KNexPHENIX: A PHENIX-Based Workflow for Improving Cryo-EM and Crystallographic Structural Models

Suparno Nandi* and Graeme L. Conn*

Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA

*Address correspondence to: nandisuparno@gmail.com or gconn@emory.edu

Key Words: macromolecular model refinement; structural biology; cryo-electron microscopy; X-ray crystallography; PHENIX

ABSTRACT

The number of structures deposited in the Protein Data Bank (PDB) has increased exponentially in recent years due to the availability of new methods to visualize increasingly complex macromolecules and their complexes in atomic detail. However, accurately refining atomic models from experimental maps remains a challenge due to limitations of current approaches for efficient model refinement. Standard PHENIX refinement can partially address these limitations with its speed and broad accessibility but often fails to yield the best model compared to those generated by more computationally intensive approaches. We therefore developed "KNexPHENIX", a customized semi-automated PHENIX-based workflow to support optimal macromolecular model building. KNexPHENIX can be used to refine macromolecular structures obtained from both cryo-electron microscopy (cryo-EM) and X-ray crystallography and is independent of the nature or size of the molecule. We evaluated KNexPHENIX on deposited structures as well as on *de novo* models chosen based on their visual fit to the deposited map or generated by molecular replacement. Compared to other approaches–PHENIX with default settings, REFMAC Servalcat, REFMAC, and CERES–KNexPHENIX consistently produced models with lower MolProbity scores, indicating improved model stereochemistry. Importantly, this was accomplished while maintaining model-to-map correlation for cryo-EM datasets and maintaining or reducing the R_{free}-R_{work} difference below accepted thresholds for X-ray crystallographic structures, limiting overfitting while preserving refinement accuracy. These results establish the KNexPHENIX workflow as a practical, accessible approach for refining both cryo-EM and crystallographic structures, enabling the generation of models with enhanced quality metrics suitable for deposition and to guide further experimental studies.

INTRODUCTION

Recent technological advancements in structural biology, particularly in single-particle cryo-electron microscopy (cryo-EM), have led to an exponential increase in the number of structures deposited in the Protein Data Bank (PDB)[1, 2]. Both X-ray crystallography and cryo-EM are now routinely employed to determine the structures of proteins, nucleic acids, and large macromolecular assemblies, such as nuclear pore complexes (up to 120 MDa), viral capsids (up to 4 MDa), ribosomes and ribosome subunits (up to 3 MDa), and large enzymes, e.g., RNA polymerase II (up to 900 kDa). While the process of obtaining three-dimensional maps for these complexes using cryo-EM or X-ray crystallography is in itself technically demanding and time-consuming, an equally significant challenge lies in the need to accurately build and refine atomic models guided by these maps. These structural models require adherence to widely accepted validation parameter values, including clashscore and Ramachandran outliers, as well as R_{work}/ R_{free} and CC_{mask} specifically for X-ray crystallographic and cryo-EM structures, respectively, to be considered appropriate by the scientific community[3-5]. Additionally, the task becomes even more challenging at resolutions below ~3 Å, where manual model building is both error-prone and labor-intensive[6]. Notably, more

than 70% of all the cryo-EM structures deposited to date in the PDB were determined with maps < 3 Å resolution.

To address these issues, several computational methods have been developed to refine models quickly, correctly, and with minimum manual intervention, such as EM-Refiner[7], Rosetta-Phenix[8], High Ambiguity Driven protein−protein Docking (HADDOCK)[9], and Correlation-Driven Molecular Dynamics (CDMD)[10]. While these approaches are effective, some rely on computationally demanding techniques such as Monte Carlo simulation in EM-Refiner[11] or molecular dynamics (MD) simulations in CDMD[12]. Other approaches also have some limitations. For example, the Rosetta-Phenix pipeline is noted in the PHENIX website documentation to perform poorly on proteins containing more than 1000 residues[13], does not refine RNA, and is computationally intensive. HADDOCK, meanwhile, focuses primarily on improvements in clashscore[9]. These issues pose significant barriers to widespread adoption of these methods, especially where high-performance computing resources are not readily available[14]. In contrast, PHENIX is one of the most widely used software suites to refine X-ray crystallographic structures due to its broad accessibility, as well as its speed over other methods due to computational efficiency[15-17]. Nevertheless, prior studies have noted that PHENIX does not always generate final models with optimal quality[8, 10, 17].

To enhance the accuracy of PHENIX-based refinement, we developed a new approach, which we term KNexPHENIX, by combining different programs available on the PHENIX suite, including ReadySet, PHENIX refinement, and geometry minimization with customized parameters. KNexPHENIX is designed to enhance both X-ray crystallographic and cryo-EM structures deposited in the PDB, as well as to refine models during the process of building *de novo* into maps where no prior structure exists. Benchmarking shows KNexPHENIX consistently outperforms default refinement available in PHENIX and REFMAC5 for crystal structures[18], REFMAC Servalcat for cryo-EM structures[19], and Cryo-EM Re-refinement System (CERES), an online tool for improving deposited cryo-EM structures by PHENIX refinement with customized parameters[20]. Therefore, KNexPHENIX represents a computationally efficient and accessible alternative for researchers to enhance the quality of their macromolecule structures.

METHODS

Overview of KNexPHENIX workflow

The general framework of KNexPHENIX refinement involves five major steps:

1) Addition of hydrogen atoms to the initial model.
2) Refinement using PHENIX, with specific parameters settings.
3) Geometry minimization, with customized parameters.
4) Removal of hydrogen atoms
5) Final refinement in PHENIX, with modified parameters.

For the studies described in this work, PHENIX version 1.21.1-5286 was used throughout, and all calculations were performed using a single processor. Workflows for refinement of existing or de novo cryoEM and X-ray crystal structures are described in the following sections, and a more detailed "how-to" guide with screenshots of the PHENIX interface is provided in the Supplementary Information.

KNexPHENIX Workflow 1: Refinement strategy for deposited cryo-EM models

After the addition of hydrogen atoms to the model using *phenix.ready_set*, we employed *phenix.real_space_refine* for five cycles using local grid search, global minimization, and B-factor refinement, with target root mean square deviation (RMSD) of 0.01Å and 1.0˚ for bonds and angles, respectively. Restraints such as secondary structure and Ramachandran were applied, but not reference model restraints. Additionally, rotamer restraints were added with the sigma value of 0.35, specifically targeting outliers (*rotamers.restraints.target=outliers*). Both *rotamers.tuneup* and *rotamers.fit* were assigned to outliers and poormap. For *pdb_interpretation*, Ramachandran restraints for only peptide bonds were added, excluding the peptide plane. The dihedral function type was set to be determined by the sign of periodicity. Next, one cycle of *phenix.geometry_minimization* was used with corrections turned on for bond length, bond angle, dihedral angle, chirality, planarity, parallelity, nonbonded distances, rotamer outliers, and secondary structure restraints. Additionally, Ramachandran restraints were applied to the model in *pdb_interpretation.* Hydrogen atoms were then removed from the minimized model using *phenix.pdbtools*. The resulting model was subject to a final round of B-factor refinement with the dihedral function type for the dihedral angles set to all harmonic in *pdb_interpretation*. For all the above steps, parameters not specified were set to their default values.

KNexPHENIX Workflow 2: Refinement strategy to improve *de novo* modeling into cryo-EM maps

To reflect a potential strategy in the determination of new structures by cryoEM, initial models were selected based on structural similarity to the existing PDB model for the cryo-EM map. The starting models were stripped of ligands and water molecules and then fit to the cryo-EM map using *fitmap* in Chimera[21]. Next, hydrogen atoms were added, and the modified model was PHENIX refined as described in the previous section, except that a single round of simulated annealing was included within the five cycles of refinement, and the dihedral function type was set as all harmonic in *pdb_interpretation*. It should be noted that if CC_{mask} values are suboptimal post-refinement, the user might consider reverting the rotamer restraints parameters to default values (rotamer restraints enabled, *rotamers.fit* set to outliers and poormap, and no specific setting for *rotamers.restraints.target*, *rotamers.tuneup*, or sigma). Parameters for geometry minimization were also kept unaltered from those used above, except that the dihedral function type was selected as all harmonic, and five cycles of minimization were used. Following the removal of hydrogen atoms from the minimized

model, five cycles of PHENIX refinement were conducted using global minimization and B-factor refinement. Reference model restraints were also implemented, with the starting model used as a reference. In *pdb_interpretation*, harmonic restraints on starting coordinates were enabled, the dihedral function type was set to all harmonic, and Ramachandran restraints and peptide plane were applied. All unspecified parameters in the procedures above were left at their default settings.

KNexPHENIX Workflow 3: Refinement strategy for previously refined X-ray crystal structures

Hydrogen atoms were added to the model using *phenix.ready_set*, and the modified structure was geometry minimized through two cycles, enforcing corrections for bond length, bond angle, dihedral angle, chirality, planarity, parallelity, nonbonded distances, rotamer outliers, and secondary structure restraints. For minimization, harmonic restraints on starting coordinates were enabled in *pdb_interpretation*, and Ramachandran restraints and peptide plane on peptide bonds were applied to the model. Post-minimization, hydrogen atoms were removed as described previously, followed by five cycles of PHENIX refinement (*phenix.refine*) using real-space, reciprocal-space, occupancy, and B-factor refinement, and included simulated annealing at the second and fourth (penultimate) rounds of refinement. Additionally, refinement cycles were stereochemistry weighted, with secondary structure restraints turned on. In *pdb_interpretation* for PHENIX refinement, harmonic restraints on starting coordinates were enabled, and Ramachandran restraints and peptide plane on peptide bonds were applied. The *dihedral_function_type* was set to all harmonic for both minimization and refinement. Parameters other than those detailed were used as provided by the default configuration.

KNexPHENIX Workflow 4: Refinement strategy to improve *de novo* modeling into X-ray crystallographic maps

For model fitting into maps derived from molecular replacement (MR)[22], an existing structure was chosen from the PDB database based on similarity (RMSD <1.0) to the structure deposited corresponding to that map, followed by removal of water molecules or ligands from the model. PHASER was used to perform MR using the entire model with sequence identity set to 100%. As before, hydrogen atoms were added to the MR model using *phenix.ready_set*, followed by *phenix.refine* with strategies as described above for refinement of deposited X-ray crystal structures. However, *reference_model.enabled* was set to “true” with the starting model used as reference. Parameters in *pdb_interpretation* were the same as above, except that *ramachandran_plot_restraints.enabled* was set to “false”. Geometry minimization parameters were also the same, except that harmonic restraints on starting coordinates were not applied. Post minimization, hydrogen atoms were removed from the model, followed by PHENIX refinement as described above. The

dihedral_function_type was set to all harmonic for both PHENIX refinement and geometry minimization. Settings not otherwise noted remained unchanged from their defaults for each step.

Default PHENIX refinement for cryo-EM and X-ray crystal structures

For performance comparisons in refinement of cryo-EM structures with KNexPHENIX, five cycles of "default" PHENIX refinement (*phenix.real_space_refine*) were performed using local grid search, global minimization, occupancy, N/Q/H flips, and B-factor refinement, with target root mean square deviation (RMSD) of 0.01Å and 1.0˚ for bonds and angles, respectively. Restraints such as secondary structure and Ramachandran were applied, but not reference model restraints. For *pdb_interpretation*, Ramachandran restraints for only peptide bonds were added, excluding the peptide plane. The dihedral function type was set to be determined by the sign of periodicity. Similarly, unless stated otherwise, for X-ray crystallographic structure refinement comparisons, five cycles of PHENIX refinement (*phenix.refine*) were performed using real-space, reciprocal-space, occupancy, and B-factor refinement. For *pdb_interpretation*, the dihedral function type was set to be determined by the sign of periodicity. For all the steps, parameters not described above were set to their default values. All the parameters in the two refinement procedures described here employed default values for parameters not described here.

Model refinement in REFMAC

Cryo-EM structures were refined using twenty cycles of REFMAC Servalcat (version 1.6.0) masked refinement with the weight and symmetry chosen automatically, and addition of hydrogen atoms. No RNA/DNA restraints were added, sharpening was not performed, and jellybody refinement was turned off. Crystal structures were refined in REFMAC5 (version 5.8.0419) through ten cycles of maximum likelihood restrained refinement using default parameters, e.g., isotropic B-factors, automatically optimized and experimental sigma stereochemistry/X-ray weights, and hydrogen atom addition. MolProbity scores[23, 24], CC_{mask}, R_{work}, and R_{free} values for the REFMAC-refined cryo-EM and crystal structures were calculated using the PHENIX validation tool. The MolProbity scores, R_{work}, and R_{free} values for the crystal structures deposited in PDB were also calculated similarly. Any options not explicitly mentioned were kept at their default values throughout.

RESULTS

KNexPHENIX improves the model quality of existing cryo-EM structures

To test the utility of KNexPHENIX for enhancing the model quality of structures previously deposited in the PDB, we chose thirteen models from the database reflecting diversity in the type of macromolecule (proteins and nucleoprotein complexes) and size (kDa to MDa range), but all exhibiting comparatively poorer model quality indicators (e.g., higher Clashscore and percentage of Ramachandran outliers; **Supplemental Table S1**). To determine improvement, we compared the MolProbity scores and CC_{mask} values between the deposited model and the new model resulting from refinement in PHENIX, Servalcat, or KNexPHENIX using Workflow 1 (**Fig.1A**; also see Methods for details). PHENIX refined structures collectively show slightly higher MolProbity scores and increased CC_{mask} values, indicating poorer model quality, but with slight improvement in model-to-map fit (**Fig. 2A,B** and **Supplemental Table S2**). Similarly, Servalcat generates the best CC_{mask} values among all the methods tested, but with the poorest MolProbity scores among all re-refined structures. In contrast, KNexPHENIX significantly lowered MolProbity scores while generally maintaining the model-to-map fit in a range similar to the deposited structures, thereby achieving a compromise between the model quality and map fit (**Fig. 2A,B**).

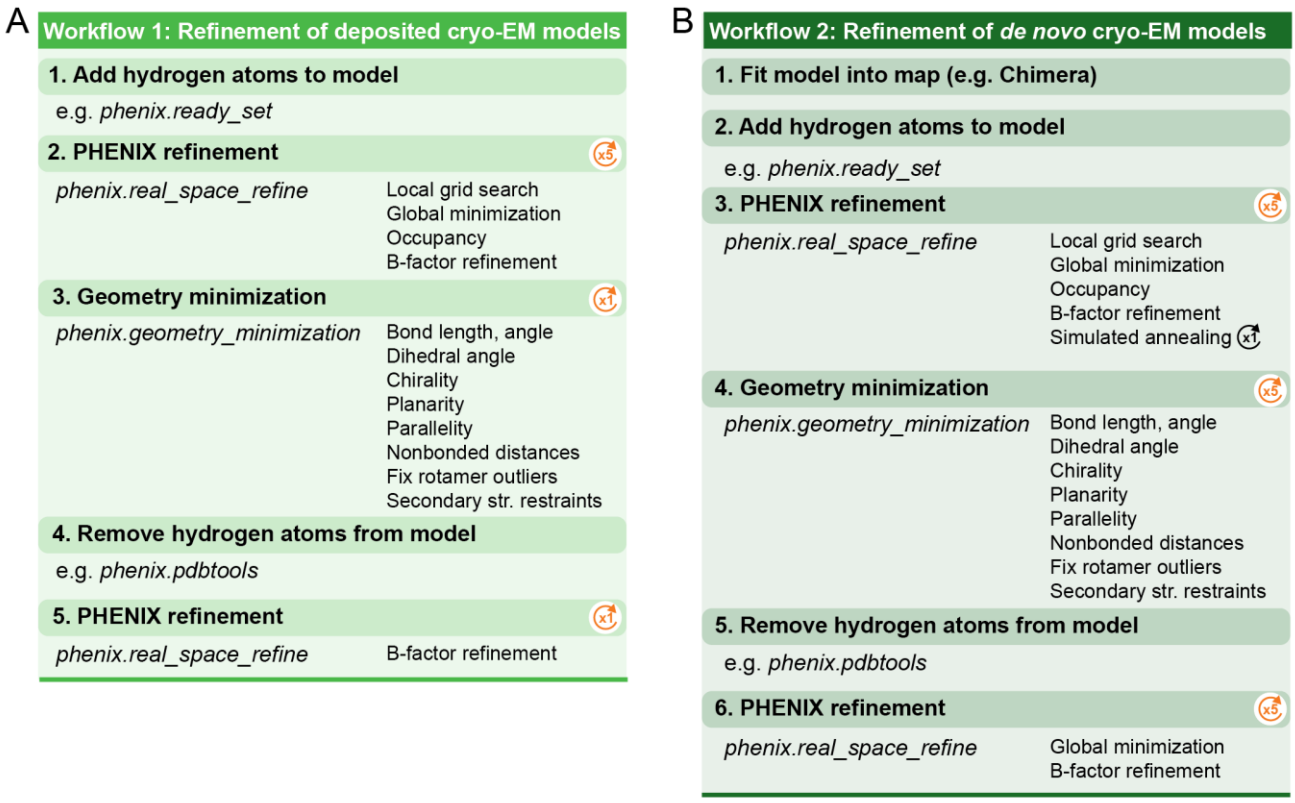

Fig. 1. KNexPHENIX workflows for refinement of structures determined by cryoEM. ***A****,* Workflow 1 for refinement of existing cryo-EM models against their corresponding maps. ***B****,* Workflow 2 for refinement of new cryo-EM models in the process of *de novo* structure determination.

Among the thirteen structures selected for our analyses, 10 were available on the CERES platform, and we therefore additionally compared MolProbity scores and CC_{mask} values on CERES to those from refinement in PHENIX, Servalcat, and KNexPHENIX. Although CERES produces improved model quality based on MolProbity scores compared to PHENIX and Servalcat, KNexPHENIX is capable of further improvement (**Fig. 2C,D** and **Supplemental Table S3**). However, the model-to-map fit for CERES-refined structures is better compared to KNexPHENIX, with Servalcat performing best in this metric (**Fig. 1C,D**). Overall, these analyses indicate that KNexPHENIX generates models that fit their corresponding maps at least comparably to the deposited model, but with improved MolProbity scores compared to PHENIX, Servalcat, or CERES.

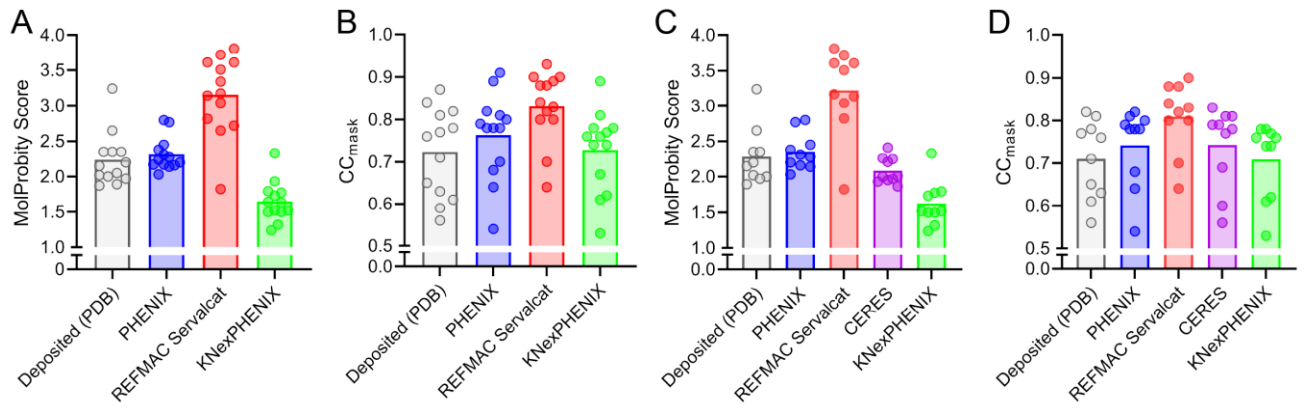

Fig. 2. KNexPHENIX enhances deposited cryo-EM model quality compared to other refinement procedures without affecting model-to-map fit. *A,* KNexPHENIX, but not PHENIX and Servalcat, reduces the MolProbity score of the 13 re-refined structures compared to the deposited PDB models. ***B,*** The CC_{mask} for structures refined with KNexPHENIX remains consistent while those for PHENIX and Servalcat show modest improvement compared to the deposited structures (PDB). ***C,*** CERES improves the MolProbity score of the 10 structures compared to PHENIX and Servalcat but not KNexPHENIX which performs best on this metric. ***D,*** The CC_{mask} values for PHENIX, CERES and KNexPHENIX refined structures remain consistent while Servalcat improves this model-to-map fit metric.

KNexPHENIX refinement of initial cryo-EM models enhances final model quality

To determine whether KNexPHENIX can assist in improving the quality of an initial cryo-EM model to produce a final model that is suitable for deposition in PDB, we tested our method (KNexPHENIX Workflow 2; **Fig.1B**) on ten structures and compared the results with those obtained from PHENIX and Servalcat. This structure set included nine of the thirteen models analyzed previously and one new model that we have recently published[25] using the method described here, but which underwent additional manual model building post-refinement (**Supplemental Table S1**). The selection was again based on ensuring diversity of macromolecule type (protein and protein-RNA complexes) and size (kDa and MDa). The initial models for fitting and refinement to the deposited maps were chosen based on their structural similarity to existing structures for those maps. As before, KNexPHENIX consistently generates models with lower MolProbity

scores compared to PHENIX and Servalcat, with Servalcat producing models with the best model-to-map fit (**Fig. 3A,B** and **Supplemental Table S4**). However, the CC_{mask} for both PHENIX and KNexPHENIX is similar (**Fig. 3B**). Where additional manual adjustment of the KNexPHENIX-refined model is necessary, refinement of the updated model may be continued using the workflow presented earlier for deposited structures for quality enhancement. Overall, KNexPHENIX generates improved model quality with negligible effect on map-to-model fit compared to standard PHENIX refinement.

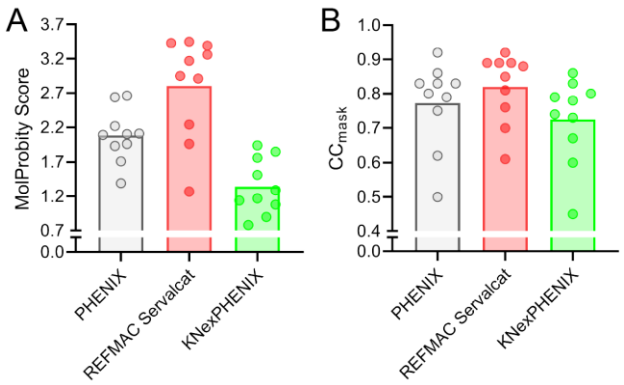

Fig. 3. Refinement of predicted initial cryo-EM models using KNexPHENIX enhances model quality. *A,* MolProbity scores of the 10 structures refined by KNexPHENIX are significantly better than the initial models processed by PHENIX or Servalcat. ***B,*** CC_{mask} from Servalcat is higher than the other two refinement procedures.

Improving model quality of deposited X-ray crystal structures using KNexPHENIX

Next, we wanted to test the performance of KNexPHENIX on existing crystal structures deposited in the PDB to assess whether the process can result in model improvement (KNexPHENIX Workflow 3; **Fig. 4A**). Sixteen protein structures were chosen from the PDB based on their type (proteins and protein-protein complexes), size (~15 to ~230 kDa), and displaying indicators of suboptimal model quality (**Supplemental Table S1**). Along with MolProbity score, values for R_{work}, R_{free}, and R_{free}-R_{work}, which can indicate model overfitting[26], were used to determine the effectiveness of KNexPHENIX in comparison to the deposited structures as well as re-refinement using PHENIX and REFMAC5. As observed previously for the KNexPHENIX refined cryo-EM structures, the procedure improves existing models and performs better than PHENIX and REFMAC5 in terms of MolProbity score (**Fig. 5A** and **Supplemental Table S5**). Both PHENIX and REFMAC5 lower the R_{work} for the existing models, while this metric is unchanged for KNexPHENIX (**Fig. 5B**); in contrast, R_{free} values of the deposited structures are comparable for both REFMAC5 and KNexPHENIX and slightly elevated by PHENIX (**Fig. 5C**). Importantly, KNexPHENIX refinement consistently restricts the difference between free and work R-factors (R_{free}-R_{work}) to below 5, widely considered an acceptable value[27, 28]. In contrast, due to their greater impact on R_{work} compared to R_{free}, the average R-factor difference for the PHENIX and REFMAC5 refined models is closer to 6 (**Fig. 5D**). Altogether, our results show that KNexPHENIX can enhance the quality of the existing structures obtained by X-ray crystallography, while limiting overfitting of the model into the map and maintaining model correctness.

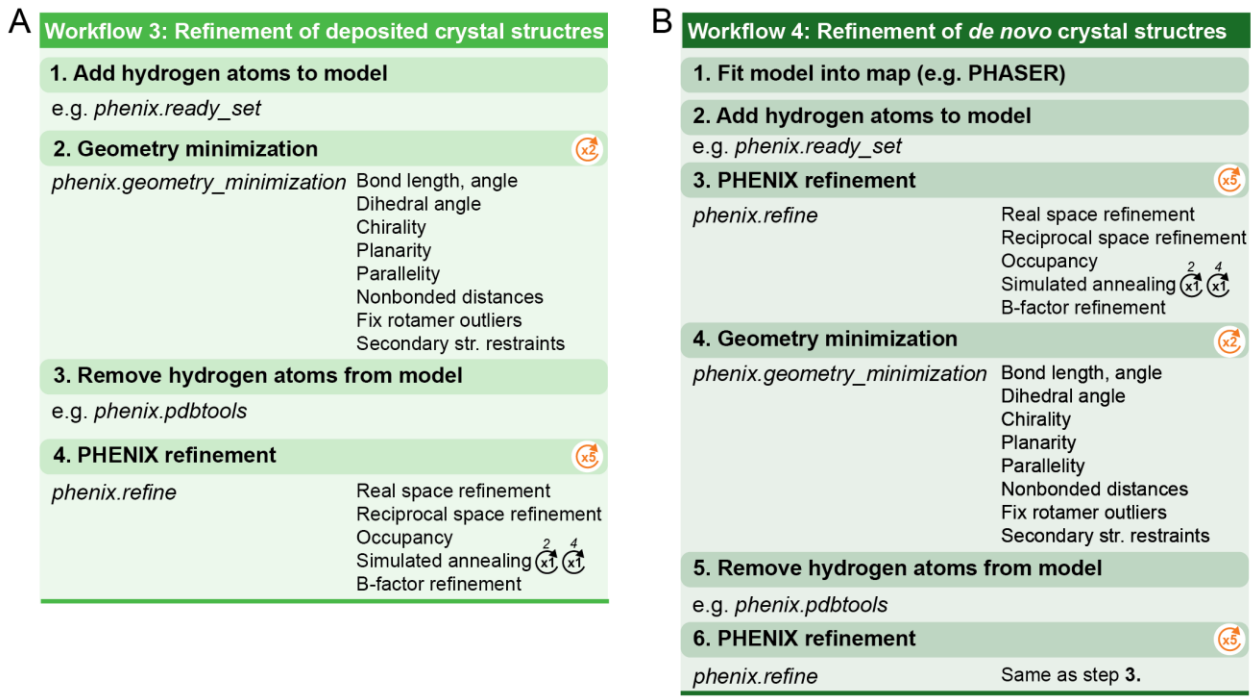

Fig. 4. KNexPHENIX workflows for refinement of structures determined by X-ray crystallography. *A*, Workflow 3 for refinement of existing X-ray crystal structures against their corresponding maps. ***B*,** Workflow 4 for refinement of new X-ray crystal structures in the process of structure determination by MR.

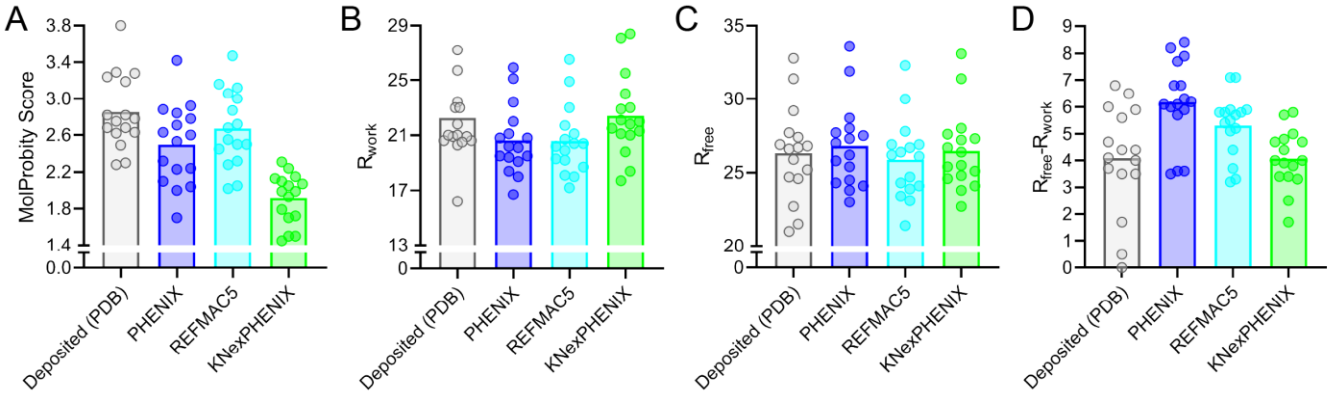

Fig. 5. KNexPHENIX improves the model quality of existing crystal structures while maintaining the fit of the model into the map. *A,* KNexPHENIX refinement significantly lowers the MolProbity score compared to PHENIX and REFMAC5 re-refined models and the original 16 deposited structures. ***B,*** Both PHENIX and REFMAC5 lower the R_{work} of the structures compared to the KNexPHENIX and deposited PDB structures ***C,*** R_{free} stays consistent for REFMAC5 and KNexPHENIX, while PHENIX increases this metric. ***D,*** PHENIX and REFMAC5 increase the difference between R_{work} and R_{free} compared to the deposited PDB structures, whereas it Is unaltered by KNexPHENIX.

MR models for crystallographic maps can be improved by KNexPHENIX

To assess whether KNexPHENIX can produce improved final models in the structure determination process by MR, we selected ten structures from the set of sixteen analyzed above. These test cases were chosen based on visual inspection of the quality of fit of the MR model to the map, as well as its similarity to the deposited model. As with other refinement methods, KNexPHENIX could not produce a better final model if this condition was not met, indicating that the method will fail in cases where MR produces a starting model with a poor model-to-map fit.

As observed previously in the refinement of deposited crystal structures, KNexPHENIX refinement of these initial models resulted in a final model with a superior MolProbity score compared to PHENIX and REFMAC5 (**Fig. 6A** and **Supplemental Table S6**). Again, the lowest R_{work} was obtained with these two methods, and REFMAC5 produced models with the lowest R_{free}, while KNexPHENIX produced models with the lowest difference between R_{free} and R_{work} (**Fig. 4B-D**). In cases where the user aims to perform further manual corrections of the model obtained from KNexPHENIX, it can subsequently be processed using the strategy outlined above for deposited crystal structures for further improvement. In general, KNexPHENIX generates final models with better quality compared to other methods, effectively preventing overfitting to the map and maintaining model accuracy.

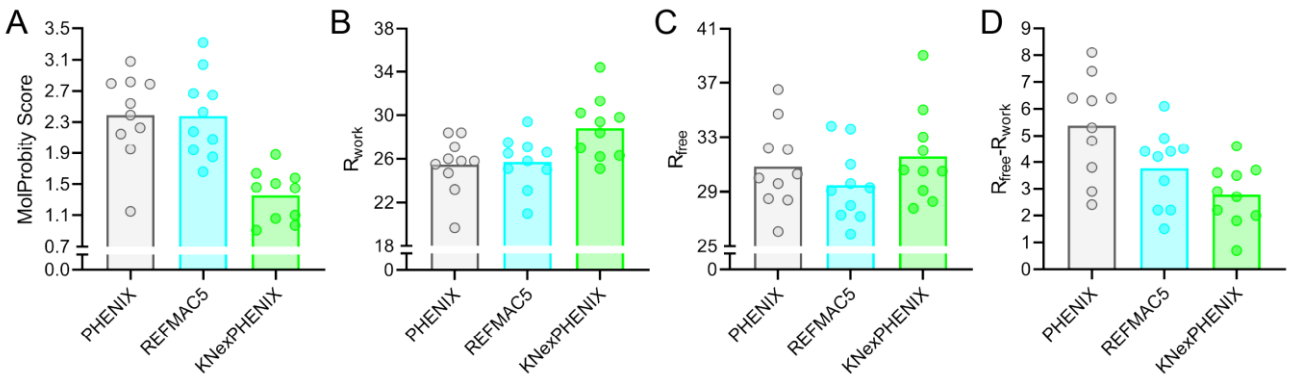

Fig. 6. KNexPHENIX refinement of MR models improves X-ray crystallographic structure quality. ***A,*** KNexPHENIX improves the *de novo* model quality of the 10 structures compared to REFMAC5 or PHENIX refinement. ***B,*** R_{work} is slightly lower for structures refined using PHENIX and REFMAC5 compared to KNexPHENIX. ***C,*** R_{free} is lower for REFMAC5 but does not vary considerably for the other two methods. ***D,*** R_{free}-R_{work} is lower in KNexPHENIX compared to the other two methods.

DISCUSSION

Structural studies using cryo-EM or X-ray crystallography require efficient and accurate model building, which is a challenge for large molecules and/ or poor-quality maps. Although several high-performance computing tools exist, many are computationally intensive and may be less readily accessible to researchers

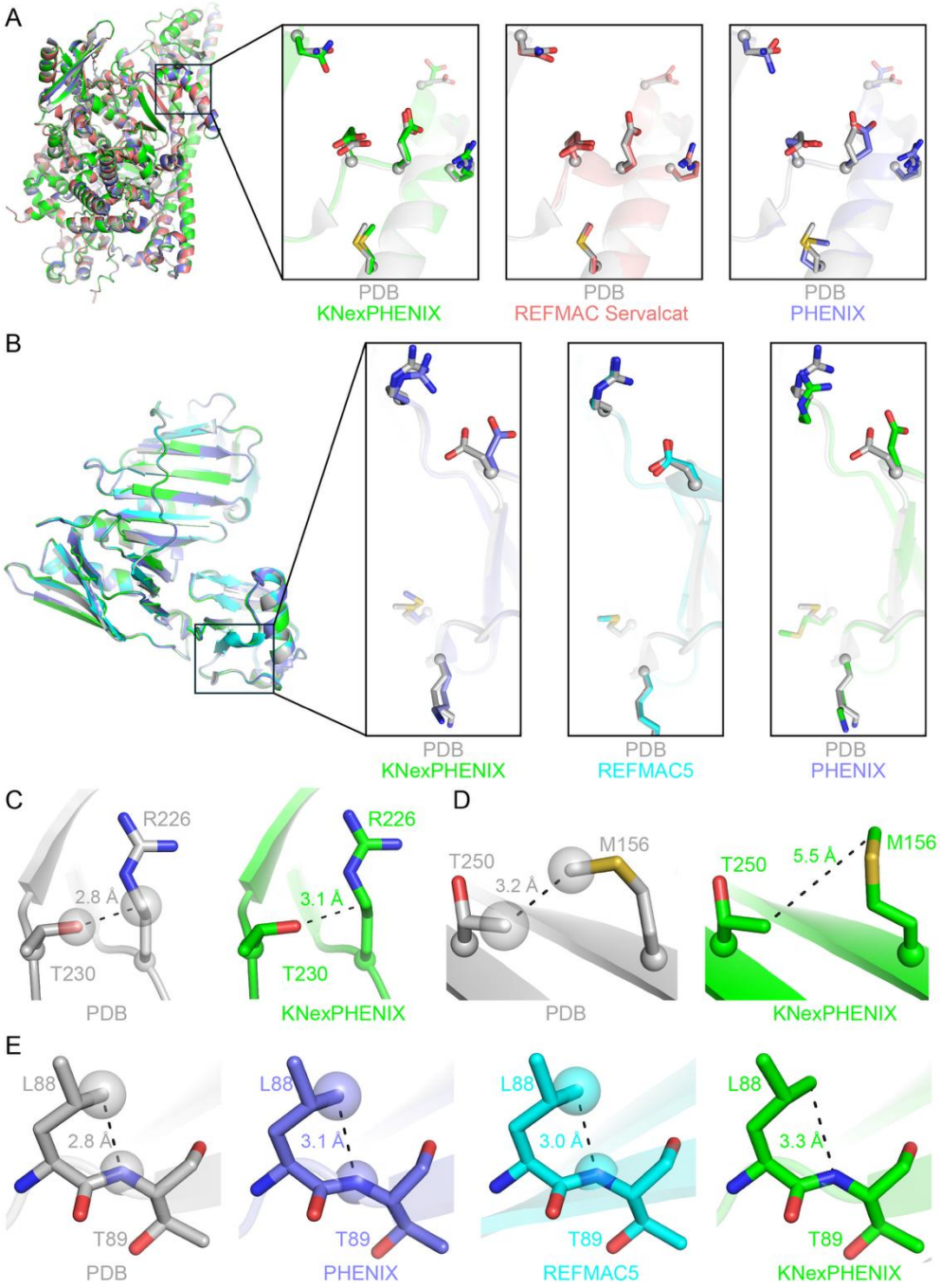

Fig. 7. KNexPHENIX refines local side chain geometry and resolves atomic clashes without compromising fit to the map. ***A,*** Protein backbone conformation is consistent across the structures refined using KNexPHENIX (green), REFMAC Servalcat (red), PHENIX (blue) for the PI3Kalpha H1047R variant determined by cryo-EM (PDB code 8GUB) and all align well with the original PDB model (gray). Sidechain position differences compared to the deposited structure are observed in the structures refined with KNexPHENIX and PHENIX, whereas REFMAC Servalcat leaves them essentially unchanged. ***B,*** The monoubiquitinated PCNA crystal structure (PDB code 3L0W) refined by KNexPHENIX (green), REFMAC5 (cyan), and PHENIX (blue) shows overall structural similarity with the PDB model (gray). REFMAC5 does not change the position of the side chains, while both PHENIX and KNexPHENIX produce more substantial movements. ***C,*** In the KNexPHENIX-refined R220A metBJFIXL HEME domain crystal structure (PDB code 1Y28), a shift of R226 resolves a steric clash with T230 present in the original PDB structure. ***D,*** Steric hindrance is also avoided by rotation of the M156 side chain from T250 by KNexPHENIX refinement. ***E,*** PHENIX and REFMAC5 do not relieve a steric clash between L88 and T89 observed in the monoubiquitinated PCNA crystal structure (PDB code 3L0W) which is successfully resolved by KNexPHENIX.

with limited resources. To address this, we developed KNexPHENIX, a set of PHENIX-based workflows that use existing programs of the PHENIX suite with customized parameters to rapidly and efficiently improve the quality of X-ray crystallographic and cryo-EM structural models.

Structural analysis of one example structure refined with KNexPHENIX, an H1047R variant of PI3Kalpha (PDB code 8GUB), suggests that while the approach did not alter the model-to-map fit, subtle adjustments of amino acid side chain positions were made that prevent clashes or improve geometry (**Fig. 7A** and **Supplemental Fig. S1A**). Interestingly, REFMAC Servalcat did not change amino acid side chain positions significantly compared to the deposited structure, with geometry quality possibly reduced, as reflected by the increased MolProbity score (**Fig. 7A**). Although the standard PHENIX refinement adjusted the side chains more extensively compared to the deposited structure, these changes did not result in a noticeable impact on the MolProbity score, underscoring the potential of KNexPHENIX as an alternative to the default PHENIX approach (**Fig. 7A**).

A similar trend was also observed for KNexPHENIX, REFMAC5, and PHENIX in refining the crystal structure of the monoubiquitinated PCNA crystal structure (PDB code 3L0W; **Fig. 7B** and **Supplemental Fig. S1B**). Again, KNexPHENIX improved the MolProbity score through localized adjustments in protein side chain positions. For instance, in the R220A metBJFIXL HEME domain crystal structure, KNexPHENIX shifts the R226 side chain away from the T230, increasing the distance between R226 Cδ and T230 OH from 2.8 to 3.1 Å and thus reducing the clash flagged by Probe in the original deposited PDB model (**Fig. 7C**). Along with translational movement, side chain rotation also plays a role in reducing the MolProbity score as observed in the repositioning of the M156 side chain away from T250 (**Fig. 7D**). Importantly, these side chain movements as well as the overall model remain visibly well fit within the density map (**Supplemental Fig. S1C,D**), as also reflected by the R_{work} and R_{free} values.

Finally, a similar inspection of the results for the monoubiquitinated PCNA structure highlights the unique way in which KNexPHENIX yields a better MolProbity score compared to the other methods. In the PDB model, Probe identified a clash between one of the methyl groups of L88 and the backbone N atom of T89 (**Fig. 7E**). While PHENIX and REFMAC5 fail to resolve this clash by moving the atoms by 3.1 Å and 3.0 Å apart, respectively, KNexPHENIX further increased this separation (to 3.3 Å), effectively resolving the clash while maintaining a good fit to the density (**Fig. 7E** and **Supplemental Fig. S1E**).

CONCLUSIONS

In summary, KNexPHENIX offers distinct advantages over default PHENIX and REFMAC refinements, particularly in optimizing local geometry without compromising the fit to experimental map, whether derived from X-ray diffraction or cryo-EM images. By improving the overall geometry of the structures, KNexPHENIX enhances the interpretability and reliability of refined models. The importance of these enhancements becomes apparent in cases where accurate side chain placement is crucial for

understanding functional mechanisms of the macromolecules or guiding downstream applications, such as designing small molecules based on specific interactions.

ASSOCIATED CONTENT

Data availability statement

All the input files to reproduce these results are available on Protein Data Bank (PDB).

Supporting Information

KNexPHENIX maintains a good model fit to map while improving model quality (Figure S1); Structures used for KNexPHENIX benchmarking against other refinement approaches (Table S1); MolProbity score (MS) and CC_{mask} calculated from 13 cryo-EM structures deposited in the PDB and after their refinement using PHENIX, REFMAC Servalcat, and KNexPHENIX (Table S2); MolProbity score (MS) and CC_{mask} calculated from 10 cryo-EM structures deposited in the PDB and after their refinement using PHENIX, REFMAC Servalcat, CERES, and KNexPHENIX (Table S3); MolProbity score (MS) and CC_{mask} calculated from *de novo* refinement of 10 cryo-EM structures by PHENIX, REFMAC Servalcat, and KNexPHENIX using maps extracted from the PDB (Table S4); MolProbity score (MS), R_{work}, R_{free}, and ΔR (R_{free}-R_{work}) calculated from 16 crystal structures deposited in the PDB and after their refinement using PHENIX, REFMAC5, and KNexPHENIX (Table S5); MolProbity score (MS), R_{work}, R_{free}, and ΔR (R_{free}-R_{work}) calculated from *de novo* refinement of 10 crystal structures by PHENIX, REFMAC5, and KNexPHENIX using maps extracted from the PDB (Table S6); KNexPHEX “How-to” guide (Workflows 1-4 with PHENIX screenshots)

AUTHOR INFORMATION

Corresponding authors

Suparno Nandi- Emory University, School of Medicine, Department of Biochemistry, Atlanta, GA, USA, 30329; orcid.org/0000-0002-3534-8002; Email- nandisuparno@gmail.com

Graeme L. Conn- Emory University, School of Medicine, Department of Biochemistry, Atlanta, GA, USA, 30329; orcid.org/0000-0002-2419-380X; Email- gconn@emory.edu

Author contributions

S.N.–conceptualization, methodology, investigation, formal analysis, validation, data curation, visualization, writing-original draft and writing-review and editing; G.L.C.–resources, supervision, funding acquisition, project administration, data curation, visualization, writing-review and editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

This work was supported by awards R01-GM130135 from the National Institute of General Medical Sciences and R01-AI088025 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

(1) Chua, E. Y. D.; Mendez, J. H.; Rapp, M.; Ilca, S. L.; Tan, Y. Z.; Maruthi, K.; Kuang, H.; Zimanyi, C. M.; Cheng, A.; Eng, E. T.; et al. Better, Faster, Cheaper: Recent Advances in Cryo-Electron Microscopy. *Annu Rev Biochem* **2022**, *91*, 1-32. DOI: 10.1146/annurev-biochem-032620-110705.

(2) Carugo, O.; Djinovic-Carugo, K. Structural biology: A golden era. *PLoS Biol* **2023**, *21* (6), e3002187. DOI: 10.1371/journal.pbio.3002187.

(3) Read, R. J.; Adams, P. D.; Arendall, W. B., 3rd; Brunger, A. T.; Emsley, P.; Joosten, R. P.; Kleywegt, G. J.; Krissinel, E. B.; Lutteke, T.; Otwinowski, Z.; et al. A new generation of crystallographic validation tools for the protein data bank. *Structure* **2011**, *19* (10), 1395-1412. DOI: 10.1016/j.str.2011.08.006.

(4) Afonine, P. V.; Klaholz, B. P.; Moriarty, N. W.; Poon, B. K.; Sobolev, O. V.; Terwilliger, T. C.; Adams, P. D.; Urzhumtsev, A. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr D Struct Biol* **2018**, *74* (Pt 9), 814-840. DOI: 10.1107/S2059798318009324.

(5) Lawson, C. L.; Kryshtafovych, A.; Adams, P. D.; Afonine, P. V.; Baker, M. L.; Barad, B. A.; Bond, P.; Burnley, T.; Cao, R.; Cheng, J.; et al. Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. *Nat Methods* **2021**, *18* (2), 156-164. DOI: 10.1038/s41592-020-01051-w.

(6) Burley, S. K.; Berman, H. M.; Chiu, W.; Dai, W.; Flatt, J. W.; Hudson, B. P.; Kaelber, J. T.; Khare, S. D.; Kulczyk, A. W.; Lawson, C. L.; et al. Electron microscopy holdings of the Protein Data Bank: the impact of the resolution revolution, new validation tools, and implications for the future. *Biophys Rev* **2022**, *14* (6), 1281-1301. DOI: 10.1007/s12551-022-01013-w.

(7) Zhang, B.; Zhang, X.; Pearce, R.; Shen, H. B.; Zhang, Y. A New Protocol for Atomic-Level Protein Structure Modeling and Refinement Using Low-to-Medium Resolution Cryo-EM Density Maps. *J Mol Biol* **2020**, *432* (19), 5365-5377. DOI: 10.1016/j.jmb.2020.07.027.

(8) DiMaio, F.; Echols, N.; Headd, J. J.; Terwilliger, T. C.; Adams, P. D.; Baker, D. Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat Methods* **2013**, *10* (11), 1102-1104. DOI: 10.1038/nmeth.2648.

(9) Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **2003**, *125* (7), 1731-1737. DOI: 10.1021/ja026939x.

(10) Igaev, M.; Kutzner, C.; Bock, L. V.; Vaiana, A. C.; Grubmuller, H. Automated cryo-EM structure refinement using correlation-driven molecular dynamics. *Elife* **2019**, *8*. DOI: 10.7554/eLife.43542.

(11) Nikolaidis, E., Pandey, V., and Mourelatos, Z. Managing the Computational Cost in a Monte Carlo Simulation by Considering the Value of Information. *SAE Technical Paper* **2012**, Technical Paper. DOI: 10.4271/2012-01-0915.

(12) Lazim, R.; Suh, D.; Choi, S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *Int J Mol Sci* **2020**, *21* (17). DOI: 10.3390/ijms21176339.

(13) *Crystallographic refinement with Rosetta (phenix.rosetta_refine)*. https://phenix-online.org/documentation/reference/rosetta_refine.html (accessed 2025 9/24).

(14) Kimera, R.; Whalen, C. J.; Economou, M. X. HPC Journey in a Resource Constrained Environment: A case of The African Center of Excellence in Bioinformatics and Data Intensive Science, Uganda. In Practice and Experience in Advanced Research Computing 2024: Human Powered Computing, Providence, RI, USA; 2024.

(15) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 2), 213-221,

(16) Afonine, P. V.; Poon, B. K.; Read, R. J.; Sobolev, O. V.; Terwilliger, T. C.; Urzhumtsev, A.; Adams, P. D. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **2018**, *74* (Pt 6), 531-544. DOI: 10.1107/S2059798318006551.

(17) Wang, R. Y.; Song, Y.; Barad, B. A.; Cheng, Y.; Fraser, J. S.; DiMaio, F. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife* **2016**, *5*. DOI: 10.7554/eLife.17219.

(18) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **2011**, *67* (Pt 4), 355-367. DOI: 10.1107/S0907444911001314.

(19) Yamashita, K.; Palmer, C. M.; Burnley, T.; Murshudov, G. N. Cryo-EM single-particle structure refinement and map calculation using Servalcat. *Acta Crystallogr D Struct Biol* **2021**, *77* (Pt 10), 1282-1291. DOI: 10.1107/S2059798321009475.

(20) Liebschner, D.; Afonine, P. V.; Moriarty, N. W.; Poon, B. K.; Chen, V. B.; Adams, P. D. CERES: a cryo-EM re-refinement system for continuous improvement of deposited models. *Acta Crystallogr D Struct Biol* **2021**, *77* (Pt 1), 48-61. DOI: 10.1107/S2059798320015879.

(21) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605-1612. DOI: 10.1002/jcc.20084.

(22) Rossmann, M. G. The molecular replacement method. *Acta Crystallogr A* **1990**, *46 (Pt 2)*, 73-82. DOI: 10.1107/s0108767389009815.

(23) Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **2010**, *66* (Pt 1), 12-21. DOI: 10.1107/S0907444909042073.

(24) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **2007**, *35* (Web Server issue), W375-383. DOI: 10.1093/nar/gkm216.

(25) Nandi, S.; Dey, D.; Srinivas, P.; Dunham, C. M.; Conn, G. L. Distant ribose 2'-O-methylation of 23S rRNA helix 69 pre-orders the capreomycin drug binding pocket at the ribosome subunit interface. *Nucleic Acids Res* **2025**, *53* (13). DOI: 10.1093/nar/gkaf618.

(26) Shabalin, I. G.; Porebski, P. J.; Minor, W. Refining the macromolecular model - achieving the best agreement with the data from X-ray diffraction experiment. *Crystallogr Rev* **2018**, *24* (4), 236-262. DOI: 10.1080/0889311X.2018.1521805.

(27) Wang, J. Estimation of the quality of refined protein crystal structures. *Protein Sci* **2015**, *24* (5), 661-669. DOI: 10.1002/pro.2639.

(28) Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J* **2008**, *275* (1), 1-21. DOI: 10.1111/j.1742-4658.2007.06178.x.