AQuaRef: Machine learning accelerated quantum refinement of protein structures

Roman Zubatyuk[1,#], Malgorzata Biczysko[2,#], Kavindri Ranasinghe[3,#], Nigel W. Moriarty[4]
Hatice Gokcan[1], Holger Kruse[6], Billy K. Poon[4], Paul D. Adams[4,5] , Mark P. Waller[6],
Adrian E. Roitberg[3], Olexandr Isayev[1,*], Pavel V. Afonine[4, *]

[1] Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213, USA

[2] Faculty of Chemistry, University of Wrocław, F. Joliot-Curie 14, 50-383 Wrocław, Poland

[3] Department of Chemistry, University of Florida, Gainesville, FL 32611, USA

[4] Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720-8235, USA

[5] Department of Bioengineering, University of California Berkeley, Berkeley, CA 94720, USA

[6] Pending.AI, Eveleigh, NSW 2015, Australia

# These authors contributed equally

*Corresponding authors: P.V.A. - PAfonine@LBL.Gov,
O.I. - olexandr@olexandrisayev.com

ABSTRACT

Cryo-EM and X-ray crystallography provide crucial experimental data for obtaining atomic-detail models of biomacromolecules. Refining these models relies on library-based stereochemical restraints, which, in addition to being limited to known chemical entities, do not include meaningful noncovalent interactions relying solely on nonbonded repulsions. Quantum mechanical (QM) calculations could alleviate these issues but are too expensive for large molecules. We present a novel AI-enabled Quantum Refinement (AQuaRef) based on AIMNet2 neural network potential mimicking QM at substantially lower computational costs. By refining 41 cryo-EM and 30 X-ray structures, we show that this approach yields atomic models with superior geometric quality compared to standard techniques, while maintaining an equal or better fit to experimental data. Notably, AQuaRef aids in determining proton positions, as illustrated in the challenging case of short hydrogen bonds in the parkinsonism-associated human protein DJ-1 and its bacterial homolog YajL.

INTRODUCTION

While advances in predictive modeling, such as AlphaFold3[1] or RoseTTAFold[2,3], have provided powerful tools for structural biology, they remain limited while experimental methods, including protein crystallography and cryo-EM, are still cornerstones of structural biology and drug development[4]. Experimental data allow for the discovery of new structures emerging in life evolution, potentially exhibiting previously unseen features. These discoveries require unbiased information provided by experiments to explore the unknown[5]. Atomic model refinement is a crucial near-final stage in crystallographic or cryo-EM structure determination aimed at producing molecular models that meet standard validation criteria while optimally fitting the experimental data[6]. Refinement heavily relies on stereochemical restraints to maintain the correct geometry of the atomic model while fitting to the experimental data[7]. These restraints originate from standard libraries that tabulate the topology and parameters of known chemical entities[8,9],

which are universally employed across popular software packages, such as CCP4[10] and Phenix[11].

The limitations of library-based restraints are manifold. Firstly, they only include terms for maintaining covalent bond lengths, bond angles, torsion angles, planes and chirality while preventing clashes through non-bonded repulsion[12]. However, it has been demonstrated that at low resolution, these restraints are insufficient to maintain realistic, chemically meaningful macromolecular geometries, making it essential to include additional restraints on protein main chain φ/ψ angles, side chain torsion χ angles, as well as hydrogen bond parameters and π-stacking interactions to stabilize protein or nucleic acid secondary structure[12–18]. These additional restraints cannot be reliably inferred from the atomic model alone and thus require manual error-prone annotation and curation using additional sources of information, such as homologous high-resolution models. Secondly, library-based restraints parametrize only known chemical entities, such as standard amino and nucleic acids as well as previously defined ligands. Consequently, any nonstandard entities or interactions, such as novel ligands or covalent cross-chain links, require manual annotation and definition, without which refinement may fail to proceed correctly or at all. Finally, deviations from standard covalent geometry due to local chemical interactions are not uncommon[19–21]. While these deviations are valid, restraints may interpret them as violations requiring 'correction'.

The advantage of using simple restraints[7] is the minimal computational cost they add to the refinement workflow. A possible next step is to use a classical force field to account for geometric elements[22]. However, these force fields have their own set of limitations: they require parametrization for new chemical species and cannot distinguish between chemically equivalent bonds in different chemical environments.

Quantum refinement is a fundamentally different approach, balancing the fitting to experimental data with a term related to the quantum mechanical energy of the system[23,24]. It has been demonstrated that the entire atomic model can benefit from a full QM treatment[25–27]. Figure 1 presents a timeline showcasing the evolution of quantum mechanics calculations for proteins, highlighting four key stages of progress and advancements in technology and methodology. Traditionally, quantum refinements were

deemed impractical for macromolecules due to the computational requirements. Methods often focused solely on the macromolecular region of interest, such as a ligand-binding pocket or enzyme active site, while employing a classical approach for the rest of the molecule[28,29]. Numerous approaches and implementations have been reported over time[30], with GPU-accelerated codes enabling QM calculations for peptides and small proteins of a few hundred atoms being one of the most prominent milestones[31]. Interaction-based model partitioning into chemically meaningful fragments[32] solved the scalability issue in quantum calculations[26], which in turn enabled the refinement of larger proteins. However, this approach remained computationally demanding.

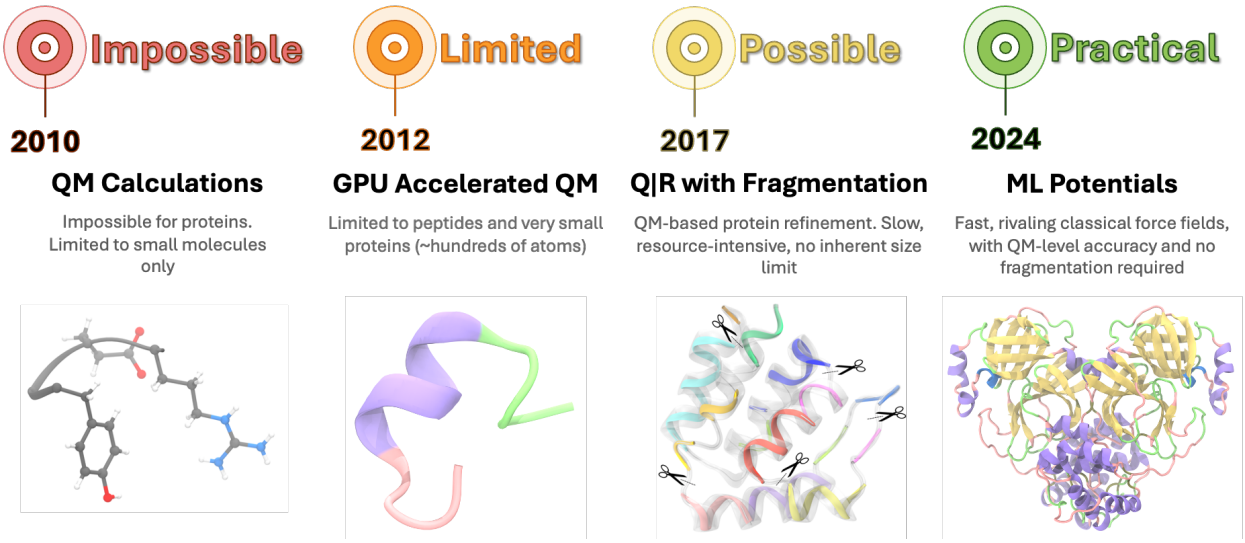


Figure 1. Timeline illustrating the progression of quantum mechanics calculations applied to proteins, emphasizing four critical stages marked by advancements in technology and methodology.

Refinement of selected cryo-EM and X-ray atomic models across various resolutions demonstrates the AQuaRef's ability to produce atomic models with superior geometric quality compared to conventional techniques while maintaining or improving agreement with experimental data. This work represents the first example where machine learning (ML) potentials have been adopted to perform quantum refinement of the entire protein,

in contrast with a recent approach where ML potentials have been combined with the ONIOM-like QM/MM partitioning[38].

RESULTS

Conceptually, quantum-based atomic model refinement is very similar to classic refinement wherein atomic model parameters are iteratively adjusted in order to minimize the residual, $T = T_{data} + w * T_{restraints}$. Here, T_{data} describes the fit of the model to the data and $T_{restraints}$ incorporates chemical restraints with an *a-priori* unknown weight, w[39].

However, there are four fundamental differences. First, in QM refinement, restraints are derived from quantum-mechanical calculations for the specific macromolecule in consideration. Second, the requirements for the initial atomic model in QM refinement are stricter compared to standard refinement: the atomic model must be correctly protonated, atom-complete and free of severe geometric violations such as steric clashes or broken covalent bonds[23,24]. Third, while crystallographic software packages inherently account for crystal symmetry, QM codes generally do not. Fourth, crystallographic software is capable of handling static disorder that is modeled with alternative conformations, whereas QM codes typically lack this capability. All these nuances specific to quantum refinement (except handling static disorder, which is a current limitation) are addressed in the Quantum Refinement package (Q|R)[23,26,27,40], which is being developed as part of this work and provides the necessary procedures to enable quantum refinement within the Phenix software.

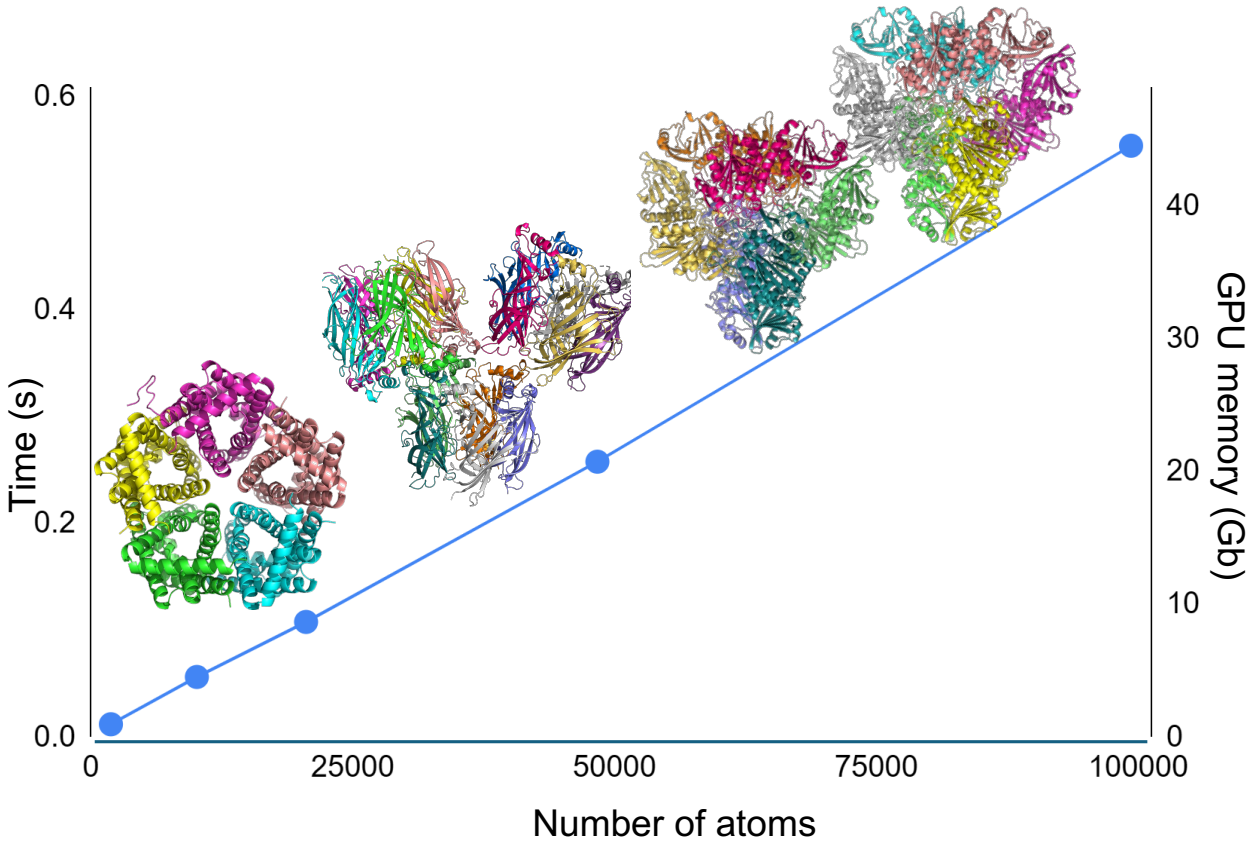


Figure 2. Computational scaling of the AIMNet2 neural network model in AQuaRef. Time to compute energy and forces (left axis) and peak GPU memory usage (right axis) versus the number of atoms in the system. Calculations are performed on a single Nvidia H100 PCIE 80GB GPU.

Conventional QM methods like density functional theory (DFT) for *N*-electron systems require $O(N^2)$ storage and $O(N^3)$ arithmetic operations. This $O(N^3)$ complexity is a critical bottleneck that limits the ability to study large realistic biological systems like proteins. Figure 2 shows the computational scaling of the AIMNet2 model, where both energy and force calculations, as well as peak GPU memory usage, scale linearly ($O(N)$) with system size. For a large protein system of 100,000 atoms, single-point energy and forces can be computed in 0.5 seconds. Overall, an atomic model consisting of approximately 180,000 atoms can fit into the 80GB memory of a single NVIDIA H100 GPU.

We tested the new quantum refinement procedure on 41 cryo-EM atomic models, 20 low-resolution and 10 very high-resolution X-ray atomic models. Standard stereochemistry[41,42] and model-to-data fit criteria[43–45], MolProbity validation tools[46] along with newly developed metrics to evaluate hydrogen bond quality[18] were used to assess the atomic models. Typically, the time needed for quantum refinement is about twice as long as standard refinement, and often shorter than the standard refinement with additional restraints such as the Ramachandran plot, secondary structure and side-chain rotamer restraints[47–50]. Quantum refinement takes under 20 minutes in about 70% of models considered in this work, with a maximum of about 1 hour (Supplementary Data: Table 6). These computations can be performed on GPU-equipped laptops, with the only limitation being available GPU memory.

Quantum refinement

The AQuaRef refinement procedure begins with a check for the completeness of the atomic model, followed by the addition of any missing atoms. This may result in steric clashes, particularly if the model was previously refined without hydrogen atoms. Models with missing atoms that cannot be trivially added (e.g., missing main chain atoms) cannot be used for quantum refinement. If clashes or other severe geometric violations are detected, quick geometry regularization is performed using standard restraints, ensuring that atoms move as little as necessary to resolve the clashes. For crystallographic refinement, to account for interactions arising from crystallographic symmetry and periodicity of unit cells, the model is expanded into a supercell by applying appropriate space group symmetry operators[25]. Subsequently, it is truncated to retain only parts of the symmetry copies within a prescribed distance from atoms of the main copy[40]. This step is unnecessary for refinement against cryo-EM data. The atom-completed and expanded model then undergoes the standard atomic model refinement protocol as implemented in Q|R package[23].

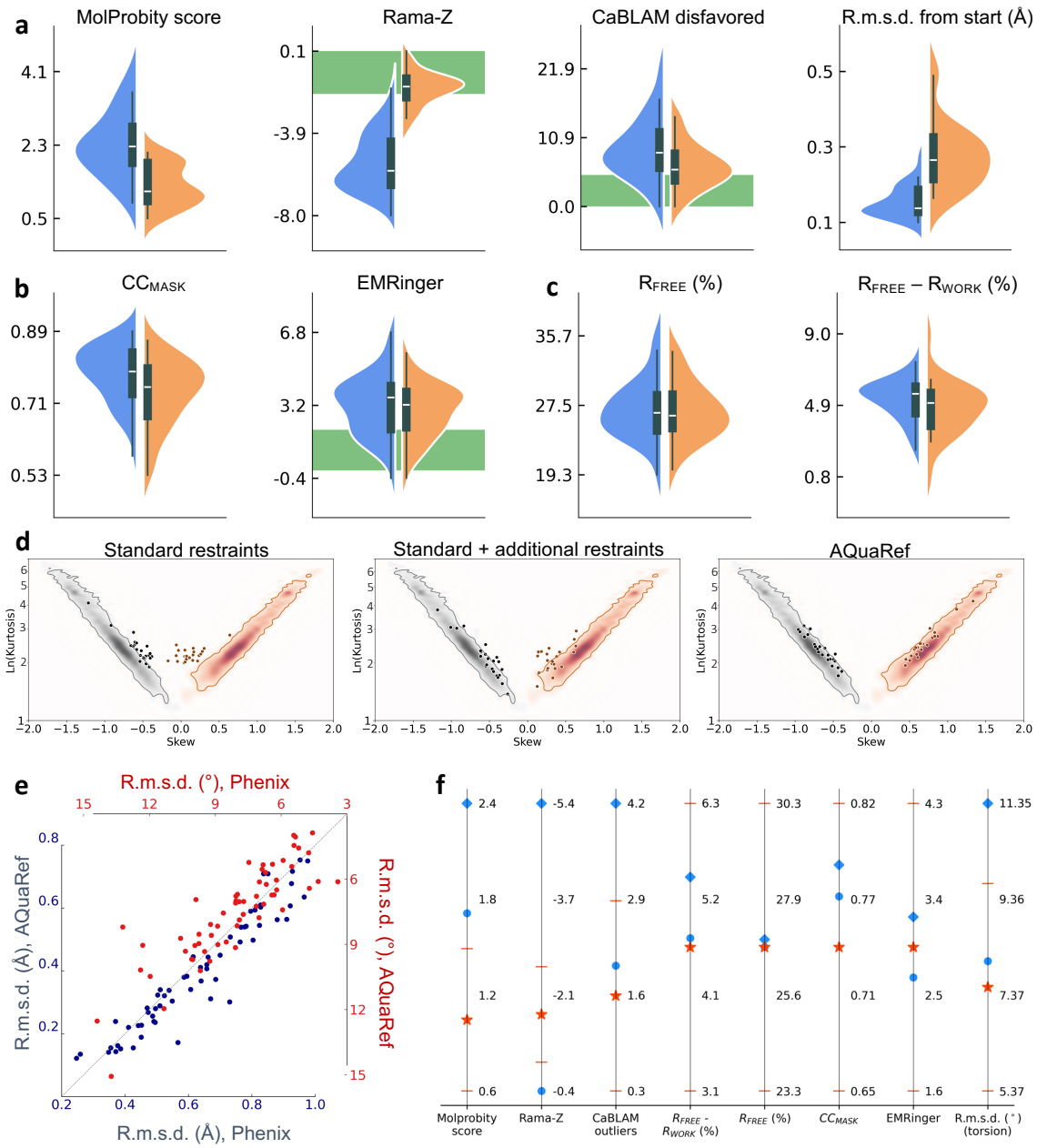


Figure 3. a-d: Summary of refinements of 41 low-resolution cryo-EM models and 20 low-resolution X-ray models using standard stereochemistry (blue) and AQuaRef (orange) restraints (Supplementary Data: Table 1). **a:** MolProbity score, Ramachandran plot Z-score, CaBLAM disfavored and r.m.s. deviation of refined model from initial model. **b:** cross-correlation between experimental and model-generated maps (CC_{mask}), and EMRinger score for cryo-EM models. **c:** R_{free} and R_{free}-R_{work} for X-ray models

(Supplementary Data: Table 3). Green band indicates favored range of corresponding values. **d:** skew-kurtosis plots for hydrogen bond parameters (Hydrogen(H)...Acceptor(A) distances and Donor-H…A angles) for refinements using (left-to-right): standard restraints; standard restraints with addition of Ramachandran plot, secondary-structure and side-chain rotamer restraints; and AQuaRef restraints. **e:** r.m.s. deviations between refined and high-resolution homology models, refinements using standard versus AQuaRef restraints, calculated using matching Cartesian coordinates (blue, lower-left) and matching torsion angles (red, upper-right) (Supplementary Data: Tables 2,4). **f:** summary of mean values, for all test refined models: MolProbity score, Ramachandran Z-score, CaBLAM outliers, r.m.s. deviation of matching torsion angles between refined and high-resolution homologous models, as well as R_{free}-R_{work} and R_{free} for X-ray models and CC_{mask} and EMRinger score for cryo-EM models for refined models with standard restraints (blue rhombi), standard restraints with addition of Ramachandran plot, secondary-structure and side-chain rotamer restraints (blue circles); and AQuaRef restraints (red stars). Red bars show standard deviations for starred values.

Application of the new refinement procedure to a set of deposited atomic models

To evaluate the performance of the new QM-based refinement, we refined 41 low-resolution cryo-EM atomic models, 20 low-resolution and 10 ultra-high-resolution X-ray atomic models, which contain only proteins. All selected 61 low-resolution atomic models have high-resolution homologs, which were used as the ground truth for comparison (Supplementary Data: Tables 2,4). Refinements were carried out using three sets of restraints: QM restraints from AIMNet2 (AQuaRef refinement); standard restraints; and standard restraints plus additional restraints on hydrogen bonds and angles involved in maintaining secondary structure, main-chain φ/ψ angles (Ramachandran plot restraints) and side-chain torsion χ angles (rotamer restraints).

Overall, low-resolution atomic models after quantum refinement exhibit systematically superior geometry quality compared to those obtained using standard restraints, as indicated by their MolProbity scores[51], Ramachandran Z-scores[52], CaBLAM disfavored[46] (Fig. 3a), and skew-kurtosis plots for hydrogen bond parameters[18] (Fig. 3d). They also

systematically deviate more from the initial coordinates. These atomic models demonstrate a very similar fit to the experimental data (Fig. 3b,c), with slightly less data overfitting for X-ray atomic models, as evidenced by a smaller R_{work}-R_{free} gap and similar R_{free}[53,54]. Since there is no equally efficient control over overfitting in cryo-EM as there is with R_{free} in crystallography, the slightly lower cross-correlation between experimental and model-calculated masked maps (CC_{mask})[43] and essentially the same EMRinger scores[55], together with significantly improved atomic model geometry, likely indicate a reduction in overfitting. Augmenting standard restraints with secondary structure, Ramachandran plot and side-chain rotamer restraints expectedly improves the geometry (Fig. 3d,f), yet using AQuaRef still produces superior atomic model geometries. With a few exceptions, atomic models refined with quantum restraints are systematically closer to their higher-resolution homologs compared to those using standard restraints alone or complemented with additional restraints (Fig. 3e,f). In some of the most remarkable cases, the local structure obtained with AQuaRef restraints closely matches the high-resolution homologs and differs from those obtained using standard restraints by up to two Angstroms (Fig. 9a-c).

Comparison with alternative state-of-the-art approaches

To further evaluate the performance of AQuaRef refinement compared to other major refinement methods and software, we refined selected low-resolution X-ray models using the AMBER force field as a source of geometric restraints[56], the Rosetta all-atom force field combined with its powerful sampling methods[57], and standard refinement as implemented in REFMAC5[58]. For cryo-EM, there are fewer refinement alternatives, with Servalcat[59] being the most popular, which we also used in this analysis.

For X-ray models, AQuaRef produced slightly better overall R_{free} values (Fig. 4a) and substantially less data overfitting, as indicated by the R_{free}-R_{work} gap (Fig. 4b). For cryo-EM models Servalcat lead to notably better CC_{mask} (Fig. 4c) and both scored the same by EMRinger method (Fig. 4d). Models refined using AQuaRef and Rosetta performed similarly well in terms of Rama-Z scores, achieving excellent results in most cases, while REFMAC5 and Servalcat had the worst scores, and AMBER fell somewhere in between (Fig. 4e). In terms of MolProbity scores and CaBLAM outliers (Fig. 4f,g), AQuaRef and

Rosetta also performed similarly well, significantly outperforming REFMAC5 and Servalcat. Rosetta-refined models were closest to the high-resolution reference models, followed by AQuaRef (Fig. 4h). This is likely due to Rosetta's use of non-gradient optimization techniques, such as sampling and local model repacking, which have a larger convergence radius compared to the gradient-driven minimization used in other programs. Finally, AQuaRef and Rosetta both produced models that fit the expected distribution of hydrogen bond parameters (Fig. 4i, Fig. 3d), followed by AMBER. REFMAC5 and Servalcat largely failed to produce models fitting this distribution, with Servalcat performing the worst.

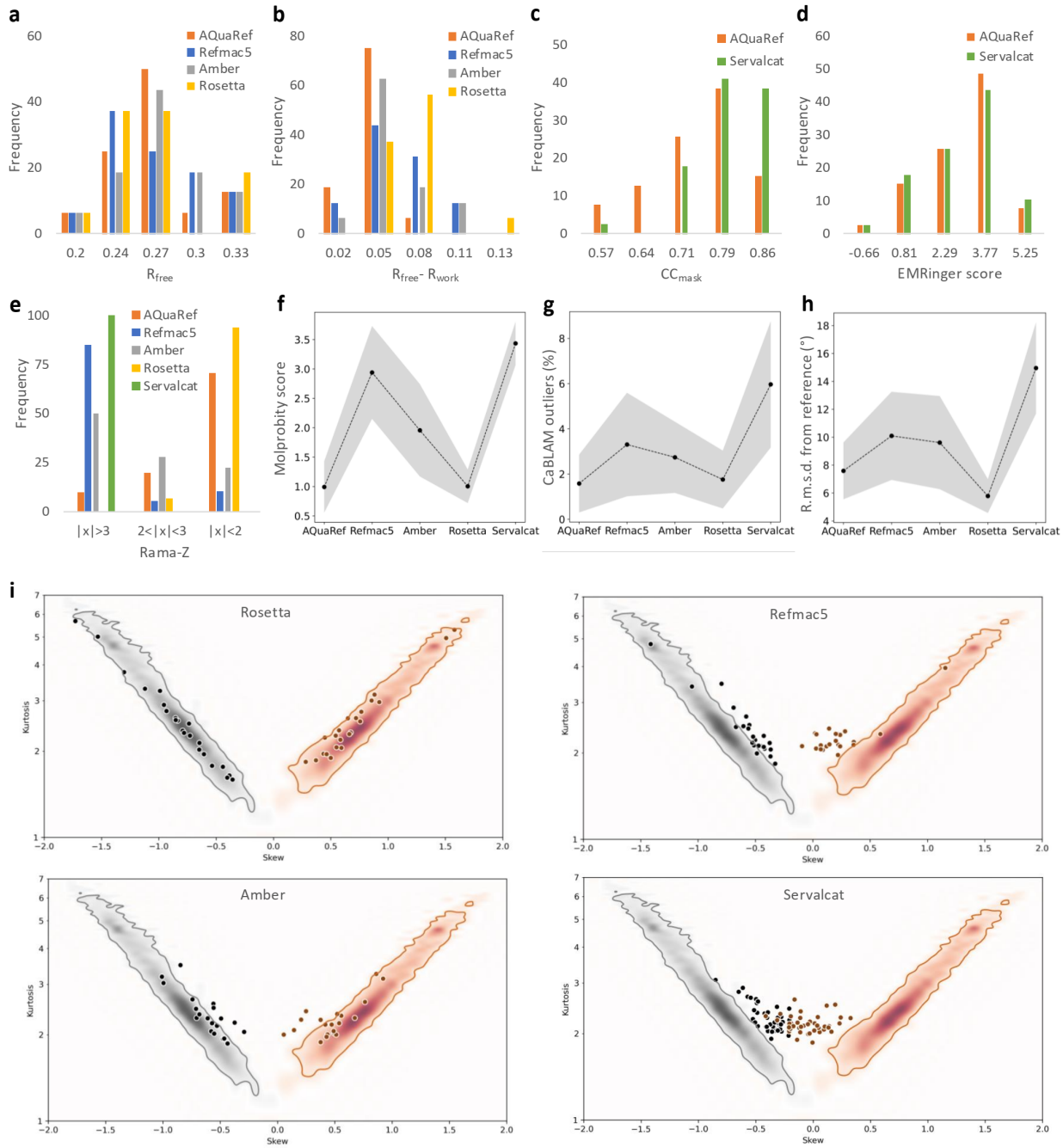


Figure 4. Summary of refinements for 41 low-resolution cryo-EM models using AQuaRef and Servalcat, and 20 low-resolution X-ray models using AQuaRef, REFMAC5, AMBER, and Rosetta. **a-e:** Distributions of R_{free}, R_{free}-R_{work}, CC_{mask}, EMRinger score, and Rama-Z, respectively. **f-h:** Mean values of MolProbity score, CaBLAM outliers, and r.m.s. deviation from the reference model, calculated across all refined models; gray bands represent the standard deviation. **i:** Skew-kurtosis plots for hydrogen bond parameters

(Hydrogen(H)...Acceptor(A) distances and Donor-H…A angles) for refinements performed using REFMAC5, AMBER, Rosetta and Servalcat.

Case study: Short hydrogen bonds in human DJ-1 and its bacterial homologue YajL

Short hydrogen bonds play a key functional role in proteins, and determining the protonation states of involved residues is critical. However, accurate location of proton positions experimentally remains challenging at resolutions near 1 Å. Lin *et al* [60] analyzed high-resolution X-ray crystal structures of human DJ-1 and its bacterial homolog YajL to determine the protonation states of carboxylic acids involved in dimer-spanning hydrogen bonds. Their approach combined bond length analysis, leveraging the distinct lengths of C=O and C–OH bonds, with qualitative interpretation of difference map peaks to identify potential evidence of protons.

This method is complicated by stereochemical restraints applied during coordinate refinement, which can bias bond lengths. For example, in E/D residues, bond length restraints for COOH groups depend on whether a hydrogen atom is explicitly modeled (Fig. 5c). To minimize this bias, Lin et al. performed final rounds of conjugate gradient least-squares refinement in SHELXL[61] without applying restraints to the residues of interest. In contrast, QM-based refinement avoids such biases entirely.

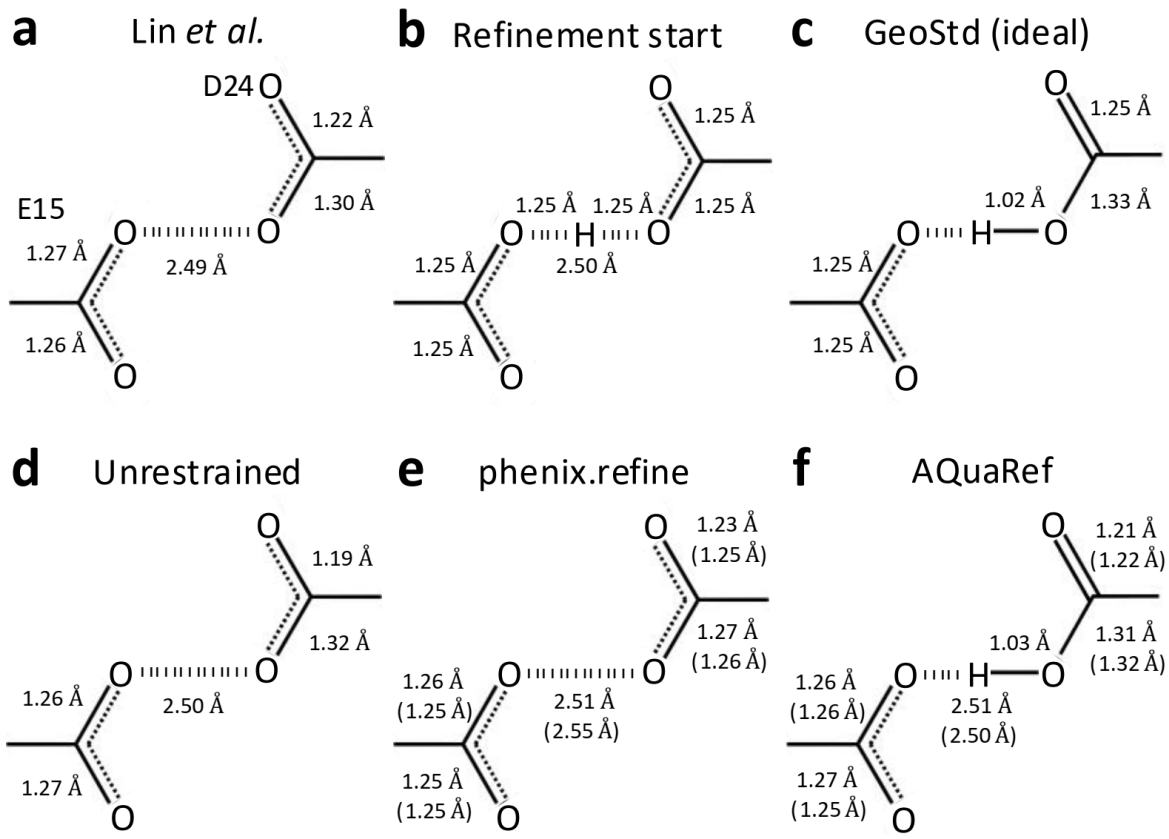


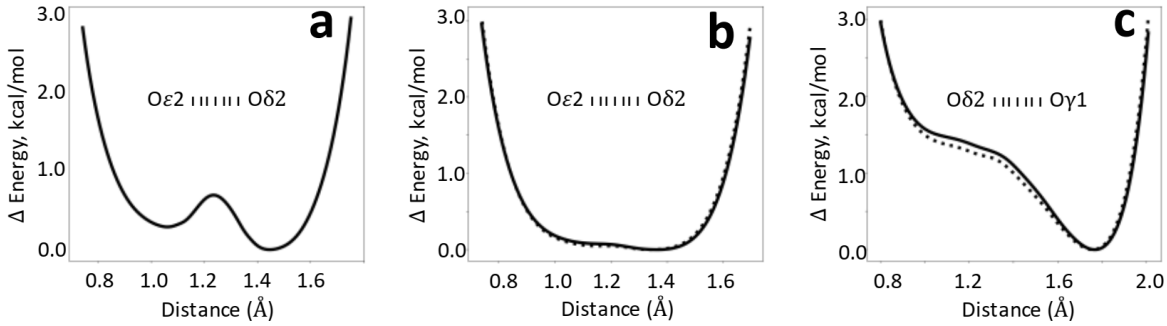
Figure 5. Wild-type DJ-1 (PDB code: 5SY6). Bond distances in the moiety of hydrogen bond between Oε2 (E15) and Oδ2 (D24), **a:** as measured in downloaded from PDB model, **b:** starting geometry for all refinements (H is present only in AQuaRef refinement), **c:** ideal library values in Phenix; geometry of –COOH or –COO groups is the same for Asp and Glu residues, **d:** unrestrained and **e:** restrained refinement with phenix.refine, **f:** refinement with AQuaRef. Distances in parentheses correspond to refinement using resolution-truncated data at 2 Å. H atom is shown only if it was explicitly modelled (present in the PDB model file).

AQuaRef refinement of DJ-1, starting with E15/D24 in an unprotonated state (Fig. 5b), produced proton positions and bond geometries (Fig 5f) consistent with Lin et al.'s findings (Fig. 5a) and unrestrained refinement using phenix.refine (Fig. 5d). However, restrained refinement with phenix.refine (Fig. 5e) yielded bond geometries that matched

289 library values assuming no proton on either COO group, highlighting the impact of
290 restraint bias.

291 To test the robustness of AIMNet2 restraints in preserving accurate geometries, the same
292 refinements were performed using experimental data truncated at 2 Å resolution. This
293 truncation removed atomic-level details that could resolve bond lengths and hydrogen
294 positions. AQuaRef produced results nearly identical to those obtained using the original
295 1.1 Å atomic resolution data, whereas restrained refinements with phenix.refine further
296 biased oxygen-carbon distances toward idealized values for the unprotonated state (Fig.
297 5e-f, values in parentheses).

298



299

300 **Figure 6.** AIMNet2 energy values relative to their minimum as a function of hydrogen
301 position between corresponding oxygen atoms, **a:** Oδ2 (D24) and Oε2 (E15) in DJ-1, **b:**
302 Oδ2 (D23) and Oε2 (E14) in YajL, and **c:** Oδ2 (D23) and Oγ1 (T16) in YajL. Solid and
303 dashed lines represent two instances of the bond in the YajL model.

304

305 Starting from an idealized symmetric arrangement (Fig. 5b), the refinement could, in
306 principle, place the proton on either E15 or D24. To explain why the proton ultimately
307 settled on Oδ2 of D24, two independent sources of evidence were considered. First,
308 sampling the hydrogen position along the Oδ2–Oε2 bond vector and computing the
309 AIMNet2 energy profile revealed a slight preference for D24 protonation (Fig. 6a).

Second, while the resolution and R-factors were insufficient for definitive proton identification in the difference map, the difference map values along the Oε2-Oδ2 axis showed elevated positive values near Oδ2, close to the prospective hydrogen position (Fig. 7a). This, together with the energetic preference from AIMNet2, may have guided the refinement to move the hydrogen toward D24.

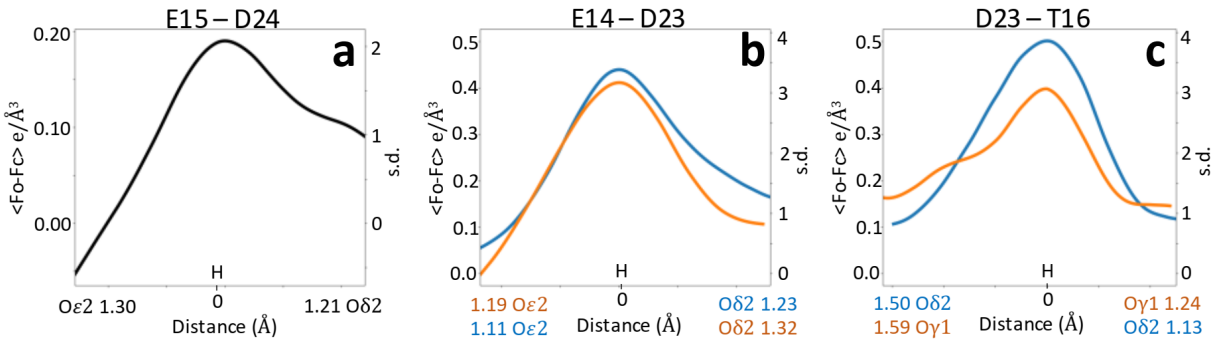


Figure 7. Mean values of the difference density map, shown in absolute units (e/Å³) and as standard deviation values along the O-H vector for the analyzed bonds for: **(a)** DJ-1 and **(b-c)** E. coli YajL models. All peak centers are aligned to the origin. Atoms belonging to chains A and B are shown in blue and orange, correspondingly.

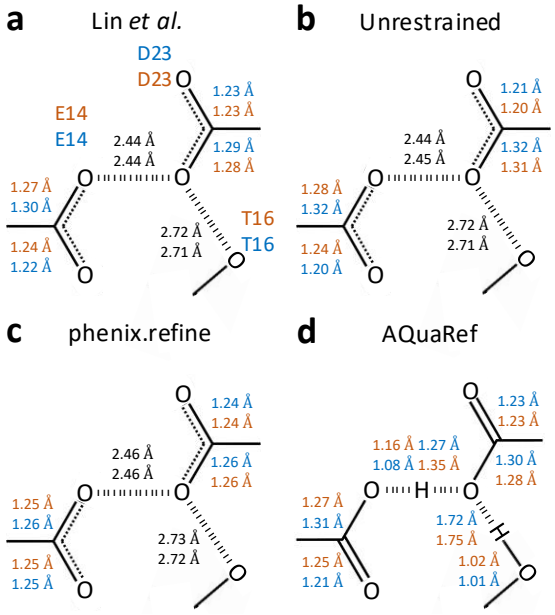


Figure 8. E. coli YajL (PDB code: 5SY4). Bond distances in the moiety of hydrogen bond between Oε2 (E14) and Oδ2 (D23) across chains A (blue) and B (orange), **a:** as measured in downloaded from PDB model, **b:** unrestrained and **c:** restrained refinement using phenix.refine, **d:** refinement with AQuaRef. H atom is shown only if it was explicitly modelled (present in the PDB model file).

Bacterial DJ-1 homologue, the YajL structure, contains two copies of the molecule in the asymmetric unit, resulting in two instances of the E14/D23 interaction. Similar to DJ-1, unrestrained refinement of YajL (Fig. 8b) yielded results consistent with Lin et al. (Fig. 8a). As with DJ-1, restrained refinement introduced significant bias in bond lengths (Fig. 8c) for both instances of the E14/D23 interaction.

Results from AQuaRef refinement aligned with Lin et al. and unrestrained phenix.refine refinement, suggesting that both D23 and E14 are protonated. However, in contrast to DJ-1, the proton in YajL does not appear to be fully associated with either of the two oxygen atoms. Instead, it seems to be shared between Oε2 and Oδ2, consistent with a Low Barrier Hydrogen Bond.

The AIMNet2 energy profile between Oε2 and Oδ2 supports this interpretation, showing a relatively flat energy landscape (Fig 6b). This indicates that the hydrogen's position could be entirely guided by the experimental data while staying within the flat region of the AIMNet2 energy well. Indeed, there is a significant difference map peaks above 3 s.d. and well above mean solvent density of 0.25 e/Å^3 very close to the position of hydrogens in the refined model in both instances of the E14/D23 interaction (Fig. 7b).

Further evidence that C-Oδ2 elongation is due to Oε2···H···Oδ2 LBHB is provided by the analysis of another hydrogen bond involving D23 and T16. All three, AQuaRef refinement (Fig 8d), the AIMNet2 energy profile (Fig. 6c), and difference map density values along the Oδ2 of D23 and Oγ1 of T16 (Fig 7c), confirm the protonation of T16 and rule out the D23 protonation in the "anti" configuration.

DISCUSSION

Here, we present AQuaRef, a novel approach to the quantum refinement of entire protein structures, made possible by using ML-accelerated quantum mechanical calculations with AIMNet2. For the first time, this allows for the refinement of full atomic models of realistic protein structures using stereochemical restraints derived from quantum mechanical calculations.

Test refinements using 61 low-resolution X-ray and cryo-EM atomic models show systematic improvements in geometric validation criteria by using QM restraints while maintaining a similar fit to the experimental data and reducing overfitting. The presence of high-resolution homologous atomic models, which are expected to better represent the actual true structures than low-resolution atomic models, allowed us to assess whether these improvements are associated with refined structures becoming closer to the true ones. With a few exceptions, we find that atomic models refined with AQuaRef restraints are systematically closer to their high-resolution matches. This indicates that QM-based refined atomic models not only improve standard validation metrics but also provide more realistic representations of the true structures compared to atomic models refined with standard restraints. Expectedly, refining 10 very high-resolution atomic models did not

significantly alter the atomic coordinates but did lead to improved *R*-factors for all ten models (Supplementary Data: Table 5). The most notable differences compared to refinement with standard restraints were observed in the position of hydrogen atoms, specifically those with rotational degrees of freedom (Fig. 9 d-g), where some of these atoms reoriented during refinement to better fit the data and, at the same time, form favorable hydrogen bonds. Another notable difference is the increased r.m.s. deviations from ideal (library) bond and angle values in the case of AQuaRef refinement (Supplementary Data: Table 5), which together with improved hydrogen positions is likely to contribute to improved *R*-factors.

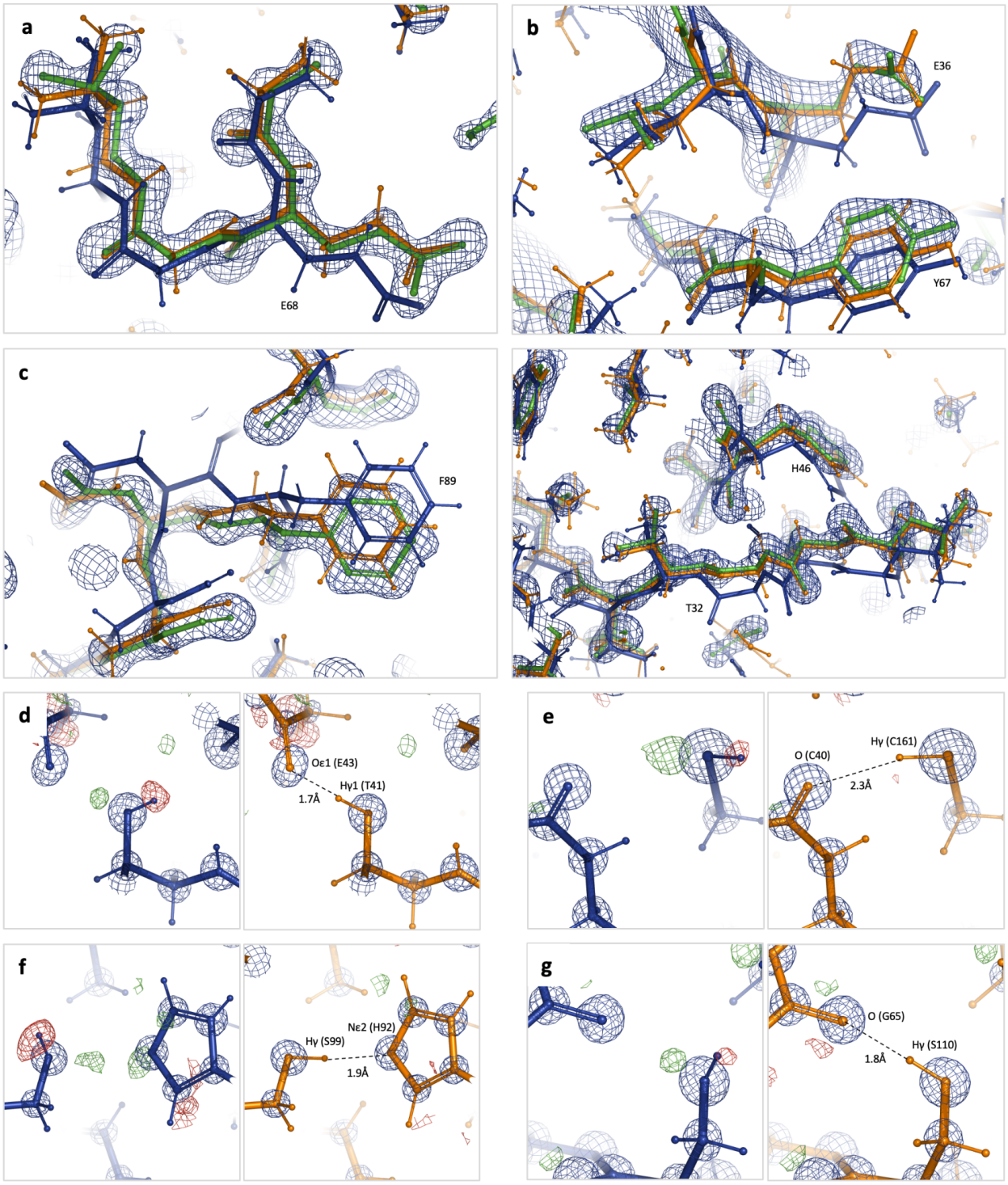
An extended comparison with popular state-of-the-art software packages and refinement methods, including the use of AMBER and Rosetta force fields as refinement restraints, as well as REFMAC5 and Servalcat from the CCP4 software suite, shows that for crystal structure refinement, only Rosetta approaches AQuaRef in terms of the quality of refined atomic model geometries. However, AQuaRef produces slightly improved R_{free} values and significantly better R_{free}-R_{work} gaps, indicating reduced data overfitting. It is also worth noting that Rosetta-based refinement is only available for crystal structures using X-ray data, and refinement times with Rosetta are up to an order of magnitude slower. Although Servalcat achieved superior CC_{mask} values compared to AQuaRef (Fig. 4d), this suggests that Servalcat overfits the map, producing higher CC_{mask} values at the cost of significantly poorer model geometry.

The case study of short hydrogen bonds in human DJ-1 and its bacterial homolog YajL, as well as the protonation states of carboxylic acids involved in these hydrogen bonds, highlights the feasibility of AQuaRef in determining proton positions consistent with experimental evidence across diverse scenarios. This process is fully automated and unbiased by the choice of restraints. Additionally, AIMNet2 energy profiles provide further information about the characteristics of hydrogen bonds and protonation states, which can be used to support specific hypotheses.

The method has been implemented in the quantum refinement software (Q|R), which is built upon the CCTBX library[62] and optionally utilizes tools from Phenix. Q|R is accessible

within Phenix, thereby making these methods readily available to the broader community of structural biologists.

Currently, AQuaRef is trained using commonly known amino acid residues, which means the method can only be applied to protein-only structures. Another main limitation is that, at present, static disorder (alternate conformations) is not handled in Q|R. Removing both limitations is the subject of future work.



405

406 **Figure 9**. **a-c:** Close-up showing models refined with standard restraints (blue) and
407 AQuaRef restraints (orange) superposed onto their higher-resolution homologous models
408 (green) with their corresponding 2mFo-DFc Fourier maps contoured at 2σ; for PDB 5YI5,
409 8R1G, and 6XMX, respectively. **d-g:** Refinement with standard AQuaRef restraints

(orange) overlaid with their corresponding 2mFo-DFc and mFo-DFc Fourier maps, contoured at 5σ (blue) and ±2σ (green, red), respectively (PDB 4O8H). The focus is on hydrogen atoms with rotational degrees of freedom that re-orient during refinement with AQuaRef restraints to satisfy the residual map and participate in hydrogen bonding.

METHODS

AIMNet2 training dataset and AQuaRef model

Since our goal was the parametrization of ML potential for polypeptides, our training dataset needed to cover chemical (amino acid sequence and protonation states), conformational, and intermolecular degrees of freedom. We began by creating a library of small peptides as SMILES strings. We used all 20 standard amino acids, 11 alternate protonation forms, three options for sequence start (ACE, NH3+, NH2), and four options for the end (NME, NHE, CBX, CBA). We enumerated all possible mono- and di-peptides and selected a random subset for tri- and tetra-peptides. Additionally, we generated SMILES for peptides linked by the cysteine-cysteine disulfide bond and their selenium counterparts. Molecular conformations were generated with OpenEye Omega[63] software using dense torsion sampling. No restrictions were applied to the configurations of the chiral centers, ensuring that the dataset and resulting model should work equally well for D-, L-, and mixed stereochemistry peptides. Intermolecular interactions were modeled by generating intermolecular complexes of 2 to 4 peptides with random orientations. No prior knowledge of preferred types of secondary structure for polypeptides was used. To manage the size of the dataset and the training process, we limited the size of peptides and complexes to less than 120 atoms, including hydrogens.

Non-equilibrium conformations of peptides and complexes were sampled with molecular dynamics simulations using the GFN-FF[64] force field. Cartesian restraints were added to keep structures near the input structure, with random torsion and intermolecular degrees of freedom. Molecular configurations for labeling (DFT calculations) and inclusion into the training dataset were selected using Query-By-Committee active learning (AL)

approach[35]. We started with a random selection of 500k samples, used an ensemble of 4 models, and performed a total of 4 iterations of AL adding new samples with high uncertainty of energy and atomic forces prediction. In the final iteration of AL, we performed uncertainty-guided optimization of the structures, minimizing the weighted difference of energy prediction and its uncertainty. This type of active sampling finds structures that balance low predicted forces and high energy uncertainty. The entire procedure resulted in a training dataset containing about one million samples, with a median number of 42 atoms per sample.

DFT calculations were performed with the B97M-D4/def2-QZVP[65-68] method using ORCA 5.0.3 software[69]. Since the Q|R does not use periodic boundary conditions, and usually not all ions and solvent molecules are resolved in the refinement, we used implicit treatment of solvent effects with CPCM[70] method using parameters for water as solvent.

The core architecture of the AQuaRef model matches the base AIMNet2 model[33], with few modifications. First, we did not use explicit long-range Coulomb and dispersion interactions, we trained to total DFT-D4 energy instead. With CPCM treatment, the Coulomb term could not be estimated using interactions between partial atomic charges, and also long-range interactions are effectively screened with a polarizable continuum. Long range dispersion interactions beyond the local cutoff of 5 Å have little effect on atomic forces, which are important in Q|R refinement. We also added explicit short-range exponential repulsion terms to make the potential more robust for the structures with clashes. The model was trained to reproduce DFT-D4 energies, forces, and Hirshfeld partial atomic charges.

Experimental data and atomic models for test cases

Protein-only, single-conformation high-to-low resolution X-ray crystallography and Cryo-EM models, along with their corresponding experimental datasets, were selected from RCSB and EMDB based on multiple criteria. These criteria include model size (between 1,000 and 10,000 non-hydrogen atoms), resolution (between 2.5 and 4 Å), geometric model quality (MolProbity clashscore better than 50, with no covalent bonds deviating by

more than 4 r.m.s.d. from ideal library values), goodness of fit between the model and the experimental data (Cryo-EM: CC_{mask} > 0.6, X-ray: R_{work} < 0.3), and the availability of a higher-resolution (better than 2 Å) homologous model (main chain superposition r.m.s.d. < 1 Å, sequence identity greater than 95%) for each considered model. Additionally, 11 ultra-high resolution single-conformation X-ray models were selected that contained only protein and ordered water atoms.

Comparison of models

All atoms were used to calculate coordinate r.m.s. deviations between models before and after refinement, as shown in Figure 3a. Coordinate r.m.s. deviations between models used for test refinements and their high-resolution homologues were calculated using the Phenix tool phenix.superpose_pdbs, which included all non-hydrogen backbone atoms plus Cβ and Cγ atoms where present. R.m.s. deviations in torsion angle space were calculated using CCTBX[62], with matching torsion angles selected as described by Headd et al.[15].

Atomic model preparation for refinement

Model preparation for refinement (e.g., adding any missing atoms) was done using qr.finalise program of Q|R, which uses the Reduce program[71] to add hydrogen atoms at geometrically predicted positions. Model geometry regularization was done using the Phenix tool phenix.geometry_minimization.

Model refinement

The exact same input models were used for all trial refinements. Real-space refinement in Phenix was performed using the phenix.real_space_refine program[12]. Four refinement runs were performed independently, starting with the same input maps (cryo-EM) or reflection data (X-ray) and models. The runs included: 1) standard restraints consisting

of restraints on bond lengths, bond angles, torsion angles, planes, chirality, and non-bonded repulsion; 2) standard restraints with the addition of secondary-structure restraints; 3) standard restraints with the addition of Ramachandran plot restraints; and 4) standard restraints with the addition of secondary-structure and Ramachandran plot restraints.

Quantum-based real- and reciprocal-space refinement was performed using the qr.refine program of Q|R, using all default settings except for the source of QM restraints (AQuaRef).

Reciprocal-space refinement in Phenix was performed using phenix.refine[72] with the exact same four choices of restraints as in real-space refinement.

Software and availability

Phenix software is available at: phenix-online.org. Quantum refinement (Q|R) software is available at qrefine.com. AQuaRef refinement is available in Phenix starting dev-5395 version. CCTBX-based Python scripts and the data (atomic models, cryo-EM maps, X-ray diffraction data) used in this study are available at: https://phenix-online.org/phenix_data/afonine/qr_aimnet2_2024/. Refinement parameters are documented in README files, as well as in the Python scripts used to run the refinements. Input data for deposited models were obtained from the Protein Data Bank[73] and Electron Microscopy Data Bank[74], either by using the Phenix tool phenix.fetch_pdb or from the CERES server[75].

Graphics software

Map and model images were prepared using PyMOL[76]. Routine inspection of maps and models was performed using Coot[77]. Plots were generated using Matplotlib[78].

Acknowledgements

PVA, NWM, and PDA acknowledge funding from the National Institutes of Health (grants R01GM071939, P01GM063210, and R24GM141254), as well as support from the Phenix Industrial Consortium and the US Department of Energy under Contract No. DE-AC02-05CH11231. OI acknowledges support from the US National Science Foundation (NSF CHE-2154447). AER acknowledges support from the US National Science Foundation (NSF CHE-1802831 and OAC-2311632). MB acknowledges support from the COST Action CA21101 “Confined molecular systems: from a new generation of materials to the stars’ (COSY)” supported by COST (European Cooperation in Science and Technology) and computer resources provided by Wroclaw Centre for Networking and Supercomputing (http://wcss.pl).

This work used Expanse at SDSC and Bridges-2 at PSC through allocation CHE200122 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by NSF grants #2138259, #2138286, #2138307, #2137603, and #2138296. This research is part of the Frontera computing project at the Texas Advanced Computing Center. Frontera is made possible by the National Science Foundation award OAC-1818253. This research, in part, was done using resources provided by the Open Science Grid, which is supported by the award 1148698 and the U.S. DOE Office of Science.

Author contributions

Conceptualization: P.V.A., A.E.R., O.I., M.B.; Methodology: R.Z., H.G., K.R.; Software: P.V.A., N.W.M., R.Z., M.P.W, H.K.; Validation: M.B., K.R., H.G., P.D.A.; Formal analysis: M.B.; Data Curation: R.Z., H.G., K.R., P.V.A.; Writing (original draft): P.V.A., M.B., R.Z.; Writing (review & editing): all authors; Visualization: P.V.A., M.B., H.G., R.Z.; Supervision, P.V.A., A.E.R., O.I.; Project administration: P.V.A., A.E.R., O.I., M.B.; Funding Acquisition: P.D.A., P.V.A., A.E.R., O.I.

Competing Interests

The authors declare no competing interests.

References

1. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
2. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
3. Krishna, R. *et al.* Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528 (2024).
4. Terwilliger, T. C. *et al.* AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat. Methods* **21**, 110–116 (2024).
5. Edich, M., Briggs, D. C., Kippes, O., Gao, Y. & Thorn, A. The impact of AlphaFold2 on experimental structure solution. *Faraday Discuss.* **240**, 184–195 (2022).
6. Urzhumtsev, A. G. & Lunin, V. Y. Introduction to crystallographic refinement of macromolecular atomic models. *Crystallogr. Rev.* **25**, 164–262 (2019).
7. Evans, P. R. An introduction to stereochemical restraints. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 58–61 (2007).
8. Vagin, A. A. *et al. REFMAC* 5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184–2195 (2004).

9. Engh, R. A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47**, 392–400 (1991).
10. Agirre, J. *et al.* The *CCP* 4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr. Sect. Struct. Biol.* **79**, 449–461 (2023).
11. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr. Sect. Struct. Biol.* **75**, 861–877 (2019).
12. Afonine, P. V. *et al.* Real-space refinement in *PHENIX* for cryo-EM and crystallography. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 531–544 (2018).
13. Kovalevskiy, O., Nicholls, R. A. & Murshudov, G. N. Automated refinement of macromolecular structures at low resolution using prior information. *Acta Crystallogr. Sect. Struct. Biol.* **72**, 1149–1161 (2016).
14. DeLaBarre, B. & Brunger, A. T. Considerations for the refinement of low-resolution crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 923–932 (2006).
15. Headd, J. J. *et al.* Use of knowledge-based restraints in *phenix.refine* to improve macromolecular refinement at low resolution. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 381–390 (2012).
16. Sobolev, O. V., Afonine, P. V., Adams, P. D. & Urzhumtsev, A. Programming new geometry restraints: parallelity of atomic groups. *J. Appl. Crystallogr.* **48**, 1130–1141 (2015).
17. De Vries, I. *et al.* New restraints and validation approaches for nucleic acid structures in *PDB-REDO*. *Acta Crystallogr. Sect. Struct. Biol.* **77**, 1127–1141 (2021).
18. Afonine, P. V., Sobolev, O. V., Moriarty, N. W., Terwilliger, T. C. & Adams, P. D.

595 Overall protein structure quality assessment using hydrogen-bonding parameters.
596 *Acta Crystallogr. Sect. Struct. Biol.* **79**, 684–693 (2023).
597 19. Moriarty, N. W., Liebschner, D., Tronrud, D. E. & Adams, P. D. Arginine off-kilter:
598 guanidinium is not as planar as restraints denote. *Acta Crystallogr. Sect. Struct. Biol.*
599 **76**, 1159–1166 (2020).
600 20. Richardson, J. S. *et al.* Model validation: local diagnosis, correction and when to
601 quit. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 132–142 (2018).
602 21. Jiang, Z., Biczysko, M. & Moriarty, N. W. Accurate geometries for “Mountain pass”
603 regions of the Ramachandran plot using quantum chemical calculations. *Proteins*
604 *Struct. Funct. Bioinforma.* **86**, 273–278 (2018).
605 22. Moriarty, N. W. *et al.* Improved chemistry restraints for crystallographic refinement
606 by integrating the Amber force field into *Phenix*. *Acta Crystallogr. Sect. Struct. Biol.*
607 **76**, 51–62 (2020).
608 23. Zheng, M., Reimers, J. R., Waller, M. P. & Afonine, P. V. *Q* | *R*: quantum-based
609 refinement. *Acta Crystallogr. Sect. Struct. Biol.* **73**, 45–52 (2017).
610 24. Bergmann, J., Oksanen, E. & Ryde, U. Combining crystallography with quantum
611 mechanics. *Curr. Opin. Struct. Biol.* **72**, 18–26 (2022).
612 25. Zheng, M. *et al.* Including crystallographic symmetry in quantum-based refinement:
613 *Q* | *R* #2. *Acta Crystallogr. Sect. Struct. Biol.* **76**, 41–50 (2020).
614 26. Zheng, M. *et al.* Solving the scalability issue in quantum-based refinement: Q|R#1.
615 *Acta Crystallogr. Sect. Struct. Biol.* **73**, 1020–1028 (2017).
616 27. Wang, L. *et al.* Real-space quantum-based refinement for cryo-EM: *Q* | *R* #3. *Acta*
617 *Crystallogr. Sect. Struct. Biol.* **76**, 1184–1191 (2020).

618 28. Borbulevych, O. Y., Plumley, J. A., Martin, R. I., Merz, K. M. & Westerhoff, L. M.
619 Accurate macromolecular crystallographic refinement: incorporation of the linear
620 scaling, semiempirical quantum-mechanics program *DivCon* into the *PHENIX*
621 refinement package. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1233–1247 (2014).
622 29. Ryde, U. Combined quantum and molecular mechanics calculations on
623 metalloproteins. *Curr. Opin. Chem. Biol.* **7**, 136–142 (2003).
624 30. Canfield, P., Dahlbom, M. G., Hush, N. S. & Reimers, J. R. Density-functional
625 geometry optimization of the 150 000-atom photosystem-I trimer. *J. Chem. Phys.*
626 **124**, 024301 (2006).
627 31. Kulik, H. J., Luehr, N., Ufimtsev, I. S. & Martinez, T. J. Ab Initio Quantum Chemistry
628 for Protein Structures. *J. Phys. Chem. B* **116**, 12501–12509 (2012).
629 32. Zheng, M. & Waller, M. P. Adaptive quantum mechanics/molecular mechanics
630 methods. *WIREs Comput. Mol. Sci.* **6**, 369–385 (2016).
631 33. Anstine, D., Zubatyuk, R. & Isayev, O. AIMNet2: A Neural Network Potential to Meet
632 your Neutral, Charged, Organic, and Elemental-Organic Needs. Preprint at
633 https://doi.org/10.26434/chemrxiv-2023-296ch-v2 (2024).
634 34. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network
635 potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–
636 3203 (2017).
637 35. Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less is more:
638 Sampling chemical space with active learning. *J. Chem. Phys.* **148**, 241733 (2018).
639 36. Devereux, C. *et al.* Extending the Applicability of the ANI Deep Learning Molecular
640 Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **16**, 4192–4202 (2020).

641 37. Zubatyuk, R., Smith, J. S., Nebgen, B. T., Tretiak, S. & Isayev, O. Teaching a neural
642 network to attach and detach electrons from molecules. *Nat. Commun.* **12**, 4870
643 (2021).
644 38. Yan, Z., Wei, D., Li, X. & Chung, L. W. Accelerating reliable multiscale quantum
645 refinement of protein–drug systems enabled by machine learning. *Nat. Commun.*
646 **15**, 4181 (2024).
647 39. Brünger, A. T., Karplus, M. & Petsko, G. A. Crystallographic refinement by simulated
648 annealing: application to crambin. *Acta Crystallogr. A* **45**, 50–61 (1989).
649 40. Wang, Y. *et al.* Optimal clustering for quantum refinement of biomolecular
650 structures: Q|R#4. *Theor. Chem. Acc.* **142**, 100 (2023).
651 41. Moriarty, N. W., Tronrud, D. E., Adams, P. D. & Karplus, P. A. Conformation-
652 dependent backbone geometry restraints set a new standard for protein
653 crystallographic refinement. *FEBS J.* **281**, 4061–4071 (2014).
654 42. Engh, R. A. & Huber, R. Structure quality and target parameters. in 474–484 (2012).
655 doi:10.1107/97809553602060000857.
656 43. Afonine, P. V. *et al.* New tools for the analysis and validation of cryo-EM maps and
657 atomic models. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 814–840 (2018).
658 44. Booth, A. D. LXXIV. An expression for following the process of refinement in X-ray
659 structure analysis using fourier series. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **36**,
660 609–615 (1945).
661 45. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy
662 of crystal structures. *Nature* **355**, 472–475 (1992).
663 46. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-

664 atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

665 47. Lovell, S. C. *et al.* Structure validation by Cα geometry: ϕ,ψ and Cβ deviation.
666 *Proteins Struct. Funct. Bioinforma.* **50**, 437–450 (2003).

667 48. Hintze, B. J., Lewis, S. M., Richardson, J. S. & Richardson, D. C. Molprobity's
668 ultimate rotamer-library distributions for model validation: MolProbity's Ultimate
669 Rotamer-Library. *Proteins Struct. Funct. Bioinforma.* **84**, 1177–1189 (2016).

670 49. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. The penultimate
671 rotamer library. *Proteins Struct. Funct. Genet.* **40**, 389–408 (2000).

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

674 51. Chen, V. B. *et al. MolProbity* : all-atom structure validation for macromolecular
675 crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

676 52. Sobolev, O. V. *et al.* A Global Ramachandran Score Identifies Protein Structures
677 with Unlikely Stereochemistry. *Structure* **28**, 1249-1258.e2 (2020).

678 53. Tickle, I. J., Laskowski, R. A. & Moss, D. S. R_{free} and the R_{free} Ratio. I. Derivation of
679 Expected Values of Cross-Validation Residuals Used in Macromolecular Least-
680 Squares Refinement. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 547–557 (1998).

681 54. Tickle, I. J., Laskowski, R. A. & Moss, D. S. R_{free} and the R_{free} ratio. II. Calculation
682 of the expected values and variances of cross-validation statistics in
683 macromolecular least-squares refinement. *Acta Crystallogr. D Biol. Crystallogr.* **56**,
684 442–450 (2000).

685 55. Barad, B. A. *et al.* EMRinger: side chain–directed model and map validation for 3D
686 cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).

687

688 56. Moriarty, N. W. *et al.* Improved chemistry restraints for crystallographic refinement by
689 integrating the Amber force field into Phenix. *Acta Cryst D* **76**, 51–62 (2020).
690 57. DiMaio, F. *et al.* Improved low-resolution crystallographic refinement with Phenix and
691 Rosetta. *Nat Methods* **10**, 1102–1104 (2013).
692 58. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal
693 structures. *Acta Cryst D* **67**, 355–367 (2011).
694 59. Yamashita, K., Palmer, C. M., Burnley, T. & Murshudov, G. N. Cryo-EM single-particle
695 structure refinement and map calculation using Servalcat. *Acta Cryst D* **77**, 1282–1291
696 (2021).
697 60. Lin, J., Pozharski, E. & Wilson, M. A. Short Carboxylic Acid–Carboxylate Hydrogen
698 Bonds Can Have Fully Localized Protons. *Biochemistry* **56**, 391–402 (2017).
699 61. Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Cryst C* **71**, 3–8
700 (2015).
701 62. Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W. & Adams, P. D. The
702 *Computational Crystallography Toolbox* : crystallographic algorithms in a reusable
703 software framework. *J. Appl. Crystallogr.* **35**, 126–136 (2002).
704 63. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T.
705 Conformer Generation with OMEGA: Algorithm and Validation Using High Quality
706 Structures from the Protein Databank and Cambridge Structural Database. *J. Chem.*
707 *Inf. Model.* **50**, 572–584 (2010).
708 64. Spicher, S. & Grimme, S. Robust Atomistic Modeling of Materials, Organometallic,
709 and Biochemical Systems. *Angew. Chem. Int. Ed.* **59**, 15665–15673 (2020).

65. Mardirossian, N. & Head-Gordon, M. Mapping the genome of meta-generalized gradient approximation density functionals: The search for B97M-V. *J. Chem. Phys.* **142**, 074111 (2015).

66. Caldeweyher, E. *et al.* A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **150**, 154122 (2019).

67. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).

68. Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **8**, 1057 (2006).

69. Neese, F., Wennmohs, F., Becker, U. & Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **152**, 224108 (2020).

70. Barone, V. & Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **102**, 1995–2001 (1998).

71. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation 1 1Edited by J. Thornton. *J. Mol. Biol.* **285**, 1735–1747 (1999).

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

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

74. The wwPDB Consortium *et al.* EMDB—the Electron Microscopy Data Bank. *Nucleic Acids Res.* **52**, D456–D465 (2024).

75. Liebschner, D. *et al. CERES* : a cryo-EM re-refinement system for continuous improvement of deposited models. *Acta Crystallogr. Sect. Struct. Biol.* **77**, 48–61 (2021).

76. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).

77. Emsley, P. & Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

78. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

SUPPLEMENTAL METHODS

Case study: Short hydrogen bonds in human DJ-1 and its bacterial homologue YajL

Atomic models and experimental data for human DJ-1 and the bacterial homologue YajL were downloaded from the Protein Data Bank (PDB) using accession codes 5SY6 and 5SY4, respectively. Reflection intensities were used in all calculations. Both atomic models were protonated using the Reduce program from Phenix suite, as the deposited models lacked hydrogen atoms.

To address uncertainty in refined atomic model parameters due to intrinsic refinement properties, particularly in the coordinates, interatomic distances are reported as averages with standard deviations, based on 100 refinement runs. Each refinement run (restrained and unrestrained with phenix.refine as well as AQuaRef) used identical settings but began with models where coordinates were randomized with an r.m.s.d. of 0.1 Å.

To avoid biases in assumptions about the protonation states of E/D and D/T residues, these residues were modeled as unprotonated with idealized geometries in the input refinement models. For AQuaRef refinements, the proton was positioned exactly between the Oε2 and Oδ2 atoms for DJ-1 (Fig. 5b) and YajL, as well as between the Oδ2 and Oγ1

759 atoms for YajL. In phenix.refine, this proton was not explicitly modeled to avoid assuming
760 its association with a particular residue.

761 Restrained phenix.refine refinements were performed with 10 macro-cycles to ensure
762 convergence. Unrestrained refinements used the same settings, except geometric
763 restraints affecting hydrogen bond interaction moieties in E/D and D/T residues were
764 eliminated. Excluded restraints were covalent bond and angle restraints for COO moieties
765 and repulsion restraints between Oε2, Oδ2, and Oγ1 atoms. AQuaRef refinements were
766 performed with all default settings.

767 The AIMNet2 energy profiles between corresponding oxygen atoms were calculated by
768 sampling the proton positions along the line connecting the oxygens at 100 equally
769 spaced points.

770 The sigma-A weighted Fo-Fc difference map value profiles were interpolated along the
771 Oδ2-H, Oε2-H, and Oγ1-H lines using tri-cubic interpolation and averaged across 100
772 difference maps from corresponding refinements. The mean bulk-solvent density value
773 was estimated using the flat bulk-solvent model, as implemented in Phenix.

774 All Python scripts using Q|R and CCTBX libraries to perform all the above calculations
775 are available in Supplemental Data.

776