the 7KR0 dataset.

Match classifications for synthetic data

Match multiconformer residues were those with at least two alternative conformations and an RMSD of less than 0.5 _ between the ground truth and qFit model conformations (for example, qFit model altloc A has an RMSD of less than 0.5 _ to ground truth model altloc A or B, and qFit model altloc B has an RMSD of less than 0.5 _ to the other ground truth model altloc A or B) (**Figure 4A**). No match multiconformer residues have at least two alternative conformations in the qFit model, but fewer conformations in the ground truth model (**Figure 4A**). Alternatively, for a no match multiconformer residue, if the ground truth model residue is also multiconformer, then the RMSD between at least one of the conformations of qFit residue and ground truth residue is more than 0.5 _ (**Figure 4A**). A match single conformer residue is when both the ground truth and qFit model have a single conformer and they have an RMSD of less than 0.5 _ (**Figure 4A**). A no match single conformer residue is when the qFit model has a single conformer but the ground truth model has more than one alternative conformer or both models have a single conformer but they have an RMSD greater than 0.5 _ (**Figure 4A**).

Data Availability

All qFit models for the PDBs discussed in this paper are included in Zenodo deposition *https://doi.org/10.5281/zenodo.10936292*.

Acknowledgements

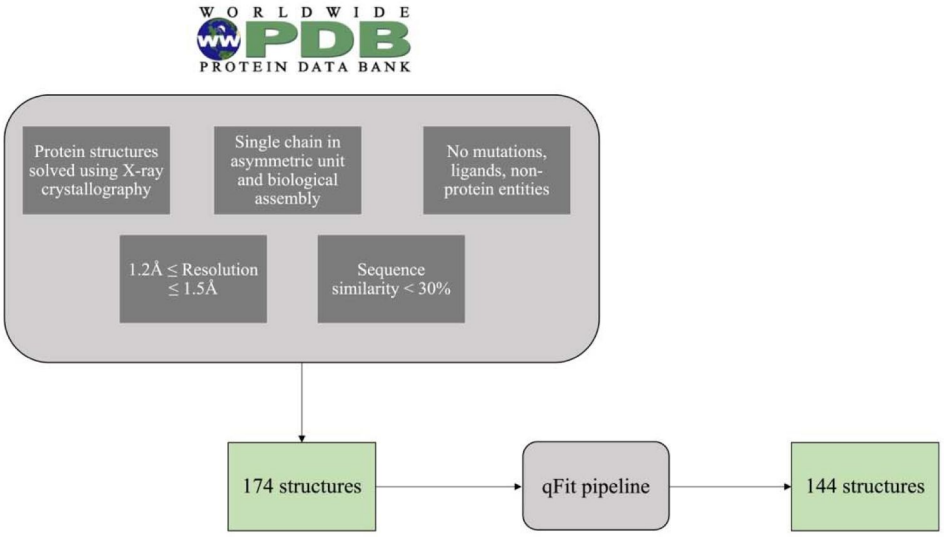
This work was supported by a National Institutes of Health (NIH) grant GM145238 and Chan Zuckerberg Initiative Essential Open Software grant to JSF and NIH R35 GM133769 to DAK. We thank Christopher Williams and Vincent Chen for help with interpretations of MolProbity score ideal side-chain geometry.

References

1. Cheng Y (2015) **Single-Particle Cryo-EM at Crystallographic Resolution** *Cell* **161**:450–457
2. Smith J. L., Hendrickson W. A., Honzatko R. B., Sheriff S (1986) **Structural heterogeneity in protein crystals** *Biochemistry* **25**:5018–5027
3. Herzik M. A., Wu M., Lander G. C (2017) **Achieving better-than-3-Å resolution by single-particle cryo-EM at 200 keV** *Nat. Methods* **14**:1075–1078
4. Keedy D. A., et al. (2018) **An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering** *Elife* **7**
5. Wankowicz S. A., de Oliveira S. H., Hogan D. W., van den Bedem H., Fraser J. S (2022) **Ligand binding remodels protein side-chain conformational heterogeneity** *Elife* **11**
6. Yabukarski F., et al. (2022) **Ensemble-function relationships to dissect mechanisms of enzyme catalysis** *Sci Adv* **8**

Supplementary Figure 1.

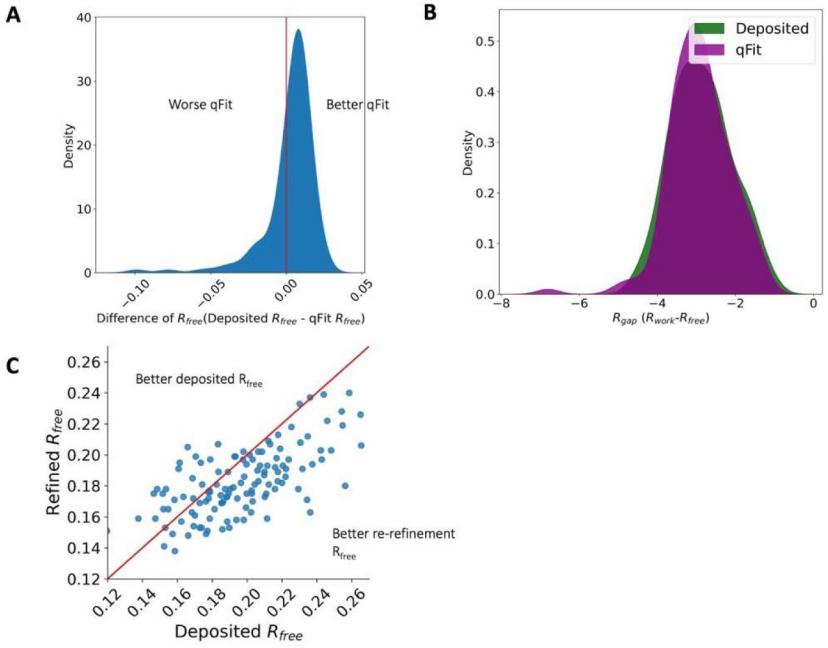
Flow diagram of the selection of the test set PDBs.

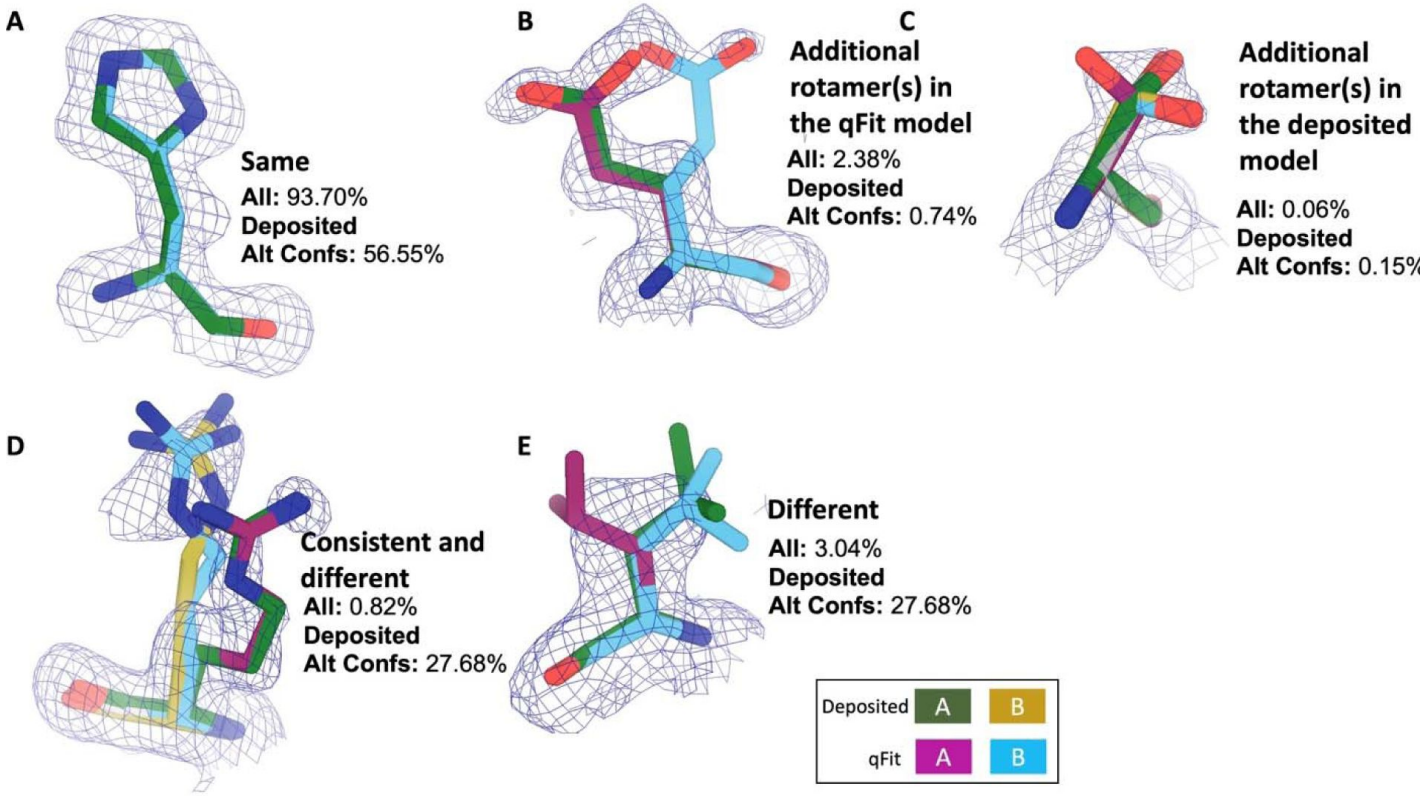


Supplementary Figure 2.

R_{free} and R-gap distributions.

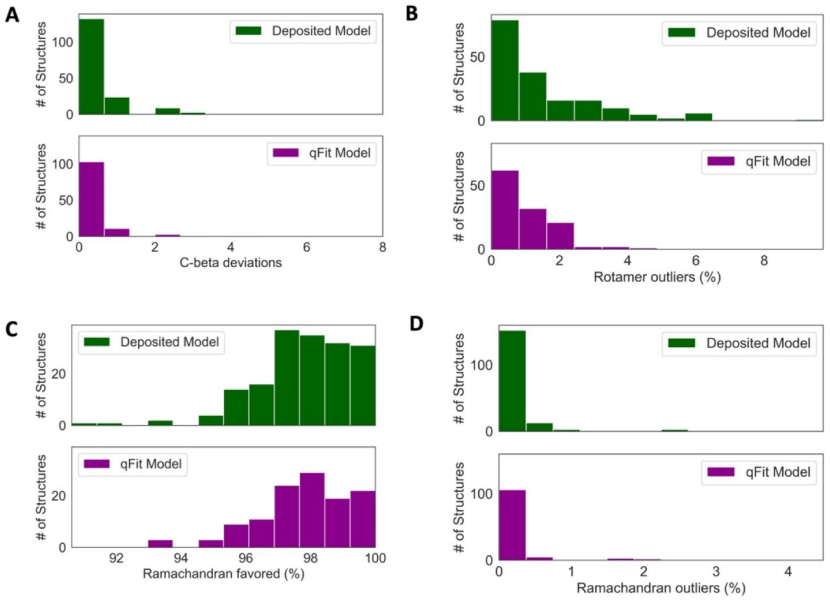
A. Distribution of difference of R_{free} between deposited and qFit models. The median difference in R_{free} is 0.6%. Median deposited models R_{free}: 18.1%, median qFit models R_{free}: 17.5%.
B. Distribution of R-gap values between deposited and qFit models (median deposited model: 3.0%, median qFit model: 3.0%).
C. Distribution of R_{free} value in PDB deposited models versus re-refined deposited models. In this manuscript, deposited models refer to the re-refined deposited models.





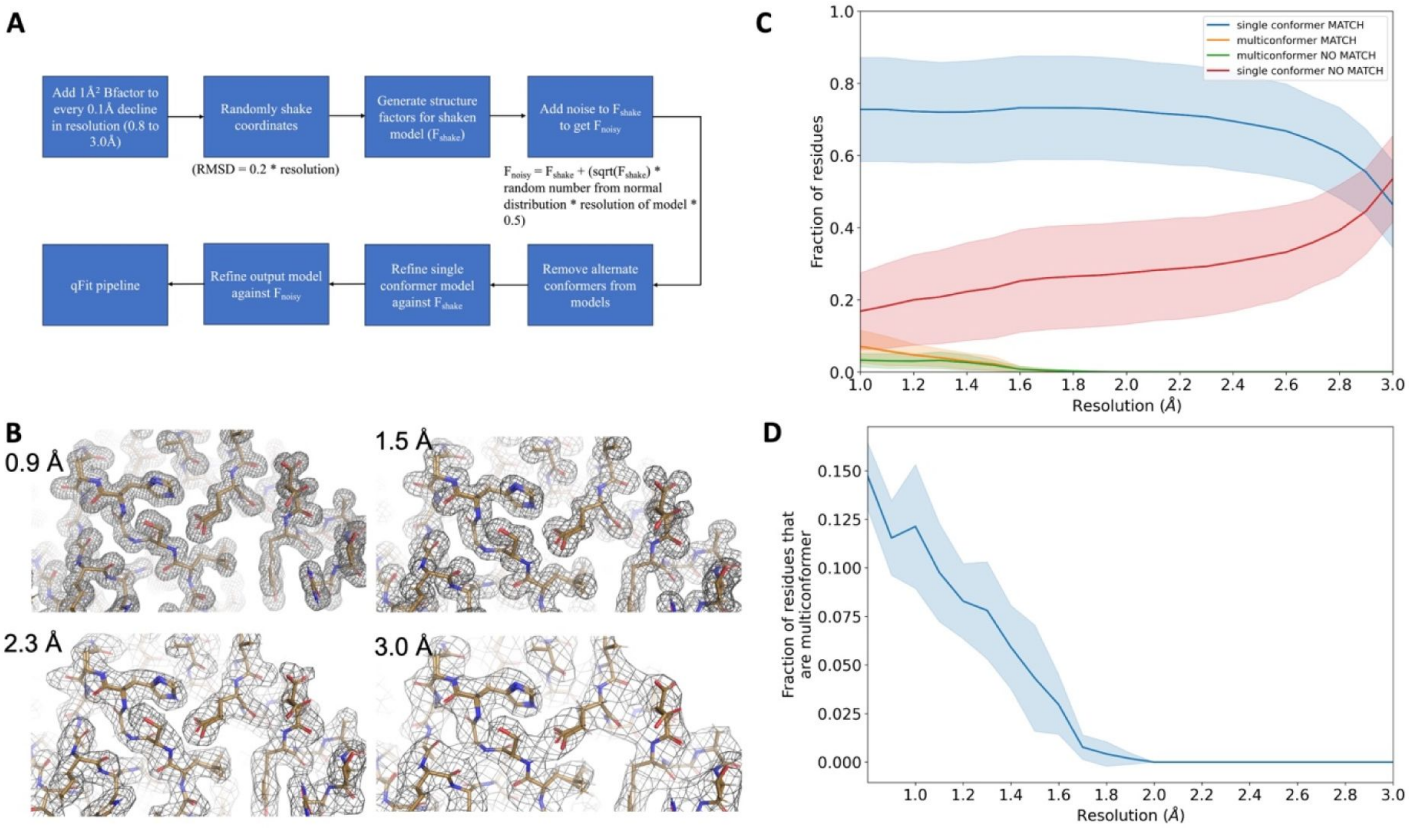
Supplementary Figure 3.

Examples of rotamer state categories
Meshes represent 2Fo-Fc density at 1 σ. Green and yellow sticks represent deposited conformer(s). Blue and magenta sticks represent qFit conformer(s).
A. Same: The entire set of rotamers identified in the deposited and qFit models are the same (PDB: 1BN6, His199).
B. Additional rotamer(s) in the qFit model: Deposited and qFit models share at least one rotamer, and at least one additional rotamer was identified in the qFit model (PDB: 3CX2, Glu165).
C. Additional rotamer(s) in the deposited model: Deposited and qFit models share at least one rotamer, and at least one additional rotamer was identified in the deposited model (PDB: 4P48, Ser6).
D. Consistent and different: Deposited and qFit models share at least one rotamer, and at least one unique additional rotamer was identified in both the deposited model and the qFit model (PDB: 3HP4, Arg81).
E. Different: The rotamers in the deposited and qFit models are all different (PDB: 1BN6, Glu110).



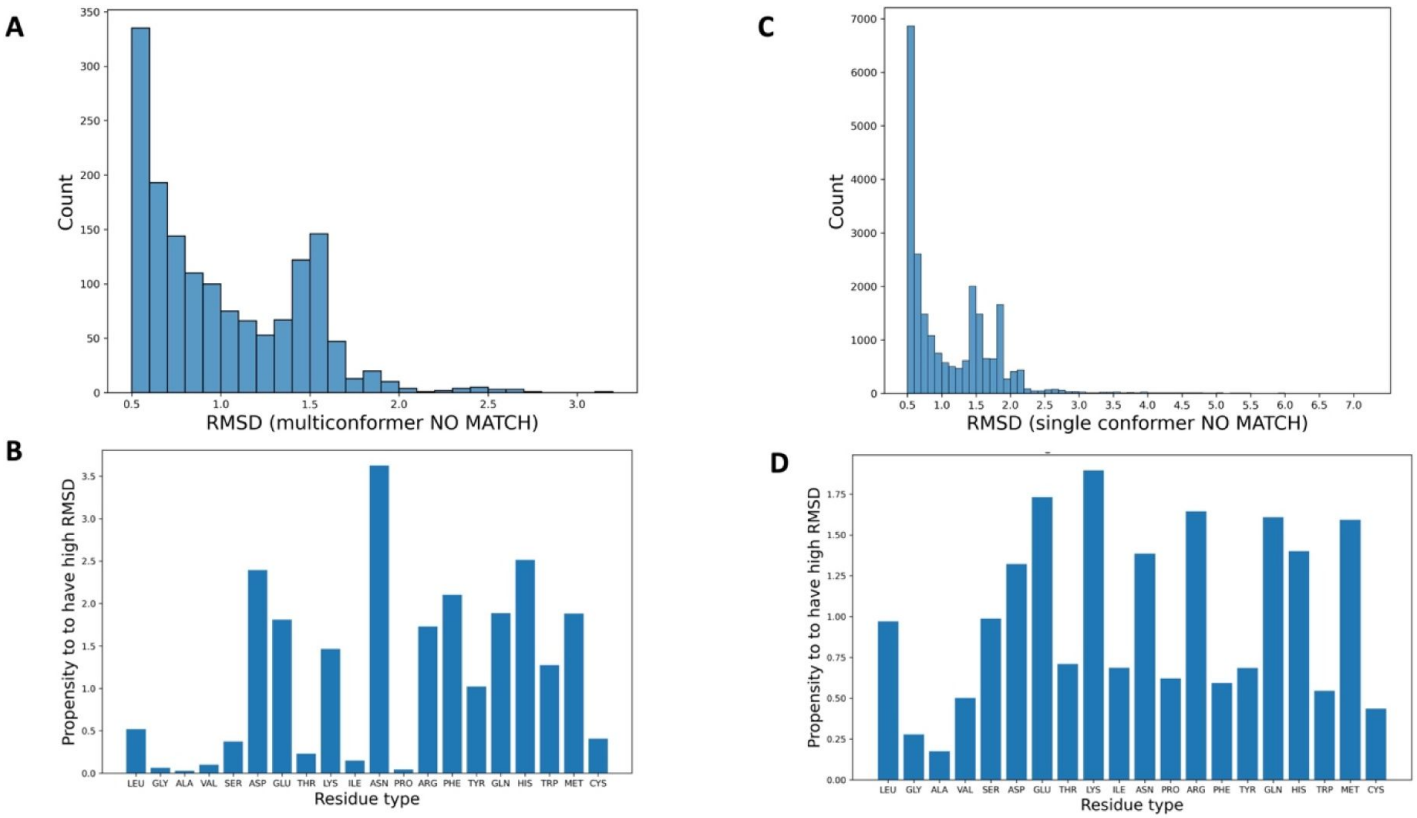
Supplementary Figure 4.

Deposited versus qFit model geometry.
A. Count of number of Cβ deviation (>0.25Å) per model (deposited model: 0.0 median [interquartile range: 0.0-0.0], qFit model: 0.0 median [interquartile range: 0.0-0.0]), p-value=0.37 from two-sided t-test.
B. Median count of number of rotamer outliers per model (deposited model: 0.94 [0.00-2.12], qFit model: 0.81 [0.35-1.60]), p-value=0.73 from two-sided t-test.
C. Percent of Ramachandran favored per model: deposited model (97.70 [96.90-98.93], qFit model: 98.0 [97.05-98.97]), p-value=0.77 from two-sided t-test.
D. Percent of Ramachandran outliers per model (deposited model 0.0 [0.0-0.0], qFit model: 0.0 [0.0- 0.0)]), p-value=0.57 from two-sided t-test.



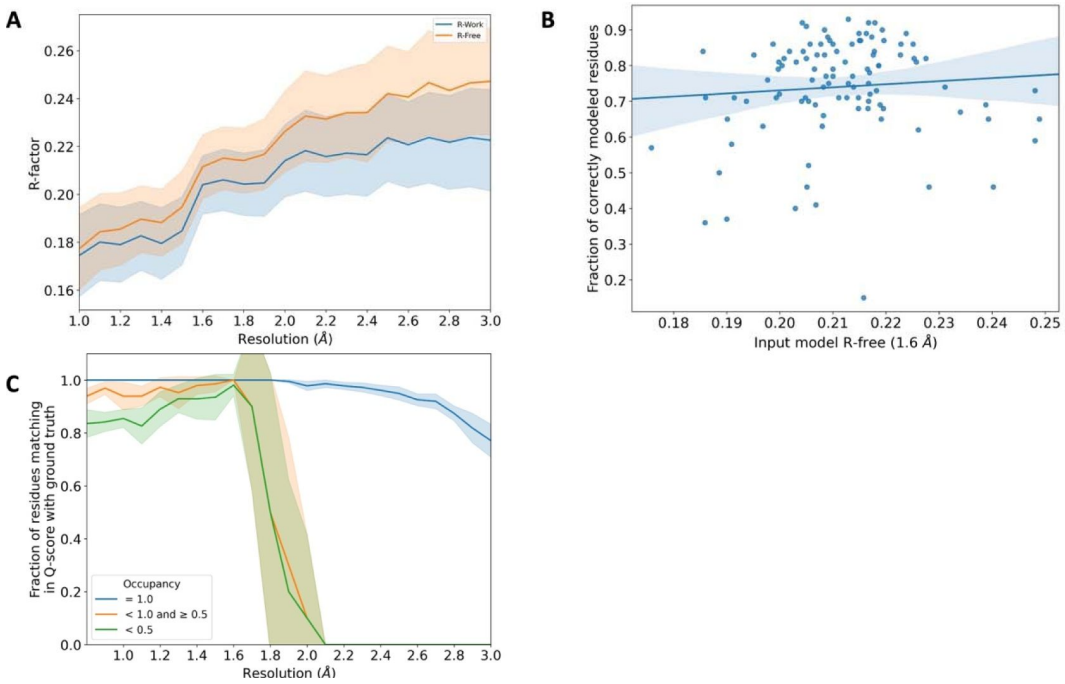
Supplementary Figure 5.

A. Protocol for generating synthetic structure factors at various resolutions starting from the ground truth model. For the 7KR0 dataset, all the steps starting from random shaking of coordinates were done 10 times for each resolution. For the larger test dataset, all steps were only done once.
B. A visualization of synthetic maps generated for the models at varying resolution. The loss in detail of density is clearly visible with worsening resolution.
C. Proportion of all residues in qFit models which have been modeled as multiconformers in the 7KR0 dataset, as a function of resolution. The shaded region around the line indicates the spread across 10 runs at every resolution step.
D. Proportion of all residues in the qFit models of qFit test dataset which are modeled as multiconformer match (orange), single conformer match (blue), multiconformer no match (green), and single conformer no match as a function of resolution of input data. The shaded region around the lines indicates the spread across the qFit test dataset which consists of 103 proteins.



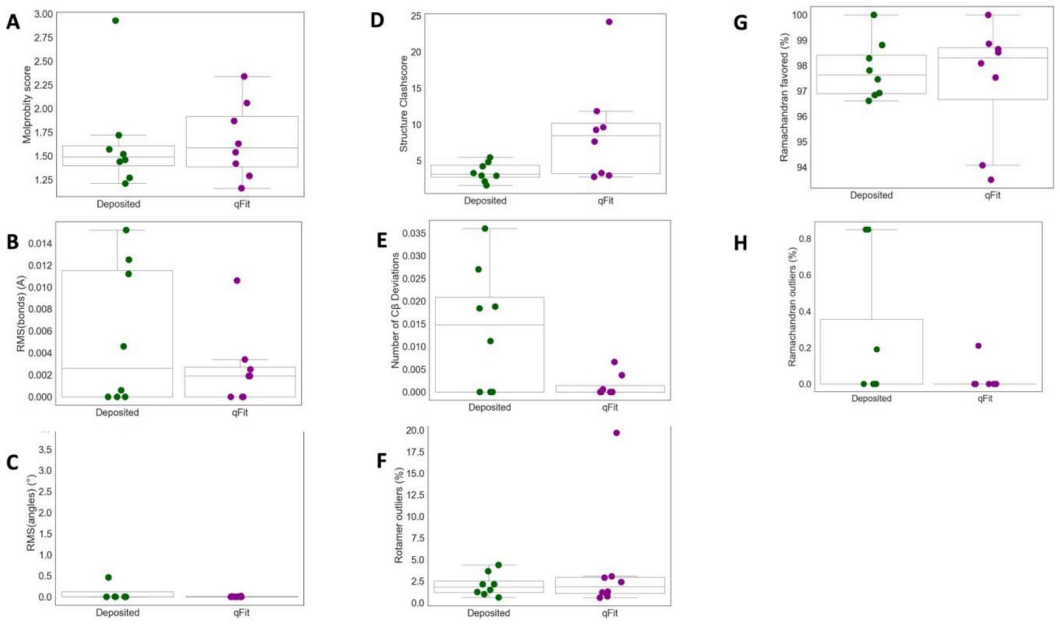
Supplementary Figure 6.

A. The distribution of RMSD between qFit residues and corresponding ground truth residues (qFit test set) whenever the RMSD is higher than the 0.5 Å cutoff, resulting in the qFit residues being classified as multiconformer no match.
B. The propensity of each residue type to be modeled with high RMSD from the ground truth (qFit test set), resulting in being classified as multiconformer no match. This propensity of a residue type x is calculated as the ratio between (i) proportion of residue type x among all the residues with a high RMSD and (ii) proportion of residue type x in the entire dataset.
C. The distribution of RMSD between qFit residues and corresponding ground truth residues (qFit test set) whenever the RMSD is higher than the 0.5 Å cutoff, resulting in the qFit residues being classified as single conformer no match.
D. The propensity of each residue type to be modeled with high RMSD from the ground truth (qFit test set), resulting in being classified as single conformer no match.



Supplementary Figure 7.

A. R_{work} (blue) and R_{free} (orange) distribution of the input model from the qFit test dataset. These correspond to the models obtained after refining against F_{noisy} structure factors (see **Supplementary Figure 5A**). The shaded region around the lines indicates the spread (standard deviation) across the qFit test dataset.
B. Fraction of correctly modeled qFit residues (match multiconformer + match single conformer) as a function of input model R_{free} for all structures in the qFit test dataset at 1.6 Å resolution (input R_{free} range: 0.17 to 0.25, n=103). The shaded region denotes the 95% confidence interval.
C. The fraction of residues in the qFit models of the 7KR0 dataset with a Q-score within 0.01 of that of the ground truth model as a function of resolution. In multiconformer residues, Q-score for every alternate conformer is calculated separately. Q-scores of residues (or) conformations which have matching occupancy (range) are compared. Occupancy of conformations were binned into three classes – occupancy equal to 1 (blue), 1 > occupancy >= 0.5 (orange) and occupancy < 0.5 (green).



Supplementary Figure 8.

A. MolProbity score(deposited model: 1.49 (median) [1.40-1.61] (interquartile range), qFit model: 1.59 (median) [1.39-1.92] (interquartile range))
B. Model average of RMSD of model bond length from idealized bond length(Å)(deposited model: 0.00 [0.00-0.01], qFit model: 0.00 [0.00-0.00])
C. Model average of RMSD of model bond angle from idealized bond angle(Å)(deposited model: 0.00 [0.00-0.11], qFit model: 0.00 [0.00-0.01])
D. Number of residues with clashscore(deposited model: 3.15 [2.74-4.39], qFit model: 8.45 [3.22-10.17])
E. Number of Cβ deviation (>0.25 Å) per model (deposited model: 0.02 [0.00-0.02], qFit model: 0.00 [0.00-0.00])
F. Number of rotamer outliers per model(deposited model: 2.0 [2.0-2.0], qFit model: 2.0 [1.0-3.0])
G. Percent of Ramachandran favored per model(deposited model: 97.6 [96.9-98.9], qFit model: 98.3 [96.7-98.7)])
H. Percent of Ramachandran outliers per model (deposited model: 0.0 [0.0-0.0], qFit model: 0.0 [0.0- 0.0])I.

7. Furnham N., Blundell T. L., DePristo M. A., Terwilliger T. C (2006) **Is one solution good enough?** *Nat. Struct. Mol. Biol* **13**:184–5

8. Fraser J. S., Lindorff-Larsen K., Bonomi M (2020) **What Will Computational Modeling Approaches Have to Say in the Era of Atomistic Cryo-EM Data?** *J. Chem. Inf. Model* **60**:2410–2412

9. Woldeyes R. A., Sivak D. A., Fraser J. S, pluribus unum E (2014) **no more: from one crystal, many conformations** *Curr. Opin. Struct. Biol* **28**:56–62

10. Burnley B. T., Afonine P. V., Adams P. D., Gros P (2012) **Modelling dynamics in protein crystal structures by ensemble refinement** *Elife* **1**

11. Ginn H. M (2021) **Vagabond: bond-based parametrization reduces overfitting for refinement of proteins** *Acta Crystallogr D Struct Biol* **77**:424–437

12. Emsley P., Lohkamp B., Scott W. G., Cowtan K (2010) **Features and development of Coot. Acta Crystallogr** *D Biol. Crystallogr* **66**:486–501

13. Ploscariu N., Burnley T., Gros P., Pearce N. M (2021) **Improving sampling of crystallographic disorder in ensemble refinement** *Acta Crystallogr D Struct Biol* **77**:1357–1364

14. Burling Temple, Brünger F. (1994) **Thermal Motion and Conformational Disorder in Protein Crystal Structures: Comparison of Multi-Conformer and Time-Averaging Models** *Isr. J. Chem* **34**:165–175

15. Weichenberger C. X., Afonine P. V., Kantardjieff K., Rupp B (2015) **The solvent component of macromolecular crystals** *Acta Crystallogr. D Biol. Crystallogr* **71**:1023–1038

16. Kabsch W. XDS. (2010) **Acta Crystallogr** *D Biol. Crystallogr* **66**:125–132

17. Karplus P. A., Diederichs K (2012) **Linking crystallographic model and data quality** *Science* **336**:1030–1033

18. Glaeser R. M (2019) **How Good Can Single-Particle Cryo-EM Become? What Remains Before It Approaches Its Physical Limits?** *Annu. Rev. Biophys* **48**:45–61

19. Stachowski T. R., Fischer M (2023) **FLEXR: automated multi-conformer model building using electron-density map sampling** *Acta Crystallogr D Struct Biol* **79**:354–367

20. Keedy D. A., Fraser J. S., van den Bedem H (2015) **Exposing Hidden Alternative Backbone Conformations in X-ray Crystallography Using qFit** *PLoS Comput. Biol* **11**

21. Riley B. T., et al. (2021) **qFit 3: Protein and ligand multiconformer modeling for X-ray crystallographic and single-particle cryo-EM density maps** *Protein Sci* **30**:270–285

22. van den Bedem, H., Dhanik, A., Latombe, J. C. & Deacon, A. M (2009) **Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers** *Acta Crystallogr. D Biol. Crystallogr* **65**:1107–1117

23. van Zundert G. C. P., et al. (2018) **qFit-ligand Reveals Widespread Conformational Heterogeneity of Drug-Like Molecules in X-Ray Electron Density Maps** *J. Med. Chem* **61**:11183–11198

24. Terwilliger T. C., et al. (2008) **Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias** *Acta Crystallogr. D Biol. Crystallogr* **64**:515–524

25. Morin A., et al. (2013) **Collaboration gets the most out of software** *Elife* **2**

26. Agrawal A., Verschueren R., Diamond S., Boyd S (2018) **A rewriting system for convex optimization problems** *Journal of Control and Decision* https://doi.org/10.1080/23307706.2017.1397554

27. Afonine P. V., et al. (2012) **Towards automated crystallographic structure refinement with phenix.refine** *Acta Crystallogr. D Biol. Crystallogr* **68**:352–367

28. Afonine P. V., Adams P. D., Sobolev O. V., Urzhumtsev A. G (2024) **Accounting for nonuniformity of bulk-solvent: A mosaic model** *Protein Sci* **33**

29. Anderson S., Crosson S., Moffat K (2004) **Short hydrogen bonds in photoactive yellow protein** *Acta Crystallogr. D Biol. Crystallogr* **60**:1008–1016

30. Berman H. M., et al. (2000) **The Protein Data Bank** *Nucleic Acids Res* **28**:235–242

31. Holton J. M., Classen S., Frankel K. A., Tainer J. A (2014) **The R-factor gap in macromolecular crystallography: an untapped potential for insights on accurate structures** *FEBS J* **281**:4046–4060

32. Vitkup D., Ringe D., Karplus M., Petsko G. A (2002) **Why protein R-factors are so large: a self-consistent analysis** *Proteins* **46**:345–354

33. Rodriguez-Corona U., Sobol M., Rodriguez-Zapata L. C., Hozak P., Castano E (2015) **Fibrillarin from Archaea to human** *Biol. Cell* **107**:159–174

34. **[No title]. https://phenix-online.org/phenixwebsite_static/mainsite/files/newsletter/CCN_2023_01.pdf#page=2.**

35. Orengo C. A., et al. (1997) **CATH--a hierarchic classification of protein domain structures** *Structure* **5**:1093–1108

36. Lovell S. C., Word J. M., Richardson J. S., Richardson D. C (2000) **The penultimate rotamer library** *Proteins* **40**:389–408

37. **https://phenix-online.org/phenixwebsite_static/mainsite/files/newsletter/CCN_2023_01.pdf#page=2.**

38. Williams C. J., et al. (2018) **MolProbity: More and better reference data for improved all-atom structure validation** *Protein Sci* **27**:293–315

39. Schuller M., et al. (2021) **Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking** *Sci Adv* **7**

40. Pintilie G., et al. (2020) **Measurement of atom resolvability in cryo-EM maps with Q-scores** *Nat. Methods* **17**:328–334

41. Nakane T., et al. (2020) **Single-particle cryo-EM at atomic resolution** *Nature* **587**:152–156

42. Xie Q., Yoshioka C. K., Chapman M. S (2020) **Adeno-Associated Virus (AAV-DJ)- Cryo-EM Structure at 1.56 Å Resolution** *Viruses* **12**

43. Chiu W., Schmid M. F., Pintilie G. D., Lawson C. L (2021) **Evolution of standardization and dissemination of cryo-EM structures and data jointly by the community, PDB, and EMDB** *J. Biol. Chem* **296**

44. Yip K. M., Fischer N., Paknia E., Chari A., Stark H (2020) **Atomic-resolution protein structure determination by cryo-EM** *Nature* **587**:157–161

45. Kleywegt G. J., et al. (2024) **Community recommendations on cryoEM data archiving and validation** *IUCrJ* **11**:140–151

46. Wolff A. M., et al. **Mapping Protein Dynamics at High-Resolution with Temperature- Jump X-ray Crystallography** *Preprint at* https://doi.org/10.1101/2022.06.10.495662

47. Dasgupta M., et al. (2019) **Dasgupta, M. et al. Mix-and-inject XFEL crystallography reveals gated conformational dynamics during enzyme catalysis. Proc. Natl. Acad. Sci. U. S. A. 116, 25634–25640 (2019).** *Proc. Natl. Acad. Sci. U. S. A* **116**:25634–25640

48. Correy G. J., et al. (2022) **The mechanisms of catalysis and ligand binding for the SARS- CoV-2 NSP3 macrodomain from neutron and x-ray diffraction at room temperature** *Sci Adv* **8**

49. Mehlman, T. (skaist), et al. **Room-temperature crystallography reveals altered binding of small-molecule fragments to PTP1B** *Preprint at* https://doi.org/10.1101/2022.11.02.514751

50. Ebrahim A., et al. **The temperature-dependent conformational ensemble of SARS- CoV-2 main protease (Mpro)** *Preprint at* https://doi.org/10.1101/2021.05.03.437411

51. Gahbauer S., et al. (2023) **Iterative computational design and crystallographic screening identifies potent inhibitors targeting the Nsp3 macrodomain of SARS-CoV-2** *Proc. Natl. Acad. Sci. U. S. A* **120**

52. Douangamath A., et al. (2020) **Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease** *Nat. Commun* **11**

53. Günther S., et al. (2021) **X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease** *Science* **372**:642–646

54. Zhong E. D., Bepler T., Berger B., Davis J. H (2021) **CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks** *Nat. Methods* **18**:176–185

55. Chen M., Ludtke S. J (2021) **Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM** *Nat. Methods* **18**:930–936

56. Kinman L. F., Powell B. M., Zhong E. D., Berger B., Davis J. H (2023) **Uncovering structural ensembles from single-particle cryo-EM data using cryoDRGN** *Nat. Protoc* **18**:319–339

57. Terashi G., Wang X., Subramaniya Maddhuri Venkata, Tesmer S. R., Kihara J. J. G. (2022) **Residue-wise local quality estimation for protein models from cryo-EM maps** *Nat. Methods* **19**:1116–1125

58. Li Q., et al. (2020) **Synthetic group A streptogramin antibiotics that overcome Vat resistance** *Nature* **586**:145–150

59. Fromm S. A., et al. (2023) **The translating bacterial ribosome at 1.55 Å resolution generated by cryo-EM imaging services** *Nat. Commun* **14**

60. Hintze B. J., Richardson J. S., Richardson D. C (2017) **Mismodeled purines: implicit alternates and hidden Hoogsteens** *Acta Crystallogr D Struct Biol* **73**:852–859

61. Wang Z., Patwardhan A., Kleywegt G. J (2022) **Validation analysis of EMDB entries** *Acta Crystallogr D Struct Biol* **78**:542–552

62. Lawson C. L., et al. (2021) **Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge** *Nat. Methods* **18**:156–164

63. Burley S. K., et al. (2022) **Electron microscopy holdings of the Protein Data Bank: the impact of the resolution revolution, new validation tools, and implications for the future** *Biophys. Rev* **14**:1281–1301

64. Davis I. W., Arendall W. B (2006) **3rd, Richardson, D. C. & Richardson** *J. S. The backrub motion: how protein backbone shrugs when a sidechain dances. Structure* **14**:265–274

65. Biel J. T., Thompson M. C., Cunningham C. N., Corn J. E., Fraser J. S (2017) **Flexibility and Design: Conformational Heterogeneity along the Evolutionary Trajectory of a Redesigned Ubiquitin** *Structure* **25**:739–749

66. Deis L. N., et al. (2014) **Multiscale conformational heterogeneity in staphylococcal protein a: possible determinant of functional plasticity** *Structure* **22**:1467–1477

67. Fraser J. S., et al. (2011) **Accessing protein conformational ensembles using room-temperature X-ray crystallography** *Proc. Natl. Acad. Sci. U. S. A* **108**:16247–16252

68. Pearce N. M., Gros P (2021) **A method for intuitively extracting macromolecular dynamics from structural disorder** *Nat. Commun* **12**

69. Wankowicz S., Fraser J (2024) **Comprehensive encoding of conformational and compositional protein structural ensembles through mmCIF data structure** *ChemRxiv* https://doi.org/10.26434/chemrxiv-2023-ggd1w-v3

70. Hancock M., et al. (2022) **Integration of software tools for integrative modeling of biomolecular systems** *J. Struct. Biol* **214**

71. Pearce N. M., Krojer T., von Delft F (2017) **Proper modelling of ligand binding requires an ensemble of bound and unbound states** *Acta Crystallogr D Struct Biol* **73**:256–266

72. Vallat B., et al. (2023) **ModelCIF: An Extension of PDBx/mmCIF Data Representation for Computed Structure Models** *J. Mol. Biol* **168021**

73. Otten R., et al. (2018) **Rescue of conformational dynamics in enzyme catalysis by directed evolution** *Nat. Commun* **9**

74. Zaragoza J. P. T., et al. (2023) **Temporal and spatial resolution of distal protein motions that activate hydrogen tunneling in soybean lipoxygenase** *Proc. Natl. Acad. Sci. U. S. A* **120**

75. Jumper J., et al. (2021) **Highly accurate protein structure prediction with AlphaFold** *Nature* **596**:583–589

76. Lane T. J (2023) **Protein structure prediction has reached the single-structure frontier** *Nat. Methods* **20**:170–173

77. Diamond S., Boyd S (2016) **CVXPY: A Python-Embedded Modeling Language for Convex Optimization** *J. Mach. Learn. Res* **17**

78. Winn M. D., et al. (2011) **Overview of the CCP4 suite and current developments** *Acta Crystallogr. D Biol. Crystallogr* **67**:235–242

Editors

Reviewer #1 (Public Review):

Summary:

Protein conformational changes are often critical to protein function, but obtaining structural information about conformational ensembles is a challenge. Over a number of years, the authors of the current manuscript have developed and improved an algorithm, qFit protein, that models multiple conformations into high resolution electron density maps in an automated way. The current manuscript describes the latest improvements to the program, and analyzes the performance of qFit protein in a number of test cases, including classical statistical metrics of data fit like Rfree and the gap between Rwork and Rfree, model geometry, and global and case-by-case assessment of qFit performance at different data resolution cutoffs. The authors have also updated qFit to handle cryo-EM datasets, although the analysis of its performance is more limited due to a limited number of high-resolution test cases and less standardization of deposited/processed data.

Strengths:

The strengths of the manuscript are the careful and extensive analysis of qFit's performance over a variety of metrics and a diversity of test cases, as well as careful discussion of the limitations of qFit. This manuscript also serves as a very useful guide for users in evaluating if and when qFit should be applied during structural refinement.

https://doi.org/10.7554/eLife.90606.2.sa2

Reviewer #2 (Public Review):

Summary

The manuscript "Uncovering Protein Ensembles: Automated Multiconformer Model building for X-ray Crystallography and Cryo-EM" by Wankowicz et al. describes updates to qFit, an algorithm for the characterization of conformational heterogeneity of protein molecules based on X-ray diffraction of Cryo-EM data. The work provides a clear description of the algorithm used by qFit. The authors then proceed to validate the performance of qFit by comparing to deposited X-ray entries in the PDB in the 1.2-1.5 Å resolution range as quantified by Rfree, Rwork-Rfree, detailed examination of the conformations introduced by

qFit, and performance on stereochemical measures (MolProbity scores). To examine the effect of experimental resolution of X-ray diffraction data, they start from an ultra high-resolution structure (SARS-CoV2 Nsp3 macrodomain) to determine how the loss of resolution (introduced artificially) degrades the ability of qFit to correctly infer the nature and presence of alternate conformations. The authors observe a gradual loss of ability to correctly infer alternate conformations as resolution degrades past 2 Å. The authors repeat this analysis for a larger set of entries in a more automated fashion and again observe that qFit works well for structures with resolutions better than 2 Å, with a rapid loss of accuracy at lower resolution. Finally, the authors examine the performance of qFit on cryo-EM data. Despite a few prominent examples, the authors find only a handful (8) of datasets for which they can confirm a resolution better than 2.0 Å. The performance of qFit on these maps is encouraging and will be of much interest because cryo-EM maps will, presumably, continue to improve and because of the rapid increase in the availability of such data for many supramolecular biological assemblies. As the authors note, practices in cryo-EM analysis are far from uniform, hampering the development and assessment of tools like qFit.

Strengths

qFit improves the quality of refined structures at resolutions better than 2.0 A, in terms of reflecting true conformational heterogeneity and geometry. The algorithm is well-designed and does not introduce spurious or unnecessary conformational heterogeneity. I was able to install and run the program without a problem within a computing cluster environment. The paper is well-written and the validation thorough.
I found the section on cryo-EM particularly enlightening, both because it demonstrates the potential for discovery of conformational heterogeneity from such data by qFit, and because it clearly explains the hurdles towards this becoming common practice, including lack of uniformity in reporting resolution, and differences in map and solvent treatment.

Weaknesses

Due to limitations of past software engineering, the paper lacks a careful comparison to past versions of qFit. In light of the extensive assessment of the current version of qFit, this is a minor concern.

Although qFit can handle supramolecular assemblies and bound organic molecules, analysis in the manuscript is limited to single-chain X-ray structures. I look forward to demonstration of its utility in such cases in future work.

Appraisal & Discussion

Overall, the authors convincingly demonstrate that qFit provides a reliable means to detect and model conformational heterogeneity within high-resolution X-ray diffraction datasets and (based on a smaller sample) in cryo-EM density maps. This represents the state of the art in the field and will be of interest to any structural biologist or biochemist seeking to attain an understanding of the structural basis of the function of their system of interest, including potential allosteric mechanisms-an area where there are still few good solutions. That is, I expect qFit to find widespread use.

https://doi.org/10.7554/eLife.90606.2.sa1

Reviewer #3 (Public Review):

Summary:

The authors address a very important issue of going beyond a single-copy model obtained by the two principal experimental methods of structural biology, macromolecular

crystallography and cryo electron microscopy (cryo-EM). Such multiconformer model is based on the fact that experimental data from both these methods represent a space- and time-average of a huge number of the molecules in a sample, or even in several samples, and that the respective distributions can be multimodal. Differently from structure prediction methods, this approach is strongly based on accurate high-resolution experimental information and requires validated single-copy high-quality models as input. In overall, the results support the authors' conclusions.

In fact, the method addresses two problems which could be considered separately:

- an automation of construction of multiple conformations when they can be identified visually;
- a determination of multiple conformations when their visual identification is difficult or impossible.

The former is a known problem, when missing alternative conformations may cost a few percent in R-factors. While these conformations are relatively easy to detect and build manually, the current procedure may save significant time being quite efficient, as the test results show. It is an indisputably useful tool for such a goal. The second problem is important from the physical point of view and has been considered first thirty years ago by Burling & Brünger. The manuscript does not specify clearly how much the current tool addresses the second case. To model such maps, the authors introduced errors in structure factors, however, being independent, as in this work, such errors, even quite high, may leave the maps reasonably well interpretable. Obviously, it is impossible to model all kinds of errors and this modeling of noise is appreciated but it would helpful for understanding if the manuscript shows, for example, the worst map when the procedure was successful.

The new procedure deals with a second-order variation in the R-factors, of about 1% or less, like placing riding hydrogen atoms, modeling density deformation or variation of the bulk solvent. In such situations, it is hard to justify model improvement. Keeping Rfree values or their marginal decreasing can be considered as a sign that the model does not overfit data but hardly as a strong argument in favor of the model.

In general, global targets are less appropriate for this kind of problems and local characteristics may be better indicators. Improvement of the model geometry is a good choice. Indeed, yet Cruickshank (1956) showed that averaged density images may lead to a shortening of covalent bonds when interpreting such maps by a single model. However, a total absence of geometric outliers is not necessarily required for the structures solved at a high resolution where diffraction data should have a more freedom to place the atoms where the experiments "see" them.

The key local characteristic for multicomformer models is a closeness of the model map to the experimental one. Actually, the procedure uses a kind of such measure, the Bayesian information criteria (BIC). Unfortunately, the manuscript does not describe how sharply it identifies the best model and how much it changes between the initial and final models; in general, there is no feeling about its values. The Q-score (page 17) can be an appropriate tool for the first problem where the multiple conformations and individual atomic images are clearly separated and not for the second problem where the contributions from neighboring conformations and atoms are merged. In addition to BIC or to even more conventional global target functions such as LS or map correlation, the extreme values of the local difference maps may help to validate, or not, the model.

This described method with the results presented is a strong argument for a need in experimental data and information they contain, differently from a pure structure prediction. This tool is important to produce user-unbiased multiconformer models rapidly

and automatically. At the same time, absence of strong density-based validation components may limit its impact.

Strengths:
Addressing an important problem and automatisation of model construction for alternative conformations using high-resolution experimental data.

Weaknesses:
An insufficient validation of the models when no discrete alternative conformations visible and insufficiency of local real-space validation indicators.

https://doi.org/10.7554/eLife.90606.2.sa0

Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *Summary:*
>
> *Protein conformational changes are often critical to protein function, but obtaining structural information about conformational ensembles is a challenge. Over a number of years, the authors of the current manuscript have developed and improved an algorithm, qFit protein, that models multiple conformations into high resolution electron density maps in an automated way. The current manuscript describes the latest improvements to the program, and analyzes the performance of qFit protein in a number of test cases, including classical statistical metrics of data fit like Rfree and the gap between Rwork and Rfree, model geometry, and global and case-by-case assessment of qFit performance at different data resolution cutoffs. The authors have also updated qFit to handle cryo-EM datasets, although the analysis of its performance is more limited due to a limited number of high-resolution test cases and less standardization of deposited/processed data.*
>
> *Strengths:*
>
> *The strengths of the manuscript are the careful and extensive analysis of qFit's performance over a variety of metrics and a diversity of test cases, as well as the careful discussion of the limitations of qFit. This manuscript also serves as a very useful guide for users in evaluating if and when qFit should be applied during structural refinement.*
>
> ***Reviewer #2 (Public Review):***
>
> *Summary*
>
> *The manuscript by Wankowicz et al. describes updates to qFit, an algorithm for the characterization of conformational heterogeneity of protein molecules based on X-ray diffraction of Cryo-EM data. The work provides a clear description of the algorithm used by qFit. The authors then proceed to validate the performance of qFit by comparing it to deposited X-ray entries in the PDB in the 1.2-1.5 Å resolution range as quantified by Rfree, Rwork-Rfree, detailed examination of the conformations introduced by qFit, and performance on stereochemical measures (MolProbity scores). To examine the effect of experimental resolution of X-ray diffraction data, they start from an ultra high-resolution structure (SARS-CoV2 Nsp3 macrodomain) to determine how the loss of resolution (introduced artificially) degrades the ability of qFit to correctly infer the nature and*

> *presence of alternate conformations. The authors observe a gradual loss of ability to correctly infer alternate conformations as resolution degrades past 2 Å. The authors repeat this analysis for a larger set of entries in a more automated fashion and again observe that qFit works well for structures with resolutions better than 2 Å, with a rapid loss of accuracy at lower resolution. Finally, the authors examine the performance of qFit on cryo-EM data. Despite a few prominent examples, the authors find only a handful (8) of datasets for which they can confirm a resolution better than 2.0 Å. The performance of qFit on these maps is encouraging and will be of much interest because cryo-EM maps will, presumably, continue to improve and because of the rapid increase in the availability of such data for many supramolecular biological assemblies. As the authors note, practices in cryo-EM analysis are far from uniform, hampering the development and assessment of tools like qFit.*
>
> *Strengths*
>
> *qFit improves the quality of refined structures at resolutions better than 2.0 A, in terms of reflecting true conformational heterogeneity and geometry. The algorithm is well designed and does not introduce spurious or unnecessary conformational heterogeneity. I was able to install and run the program without a problem within a computing cluster environment. The paper is well written and the validation thorough.*
>
> *I found the section on cryo-EM particularly enlightening, both because it demonstrates the potential for discovery of conformational heterogeneity from such data by qFit, and because it clearly explains the hurdles towards this becoming common practice, including lack of uniformity in reporting resolution, and differences in map and solvent treatment.*
>
> *Weaknesses*
>
> *The authors begin the results section by claiming that they made "substantial improvement" relative to the previous iteration of qFit, "both algorithmically (e.g., scoring is improved by BIC, sampling of B factors is now included) and computationally (improving the efficiency and reliability of the code)" (bottom of page 3). However, the paper does not provide a comparison to previous iterations of the software or quantitation of the effects of these specific improvements, such as whether scoring is improved by the BIC, how the application of BIC has changed since the previous paper, whether sampling of B factors helps, and whether the code faster. It would help the reader to understand what, if any, the significance of each of these improvements was.*

Indeed, it is difficult (embarrassingly) to benchmark against our past work due to the dependencies on different python packages and the lack of software engineering. With the infrastructure we've laid down with this paper, made possible by an EOSS grant from CZI, that will not be a problem going forward. Not only is the code more reliable and standardized, but we have developed several scientific test sets that can be used as a basis for broad comparisons to judge whether improvements are substantial. We've also changed with "substantial improvement" to "several modifications" to indicate the lack of comparison to past versions.

> *The exclusion of structures containing ligands and multichain protein models in the validation of qFit was puzzling since both are very common in the PDB. This may convey the impression that qFit cannot handle such use cases. (Although it seems that qFit has an algorithm dedicated to modeling ligand heterogeneity and seems to be able to handle multiple chains). The paper would be more effective if it explained how a user of the software would handle scenarios with ligands and multiple chains, and why these would be excluded from analysis here.*

qFit can indeed handle both. We left out multiple chains for simplicity in constructing a dataset enriched for small proteins while still covering diversity to speed the ability to rapidly iterate and test our approaches. Improvements to qFit ligand handling will be discussed in a forthcoming work as we face similar technical debt to what we saw in proteins and are undergoing a process of introducing “several modifications” that we hope will lead to “substantial improvement” - but at the very least will accelerate further development.

> *It would be helpful to add some guidance on how/whether qFit models can be further refined afterwards in Coot, Phenix, ..., or whether these models are strictly intended as the terminal step in refinement.*

We added to the abstract:

“Importantly, unlike ensemble models, the multiconformer models produced by qFit can be manually modified in most major model building software (e.g. Coot) and fit can be further improved by refinement using standard pipelines (e.g. Phenix, Refmac, Buster).”

and introduction:

“Multiconformer models are notably easier to modify and more interpretable in software like Coot12 unlike ensemble methods that generate multiple complete protein copies(Burnley et al. 2012; Ploscariu et al. 2021; Temple Burling and Brünger 1994).”

and results:

“This model can then be examined and edited in Coot12 or other visualization software, and further refined using software such as phenix.refine, refmac, or buster as the modeler sees fit.”

and discussion

“qFit is compatible with manual modification and further refinement as long as the subsequent software uses the PDB standard altloc column, as is common in most popular modeling and refinement programs. The models can therefore generally also be deposited in the PDB using the standard deposition and validation process.”

> *Appraisal & Discussion*
>
> *Overall, the authors convincingly demonstrate that qFit provides a reliable means to detect and model conformational heterogeneity within high-resolution X-ray diffraction datasets and (based on a smaller sample) in cryo-EM density maps. This represents the state of the art in the field and will be of interest to any structural biologist or biochemist seeking to attain an understanding of the structural basis of the function of their system of interest, including potential allosteric mechanisms-an area where there are still few good solutions. That is, I expect qFit to find widespread use.*
>
> ***Reviewer #3 (Public Review):***
>
> *Summary:*
>
> *The authors address a very important issue of going beyond a single-copy model obtained by the two principal experimental methods of structural biology, macromolecular crystallography and cryo electron microscopy (cryo-EM). Such multiconformer model is based on the fact that experimental data from both these methods represent a space- and time-average of a huge number of the molecules in a sample, or even in several samples, and that the respective distributions can be multimodal. Different from structure prediction methods, this approach is strongly*

> *based on high-resolution experimental information and requires validated single-copy high-quality models as input. Overall, the results support the authors' conclusions.*
>
> *In fact, the method addresses two problems which could be considered separately:*
>
> *- An automation of construction of multiple conformations when they can be identified visually;*

- A determination of multiple conformations when their visual identification is difficult or impossible.

We often think about this problem similarly to the reviewer. However, in building qFit, we do not want to separate these problems - but rather use the first category (obvious visual identification) to build an approach that can accomplish part of the second category (difficult to visualize) without building "impossible"/nonexistent conformations - *with a consistent approach/bias.*

> *The first one is a known problem, when missing alternative conformations may cost a few percent in R-factors. While these conformations are relatively easy to detect and build manually, the current procedure may save significant time being quite efficient, as the test results show.*

We agree with the reviewers' assessment here. The "floor" in terms of impact is automating a tedious part of high resolution model building and improving model quality.

> *The second problem is important from the physical point of view and has been addressed first by Burling & Brunger (1994; https://doi.org/10.1002/ijch.199400022). The new procedure deals with a second-order variation in the R-factors, of about 1% or less, like placing riding hydrogen atoms, modeling density deformation or variation of the bulk solvent. In such situations, it is hard to justify model improvement. Keeping Rfree values or their marginal decreasing can be considered as a sign that the model is not overfitted data but hardly as a strong argument in favor of the model.*

We agree with the overall sentiment of this comment. What is a significant variation in R-free is an important question that we have looked at previously (http://dx.doi.org/10.1101/448795) and others have suggested an R-sleep for further cross validation (https://pubmed.ncbi.nlm.nih.gov/17704561/). For these reasons it is important to get at the significance of the changes to model types from large and diverse test sets, as we have here and in other works, and from careful examination of the *biological* significance of alternative conformations with experiments designed to test their importance in mechanism.

> *In general, overall targets are less appropriate for this kind of problem and local characteristics may be better indicators. Improvement of the model geometry is a good choice. Indeed, yet Cruickshank (1956; https://doi.org/10.1107/S0365110X56002059) showed that averaged density images may lead to a shortening of covalent bonds when interpreting such maps by a single model. However, a total absence of geometric outliers is not necessarily required for the structures solved at a high resolution where diffraction data should have more freedom to place the atoms where the experiments "see" them.*

Again, we agree—geometric outliers should not be completely absent, but it is comforting when they and model/experiment agreement both improve.

> *The key local characteristic for multi conformer models is a closeness of the model map to the experimental one. Actually, the procedure uses a kind of such measure, the Bayesian information criteria (BIC). Unfortunately, there is no information about how*

> *sharply it identifies the best model, how much it changes between the initial and final models; in overall there is not any feeling about its values. The Q-score (page 17) can be a tool for the first problem where the multiple conformations are clearly separated and not for the second problem where the contributions from neighboring conformations are merged. In addition to BIC or to even more conventional target functions such as LS or local map correlation, the extreme and mean values of the local difference maps may help to validate the models.*

We agree with the reviewer that the problem of "best" model determination is poorly posed here. We have been thinking a lot about htis in the context of Bayesian methods (see: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9278553/); however, a major stumbling block is in how variable representations of alternative conformations (and compositions) are handled. The answers are more (but by no means simply) straightforward for ensemble representations where the entire system is constantly represented but with multiple copies.

> *This method with its results is a strong argument for a need in experimental data and information they contain, differently from a pure structure prediction. At the same time, absence of strong density-based proofs may limit its impact.*

We agree - indeed we think it will be difficult to further improve structure prediction methods without much more interaction with the experimental data.

> *Strengths:*
>
> *Addressing an important problem and automatization of model construction for alternative conformations using high-resolution experimental data.*
>
> *Weaknesses:*
>
> *An insufficient validation of the models when no discrete alternative conformations are visible and essentially missing local real-space validation indicators.*

While not perfect real space indicators, local real-space validation is implicit in the MIQP selection step and explicit when we do employ Q-score metrics.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations For The Authors):***
>
> *A point of clarification: I don't understand why waters seem to be handled differently in for cryo-EM and crystallography datasets. I am interested about the statement on page 19 that the Molprobity Clashscore gets worse for cryo-EM datasets, primarily due to clashes with waters. But the qFit algorithm includes a round of refinement to optimize placement of ordered waters, and the clashscore improves for the qFit refinement in crystallography test cases. Why/how is this different for cryo-EM?*

We agree that this was not an appropriate point. We believe that the high clash score is coming from side chains being incorrectly modeled. We have updated this in the manuscript and it will be a focus of future improvements.

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *- It would be instructive to the reader to explain how qFit handles the chromophore in the PYP (1OTA) example. To this end, it would be helpful to include deposition of the multiconformer model of PYP. This might also be a suitable occasion for discussion of*

> *potential hurdles in the deposition of multiconformer models in the PDB (if any!). Such concerns may be real concerns causing hesitation among potential users.*

Thank you for this comment. qFit does not alter the position or connectivity of any HETATM records (like the chromophore in this structure). Handling covalent modifications like this is an area of future development.

Regarding deposition, we have noted above that the discussion now includes:

"qFit is compatible with manual modification and further refinement as long as the subsequent software uses the PDB standard altloc column, as is common in most popular modeling and refinement programs. The models can therefore, generally also be deposited in the PDB using the standard deposition and validation process."

Finally, we have placed all PDBs in a Zenodo deposition (XXX) and have included that language in the manuscript. It is currently under a separate data availability section (page XXX). We will defer to the editor as to the best header that should go under.

> *- It may be advisable to take the description of true/false pos/negatives out of the caption of Figure 4, and include it in a box or so, since these terms are important in the main text too, and the caption becomes very cluttered.*

We think adding the description of true/false pos/negatives to the Figure panel would make it very cluttered and wordy. We would like to retain this description within the caption. We have also briefly described each in the main text.

> *- page 21, line 4: some issue with citation formatting.*

We have updated these citations.

> *- page 25, second paragraph: cardinality is the number of members of a set. Perhaps "minimal occupancy" is more appropriate.*

Thank you for pointing this out. This was a mistake and should have been called the occupancy threshold.

> *- page 26: it's - its*

Thank you, we have made this change.

> *- Font sizes in Supplementary Figures 5-7 are too small to be readable.*

We agree and will make this change.

> ***Reviewer #3 (Recommendations For The Authors):***
>
> *General remarks*
>
> *(1) As I understand, the procedure starts from shifting residues one by one (page 4; A.1). Then, geometry reconstruction (e.g., B1) may be difficult in some cases joining back the shifted residues. It seems that such backbone perturbation can be done more efficiently by shifting groups of residues ("potential coupled motions") as mentioned at the bottom of page 9. Did I miss its description?*

We would describe the algorithm as sampling (which includes minimal shifts) in the backbone residues to ensure we can link neighboring residues. We agree that future

iterations of qFit should include more effective backbone sampling by exploring motion along the Cβ-Cα, C-N, and (Cβ-Cα × C-N) bonds and exploring correlated backbone movements.

> *(2) While the paper is well split in clear parts, some of them seem to be not at their right/optimal place and better can be moved to "Methods" (detailed "Overview of the qFit protein algorithm" as a whole) or to "Data" missed now (Two first paragraphs of "qFit improves overall fit...", page 8, and "Generating the qFit test set", page 22, and "Generating synthetic data ..." at page 26; description of the test data set), At my personal taste, description of tests with simulated data (page 15) would be better before that of tests with real data.*

Thank you for this comment, but we stand by our original decision to keep the general flow of the paper as it was submitted.

> *(3) I wonder if the term "quadratic programming" (e.g., A3, page 5) is appropriate. It supposes optimization of a quadratic function of the independent parameters and not of "some" parameters. This is like the crystallographic LS which is not a quadratic function of atomic coordinates, and I think this is a similar case here. Whatever the answer on this remark is, an example of the function and its parameters is certainly missed.*

We think that the term quadratic programming is appropriate. We fit a function with a loss function (observed density - calculated density), while satisfying the independent parameters. We fit the coefficients minimizing a quadratic loss. We agree that the quadratic function is missing from the paper, and we have now included it in the Methods section.

> *Technical remarks to be answered by the authors :*
>
> *(1) Page 1, Abstract, line 3. The ensemble modeling is not the only existing frontier, and saying "one of the frontiers" may be better. Also, this phrase gives a confusing impression that the authors aim to predict the ensemble models while they do it with experimental data.*

We agree with this statement and have re-worded the abstract to reflect this.

> *(2) Page 2. Burling & Brunger (1994) should be cited as predecessors. On the contrary, an excellent paper by Pearce & Gros (2021) is not relevant here.*

While we agree that we should mention the Burling & Brunger paper and the Pearce & Gros (2021) should not be removed as it is not discussing the method of ensemble refinement.

> *(3) Page 2, bottom. "Further, when compared to ..." The preference to such approach sounds too much affirmative.*

We have amended this sentence to state:

“Multiconformer models are notably easier to modify and more interpretable in software like Coot(Emsley et al. 2010) unlike ensemble methods that generate multiple complete protein copies(Burnley et al. 2012; Ploscariu et al. 2021; Temple Burling and Brünger 1994).”

“The point we were trying to make in this sentence was that ensemble-based models are much harder to manually manipulate in Coot or other similar software compared to multiconformer models. We think that the new version of this sentence states this point more clearly.”

> *(4) Page 2, last paragraph. I do not see an obvious relation of references 15-17 to the phrase they are associated with.*

We disagree with this statement, and think that these references are appropriate.

“Multiconformer models are notably easier to modify and more interpretable in software like Coot12 unlike ensemble methods that generate multiple complete protein copies(Burnley et al. 2012; Ploscariu et al. 2021; Temple Burling and Brünger 1994).”

> *(5) Page 3, paragraph 2. Cryo-EM maps should be also "high-resolution"; it does not read like this from the phrase.*

We agree that high-resolution should be added, and the sentence now states:

“However, many factors make manually creating multiconformer models difficult and time-consuming. Interpreting weak density is complicated by noise arising from many sources, including crystal imperfections, radiation damage, and poor modeling in X-ray crystallography, and errors in particle alignment and classification, poor modeling of beam induced motion, and imperfect detector Detector Quantum Efficiency (DQE) in high-resolution cryo-EM.”

> *(6) Page 3, last paragraph before "results". The words "... in both individual cases and large structural bioinformatic projects" do not have much meaning, except introducing a self-reference. Also, repeating "better than 2 A" looks not necessary.*

We agree that this was unnecessary and have simplified the last sentence to state:

“With the improvements in model quality outlined here, qFit can now be increasingly used for finalizing high-resolution models to derive ensemble-function insights.”

> *(7) Page 3. "Results". Could "experimental" be replaced by a synonym, like "trial", to avoid confusing with the meaning "using experimental data"?*

We have replaced experimental with exploratory to describe the use of qFit on CryoEM data. The statement now reads:

“For cryo-EM modeling applications, equivalent metrics of map and model quality are still developing, rendering the use of qFit for cryo-EM more exploratory.”

> *(8) Page 4, A.1. Should it be "steps +/- 0.1" and "coordinate" be "coordinate axis"? One can modify coordinates and not shift them. I do not understand how, with the given steps, the authors calculated the number of combinations ("from 9 to 81"). Could a long "Alternatively, ...absent" be reduced simply to "Otherwise"?*

We have simplified and clarified the sentence on the sampling of backbone coordinates to state:

“If anisotropic B-factors are absent, the translation of coordinates occurs in the X, Y, and Z directions. Each translation takes place in steps of 0.1 along each coordinate axis, extending to 0.3 Å, resulting in 9 (if isotropic) or to 81 (if anisotropic) distinct backbone conformations for further analysis.”

> *(9) Page 6, B.1, line 2. Word "linearly" is meaningless here.*

We have modified this to read:

“Moving from N- to C- terminus along the protein,”

> *(10) Page 9, line 2. It should be explained which data set is considered as the test set to calculate Rfree.*

We think this is clear and would be repetitive if we duplicated it.

> *(11) Page 9, line 7. It should be "a valuable metric" and not "an"*

We agree and have updated the sentence to read:

“Rfree is a valuable metric for monitoring overfitting, which is an important concern when increasing model parameters as is done in multiconformer modeling.”

> *(12) Page 10, paragraph 3. "... as a string (Methods)". I did not find any other mention of this term "string", including in "Methods" where it supposed to be explained. Either this should be explained (and an example is given?), or be avoided.*

We agree that string is not necessary (discussing the programmatic datatype). We have removed this from the sentence. It now reads:

“To quantify how often qFit models new rotameric states, we analyzed the qFit models with *phenix.rotalyze*, which outputs the rotamer state for each conformer (**Methods**).”

> *(13) Page10, lines 3-4 from bottom. Are these two alternative conformations justified?*

We are unsure what this is referring to.

> *(14) Page 12, Fig. 2A. In comparison with Supplement Fig 2C, the direction of axes is changed. Could they be similar in both Figures?*

We have updated Supplementary Figure 2C to have the same direction of axes as Figure 2A.

> *(15) Page 15, section's title. Choose a single verb in "demonstrate indicate".*

We have amended the title of this section to be:

“Simulated data demonstrate qFit is appropriate for high-resolution data.”

> *(16) Page 15, paragraph 2. "Structure factors from 0.8 to 3.0 A resolution" does not mean what the author wanted apparently to tell: "(complete?) data sets with the high-resolution limit which varied from 0.8 to 3.0 A ...". Also, a phrase of "random noise increasing" is not illustrated by Figs.5 as it is referred to.*

We have edited this sentence to now read:

“To create the dataset for resolution dependence, we used the ground truth 7KR0 model, including all alternative conformations, and generated artificial structure factors with a high resolution limit ranging from 0.8 to 3.0 Å resolution (in increments of 0.1 Å).”

> *(17) Page 15, last paragraph is written in a rather formal and confusing way while a clearer description is given in the figure legend and repeated once more in Methods. I would suggest to remove this paragraph.*

We agree that this is confusing. Instead of create a true positive/false positive/true negative/false negative matrix, we have just called things as they are, multiconformer or single conformer and match or no match. We have edited the language the in the manuscript and figure legends to reflect these changes.

> *(18) Page 16. Last two paragraphs start talking about a new story and it would help to separate them somehow from the previous ones (sub-title?).*

We agree that this could use a subtitle. We have included the following subtitle above this section:

“Simulated multiconformer data illustrate the convergence of qFit.”

> *(19) Page 20. "or static" and "we determined that" seem to be not necessary.*

We have removed static and only used single conformer models. However, as one of the main conclusions of this paper is determining that qFit can pick up on alternative conformers that were modeled manually, we have decided to the keep the “we determined that”.

> *(20) Page 21, first paragraph. "Data" are plural; it should be "show" and "require"*

We have made these edits. The sentence now reads:

“However, our data here shows that not only does qFit need a high-resolution map to be able to detect signal from noise, it also requires a very well-modeled structure as input.”

> *(21) Page 21, References should be indicated as [41-45], [35,46-48], [55-57]. A similar remark to [58-63] at page 22.*

We have fixed the reference layout to reflect this change.

> *(22) Page 21, last paragraph. "Further reduce R-factors" (moreover repeated twice) is not correct neither by "further", since here it is rather marginal, nor as a goal; the variations of R-factors are not much significant. A more general statement like "improving fit to experimental data" (keeping in mind density maps) may be safer.*

We agree with the duplicative nature of these statements. We have amended the sentence to now read:

“Automated detection and refinement of partial-occupancy waters should help improve fit to experimental data further reduce Rfree15 and provide additional insights into hydrogen-bond patterns and the influence of solvent on alternative conformations.”

> *(23) Page 22. Sub-sections of "Methods" are given in a little bit random order; "Parallelization of large maps" in the middle of the text is an example. Put them in a better order may help.*

We have moved some section of the Methods around and made better headings by using an underscore to highlight the subsections (*Generating and running the qFit test set, qFit improved features, Analysis metrics, Generating synthetic data for resolution dependence).*

> *(24) Page 24. Non-convex solution is a strange term. There exist non-convex problems and functions and not solutions.*

We agree and we have changed the language to reflect that we present the algorithm with non-convex problems which it cannot solve.

> *(25) Page 26, "Metrics". It is worthy to describe explicitly the metrics and not (only) the references to the scripts.*

For all metrics, we describe a sentence or two on what each metric describes. As these metrics are well known in the structural biology field, we do not feel that we need to elaborate on them more.

> *(26) Page 26. Multiplying B by occupancy does not have much sense. A better option would be to refer to the density value in the atomic center as occ*(4*pi/B)^1.5 which gives a relation between these two entities.*

We agree and have update the B-factor figures and metrics to reflect this.

> *(27) Page 40, suppl. Fig. 5. Due to the color choice, it is difficult to distinguish the green and blue curves in the diagram.*

We have amended this with the colors of the curves have been switched.

> *(28) Page 42, Suppl. Fig. 7. (A) How the width of shaded regions is defined? (B) What the blue regions stand for? Input Rfree range goes up to 0.26 and not to 0.25; there is a point at the right bound. (C) Bounds for the "orange" occupancy are inversed in the legend.*

(A) The width of the shaded region denotes the standard deviations among the values at every resolution. We have made this clearer in the caption

(B) The blue region denotes the confidence interval for the regression estimate. Size of the confidence interval was set to 95%. We have made this clearer in the caption

(C) This has been fixed now

The maximum R-free value is 0.2543, which we rounded down to 0.25.

> *(29) Page 43. Letters E-H in the legend are erroneously substituted by B-E.*

We apologize for this mistake. It is now corrected.

https://doi.org/10.7554/eLife.90606.2.sa4