

UC Berkeley

UC Berkeley Previously Published Works

Title

Macromolecular refinement of X-ray and cryoelectron microscopy structures with Phenix/OPLS3e for improved structure and ligand quality.

Permalink

<https://escholarship.org/uc/item/2v21391b>

Journal

Structure (London, England : 1993), 29(8)

ISSN

0969-2126

Authors

van Zundert, Gydo CP
Moriarty, Nigel W
Sobolev, Oleg V
[et al.](#)

Publication Date

2021-08-01

DOI

10.1016/j.str.2021.03.011

Peer reviewed



HHS Public Access

Author manuscript

Structure. Author manuscript; available in PMC 2022 August 05.

Published in final edited form as:

Structure. 2021 August 05; 29(8): 913–921.e4. doi:10.1016/j.str.2021.03.011.

Macromolecular refinement of X-ray and cryo-electron microscopy structures with Phenix / OPLS3e for improved structure and ligand quality

Gydo C.P. van Zundert¹, Nigel W. Moriarty², Oleg V. Sobolev², Paul D. Adams^{2,3}, Kenneth W. Borrelli^{1,4}

¹Schrödinger, New York, NY 10036, USA

²Molecular Biosciences and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

³Department of Bioengineering, University of California at Berkeley, Berkeley, CA 94720, USA

⁴Lead Contact

Summary

With the advent of the resolution revolution in cryo-electron microscopy, low resolution refinement is common, and likewise increases the need for a reliable force field. Here we report on the incorporation of the OPLS3e force field with the VSGB2.1 solvation model in the structure determination package Phenix. Our results show significantly improved structure quality and reduced ligand strain at lower resolution for X-ray refinement. For refinement of cryo-EM based structures, we find comparable quality structures, goodness-of-fit and reduced ligand strain. We also show how structure quality and ligand strain is related to the map-model cross correlation as a function of data weight, and how it can detect overfitting. Signs of overfitting are found in over half of our cryo-EM dataset, which can be remedied by a re-refinement at a lower data weight. Finally, a start-to-end script for refining structures with Phenix/OPLS3e is available in the Schrödinger 2020–3 distribution.

Graphical Abstract

Correspondence: kenneth.borrelli@schrodinger.com.

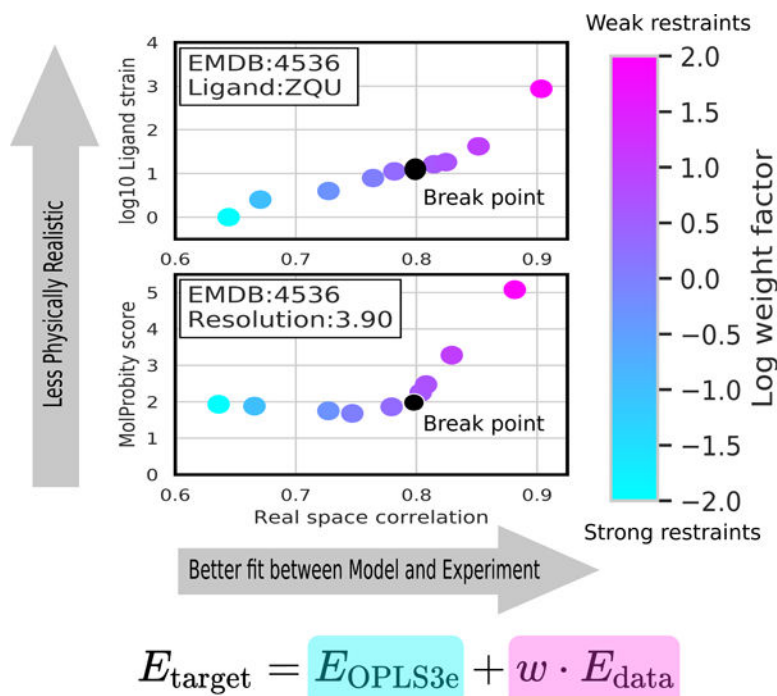
Author Contributions

Conceptualization, G.C.P.v.Z and K.W.B.; Methodology, G.C.P.v.Z., N.W.M., O.V.S. and K.W.B.; Software, G.C.P.v.Z., N.W.M., O.V.S. and K.W.B.; Validation, G.C.P.v.Z.; Formal Analysis, G.C.P.v.Z.; Investigation, G.C.P.v.Z.; Resources, G.C.P.v.Z. and K.W.B.; Writing - Original Draft, G.C.P.v.Z. and K.W.B.; Writing - Review & Editing, G.C.P.v.Z., N.W.M., O.V.S., P.D.A., and K.W.B.; Funding Acquisition, K.W.B. and P.D.A.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of Interests

GCPvZ and KWB have financial holdings in Schrödinger.



e-TOC blurb

Van Zundert et al. have implemented the OPLS3e forcefield and VSGB2.1 solvation model into the structure determination package Phenix, resulting in improved structure and ligand quality for refinement at lower resolution. Signs of overfitting were found in over half of cryo-EM cases investigated and a straightforward remedy is proposed.

Introduction

X-ray crystallography has been the main source of structural information accounting for about 90% of all entries in the PDB and has been the backbone of structural biology for the last decades. With the advent of the “resolution revolution” due to improved hardware and software, cryo-electron microscopy (cryo-EM) has become a viable approach for investigating larger molecular complexes and membrane proteins, potentially vastly expanding new drug targets amenable to structure based drug design (SBDD). Many SBDD methods require atomic accuracy for at least all of the heavy atoms. Unfortunately, the models fit to many X-ray crystallography datasets and virtually all cryo-EM datasets are hampered by low resolution, and contain some ambiguities at this scale.

Due to the low data to parameter ratio found in macromolecular refinement, more so at lower resolutions, knowledge based restraints are introduced to reduce overfitting, the most simple form being bond lengths and angles. At lower resolution, additional restraints can be included, such as secondary structure (Head et al. 2012), and deformable elastic network restraints (Schröder et al., 2014). The former is currently routinely used during cryo-EM structure refinement and low-resolution X-ray crystallography. Accurate restraints are thus paramount for obtaining physically realistic structures and ligands when data is relatively

sparse. The problem is exacerbated in cryo-EM as no robust validation metric is generally agreed upon to detect overfitting, even though there is considerable concern and several methods have been proposed (DiMaio et al., 2013; Falkner and Schröder, 2013; Lagerstedt et al., 2020; Volkman, 2009). This is in contrast to crystallographic refinement where the concept of R_{free} is universally used (Brunger, 1992), though the precise practicalities are not on a firm theoretical footing (Tickle et al., 1998).

Knowledge based structural restraints for the standard amino-acid residues are well established from the wealth of high-resolution data that is available in the Protein DataBank (wwPDB consortium, 2019) and small molecule databases (Groom et al., 2016). The Engh and Huber equilibrium bond lengths and angles (Engh and Huber, 1991, 2001) are used for several decades, but have been replaced by conformation dependent restraint libraries for the backbone atoms in Phenix (Moriarty et al., 2016; Tronrud et al., 2010). Generating restraints for small molecules, however, is significantly harder, due to the vastly increased chemical space that they can occupy and limited available data; in conjunction with the observation that ligand densities are typically less resolved compared to their surroundings, either due to a superposition of states, partial occupancy, increased mobility, or any combination of these, this has led to a number of critical publications addressing the quality of ligands deposited in the PDB (Deller and Rupp, 2015; Liebeschuetz et al., 2012; Peach et al., 2017; Reynolds, 2014, Sitzmann et al., 2012). Graphical interface tools have been developed to help assess and validate the ligand and its binding pocket (Cereto-Massagué et al., 2013). Moreover, the wwPDB formed a working group to define proper ligand validation protocols to combat these issues, with several recommendations already being implemented (Adams et al., 2016).

Knowledge based restraints for ligands are usually provided through restraint dictionaries in CIF format, derived from high-resolution structure models extracted from small molecule crystallography databases, or quantum mechanical calculations, with several computer programs having been developed to automate the process (Steiner and Tucker, 2017). Large libraries containing restraints for monomers present in the PDB such as the REFMAC5 monomer library (Vagin et al., 2004) are extensively used in REFMAC (Kovalevskiy et al., 2018) and Phenix (Liebschner et al., 2019). In addition, the latter uses the GeoStd library (<http://sourceforge.net/projects/geostd>) that is curated for accurate amino acid restraints. The restraints comprise the above mentioned bond length and angle restraints, but also include torsion angle, chirality, and planarity restraints, usually combined with a repulsive interaction energy term to prevent serious atomic clashes. However, for more complicated molecules such as macrocyclic peptides, generating an accurate restraint dictionary can still be a tedious, error-prone, and time consuming process. Furthermore, electrostatics and attractive Van der Waals forces are typically not taken into account, though it has been shown that electrostatics is important for modeling the hydration of DNA (Fenn et al., 2011), while including an implicit solvent improves stereochemistry for solvent accessible residues (Moulinier, 2003). Moreover, since cryo-EM results in Coulomb potential maps, the inclusion of electrostatics during refinement will ultimately be important to accurately represent the experimental data (Wang and Moore, 2017).

More sophisticated approaches introduce a physics-based force field or low-level quantum mechanics Hamiltonian to calculate energies and gradients during structure refinement. The use of a force field in refinement has already been introduced several decades ago in the well-known Xplor and CNS programs using the OPLS-AA parameters (Brunger et al., 1998). Also, the Schrödinger developed PrimeX refinement protocol uses the Prime force field, a combination of the OPLS3 and VSGB2.0 energy model, and we have shown that it results in significantly improved structures based on standard quality measures such as the number of clashes (Bell et al., 2012; Li et al., 2011). The Amber force field (Ponder and Case, 2003) is available in Phenix and a recent analysis likewise showed that structure quality is improved, more so at lower resolutions (Moriarty et al., 2020). The QR package applies QM to the entire protein (reciprocal or real space) improving geometries and hydrogen bonding networks (Wang et al. 2020; Zheng et al., 2017, 2020). A number of publications focused particularly on ligand geometries in reciprocal space refinement and are mainly effective in relieving ligand strain: Phenix-AFITT supplements the Phenix energy model with the MMFF-94 force field for ligands only (Janowski et al., 2009); Phenix-DIVCON takes a different approach by applying the AM1 quantum mechanical Hamiltonian (Borbulevych et al., 2014, 2016).

The recently released OPLS3e force field combines accurate fixed partial charges from on-the-fly semi-empirical quantum mechanics calculations with accurate torsional profiles from a combination of a database covering a substantial portion of medicinal chemistry space and the ability to easily extend that database with high-level quantum mechanics calculations when encountering functional groups not already covered (Roos et al., 2019). To energetically account for the macromolecule - bulk solvent interaction, we have previously developed the VSGB energy model for accurate protein structure modeling of loops (Li et al., 2011). The VSGB model combines an optimized GBSA-based implicit solvent model with additional corrections for hydrogen bonding, π - π , self-contact, and hydrophobic interactions. Here we introduce the implementation of an interface between the principal refinement programs in the widely used macromolecular structure determination package Phenix and the OPLS3e force field and VSGB2.1 implicit solvation model through Schrödinger's Prime software, which we refer to as Phenix/OPLS3e. The implementation is versatile supporting covalently bound ligands and multiconformer complexes, and can be used in reciprocal and real space refinement for either the whole or part of the structure, using `phenix.refine` (Afonine et al., 2012) and `phenix.real_space_refine` (Afonine et al., 2018a). It furthermore removes the need for accurate restraints files as ligands are automatically parameterized by the OPLS3e force field, either through its internal database or generated by the ForceFieldBuilder available in Maestro (www.schrodinger.com), the main graphical user interface for the Schrödinger software suite. We benchmarked our approach on 2284 cases using X-ray data and show that refinement with Phenix/OPLS3e results in structures with statistically significant improved Molprobit scores and Ramachandran Z-scores (Sobolev et al., 2020) at resolutions better than 1.5Å, and reduced whole structure and ligand strain energies across all resolutions. Larger improvements are seen at lower resolutions, with models showing a small increase in their R_{free} values, that are only statistically significant at resolutions between 2.0 and 3.0Å. Interestingly, when we applied our method to 15 cryo-EM structural models, this resulted in models with similar

MolProbity scores and real space cross correlation values but reduced ligand strain. We furthermore show that the inclusion of a high quality ligand force field provides an additional measure to reduce overfitting by constructing a ligand strain versus cross correlation curve by sampling over weight space through multiple refinements. Our analysis highlights some instances where structure quality and goodness-of-fit could be improved in current deposited PDB structures, while signs of overfitting are present in 8 of the 15 cases based on the MolProbity score and Ramachandran Z-score. These cases can be straightforwardly remedied by re-refinement at a lower data weight factor.

The Phenix/OPLS3e implementation has already seen applications in cryo-EM enabled SBDD of macrocycles in ribosomes (Qi et al., 2019) and determining high-confidence ligand binding poses in ambiguous cryo-EM structures in the GemSpot pipeline (Robertson et al., 2020). We foresee our approach as an accessible and user-friendly implementation of a high quality force field to improve structure and ligand quality especially when using lower resolution data provided either by crystallography and cryo-EM, and as an additional measure for reducing overfitting in cryo-EM structure refinement and cryo-EM ligand fitting. () *d r r d r c ff other ff other ff other*

Results and discussion

Structure quality improvement increases with decreasing resolution

To measure the impact of the OPLS3e/VSGB force field on refinement, we applied Phenix/OPLS3e and standard Phenix reciprocal space refinement on 2284 protein ligand complexes at varying X-ray diffraction resolutions as described above. For each of the resolution bins we calculated the difference in R-factors, and MolProbity score between the two refinement protocols where differences are calculated by subtracting standard Phenix refined structure metrics from Phenix/OPLS3e refined structure metrics (Figure 1a–b). The absolute values for the R-factors, individual MolProbity metrics, the bond length and bond angle RMSZs, and Ramachandran Z-scores for both protocols can be found in Figure S1 and Data S1.

In general there is a slight increase in the median R_{free} value, ranging from 0.003 (0.3%) in the 1.0–1.5Å resolution bin to 0.010 at the 3.5–4.0Å range for models re-refined starting from the initial PDB structure. For R_{gap} we observe a different behavior, with increased values of 0.006 at high resolution and decreased values for lower resolution refinement starting at 2.0Å. However, the difference in the mean is statistically only significant within the 2.0 – 3.0Å range for R_{free} and the 1.5 – 3.0Å for R_{gap} ($p < 0.05$). The MolProbity score increases by 0.21 for high resolution refinements (1.0–1.5Å), but improves at lower resolutions with a small decrease of 0.09 at 1.5–2.0Å and 0.32–0.66 improvement at lower resolutions. As the MolProbity score is a combination of several factors, we also inspected other geometric metrics (see Figure S1). All terms except for C-beta deviations have improved median values for structures at resolutions lower than 1.5Å, including the Ramachandran Z-score, where the median ranges from 0.0 down to –2.0, within the bounds that correspond to high quality structures. Unsurprisingly, the structure energy is markedly reduced after refinement with inclusion of the OPLS3e energies indicating that the improvement in the structure quality is indeed due to the force field (Figure 1c), with the mode found at -1.5×10^3 kcal/mol. As observed previously, the bond length and bond angle

RMSD values increase by 0.09Å and 1.3–1.5°, with absolute RMSD values of 0.1Å and 2.3°, which is well in-line with observed values in high-resolution structures (Bell et al., 2012; Jaskolski et al. 2007). However, a better comparison is provided by the RMSZ scores, which increases by 0.5 – 0.8 and 0.6 – 1.0 for the bond length and bond angle, but globally stays below 1.4 for both terms after Phenix/OPLS3e refinement. The only structure metric that suffers is the number of C-beta outliers, which is typically 0 or 1 in standard Phenix refinements, due to tight restraints. Phenix/OPLS3e typically results in an additional 1–3 C-beta outliers depending on resolution. All structure based metric differences are statistically significant ($p < 0.05$) at resolutions worse than 1.5Å, apart from the Ramachandran Z-score which is at the border of significance in the 1.5 – 2.0Å resolution bin ($p = 0.05$).

The improvement of structure quality in terms of the clash score has already been observed in our previous work describing PrimeX, with similar bond length and bond angle RMSDs found compared to ideality (Bell et al., 2012). Furthermore, during the development of this work and manuscript, Moriarty et al. (2020) reported on the inclusion of the Amber force field in Phenix reciprocal space refinement. The results show similar trends both in R-factors and MolProbity statistics showing increased R_{work} and R_{free} values, with a reduction in the MolProbity score and an increase in the C-beta outliers.

To properly compare the output structures of Phenix/OPLS3e and standard Phenix, it is important to take into account the weight optimization protocol used in phenix.refine. The weight optimization scheme is essentially a grid search where several weights are sampled, and the weight is chosen in such a way that certain structure based metrics fall within set cutoffs and within those bounds a specific metric is minimized. For example, for structures with a resolution better than 1.5Å, cutoffs are used for the bond length and angle RMSDs, and the weight resulting in the lowest R_{free} is chosen; this is in contrast to refinement at lower resolution such as between 2.5 and 3.5Å where besides the bond length and angle RMSDs also an acceptable R_{gap} and R_{free} range is compared to the lowest values found among all weights, and the structure containing the lowest clash score is chosen. The acceptable R_{gap} and R_{free} ranges are set to 6% and 1.5% points respectively. Taking these optimization parameters into account, we note that most Phenix/OPLS3e structures fall within the 1.5% R_{free} cutoff (Figure 1a), i.e. the median falls well below 1.5%, while nearly all exhibit an improved Clashscore (see Figure S1), where the 95% percentile whisker is well below zero. Thus, when the default optimization scheme would be provided with both the Phenix/OPLS3e re-refined structure and the standard Phenix structure, it would more often pick the Phenix/OPLS3e structure. The same arguments would hold for the other resolution bins starting from 2.0Å onward. In effect, this could also explain the behaviour between these resolution bins in Figure 1a where the R_{work} and R_{free} range is markedly increased compared to refinement at higher resolution. Obviously, changing the weight selection parameters might change the preferred force field used, as the standard Phenix restraints model seems better at finding lower R-factors, especially at higher resolution.

Next we inspected ligand quality by calculating the reduction in strain energy, i.e. energy before and after refinement, and their real space cross correlation for all 4928 ligand cases. Although the RMSDs between the starting and end conformation is typically smaller than 1Å (Figure 2a), and real space cross correlations remain about the same with a median

difference of -0.001 (Figure 2b), in 99% of the cases the strain was markedly reduced (Figure 2c).

In general, our results are in agreement with previous studies for both full structure refinement as well as ligand focused refinement protocols, showing that the impact of including the force field is mainly improving the detail of the models, inherent to the Phenix global minimization strategy employed during refinement (`refine.strategy=individual_sites+individual_adp+occupancies`) which applies a gradient based local minimization, a point that was adequately made recently by Moriarty et al. (2020).

Finally, we investigated the computational performance of the Phenix/OPLS3e refinement against standard Phenix refinement. In general, a Phenix/OPLS3e refinement takes about an order of magnitude more time (Figure S2), where the bulk of the time is spent in calculating the energies and gradients, while a negligible part is spent on input/output communication between the main refinement instance and the OPLS3e energy server process. We would recommend to use Phenix/OPLS3e at the final stages of the model building process, which coincides also with the refinement strategy employed.

Re-refining cryo-EM structures with Phenix / OPLS3e

After validating our implementation in reciprocal space refinement using X-ray diffraction data, we applied our protocol on 15 deposited cryo-EM structures and maps, previously used in our GemSpot pipeline (Robertson et al., 2020). Since cryo-EM refinement lacks a robust cross-validation metric, we performed a grid scan over the map weight parameters for our cases, using both the Phenix standard restraints model and the OPLS3e force field. For each refinement, the MolProbity score, Ramachandran Z-score, and cross correlation values were calculated and are shown in Figure 3 and 4 for Phenix/OPLS3e. Individual Clashscores are shown in Figure S3.

It is well known that refinement is a balance between restraints and the available data, implying a trade-off between geometric quality and fit to the data. Interestingly, however, we find that although the cross-correlation value between the model and the map typically increases with increasing weight for the experimental data, structure quality proxied by the MolProbity score remains relatively stable at lower weights, after which the score increases dramatically after a critical weight. Also, when inspecting other proxies, such as the Ramachandran Z-score and Clashscore, a similar behavior is observed (Figure 4 and S3). In cryo-EM modeling it is typically assumed that with equal structure quality the model with higher cross-correlation is preferred, while at equal cross-correlation the model with higher structure quality is preferred. Although there is no current metric that can formally decide how to balance cross-correlation and structure quality, and thus what weight to pick to generate the “most correct” structure, it is clear that there is a reasonable narrow range of acceptable weight values as can be seen in Figure 3, where a noticeable kink is observed in the quality-correlation plots. We therefore propose that providing a single cross correlation value and MolProbity score is not convincing to indicate that the best refined structure is obtained. Moreover, the shape of the quality-correlation curve indicates that the modeler is essentially free to choose the correlation value, reminiscent of earlier X-ray refinement

before the advent of R_{free} where exceedingly low R-factors were obtained. However, cryo-EM lacking a cross-validation metric, a minimal check can be performed by staying within the stable structure quality regime. However, we note that in most cases a weight factor of 3 provides a reasonable estimate for the location of the break point.

To further narrow down the best weight, we also investigated the ligand energy against its local map-model correlation as a function of the weight factor (Figure 5). The shape of the curve follows a similar pattern as the MolProbity score with relatively low ligand energies, where the increase in energy is compared to the lowest energy conformer, at low to intermediate weights together with low real space correlations; while at high weights the correlation values and the energies increase. In addition to the knowledge based MolProbity metrics, the ligand energy provides an additional physics based check on overfitting as highly strained ligands are unlikely to be found. Since the resolution of cryo-EM maps is often non-uniform, the weight of the data during refinement could also be varied locally for optimal model to map fitting, even though this has not been investigated so far. Inspecting the ligand energy versus correlation at different weights provides one route to investigating local weight optimization, which is especially important for the binding site, as here accuracy is key. Based on our findings here coupled with the above observation for the global structure metrics, we advise an energy increase of the ligand to be no more than 1 log unit, i.e. 10 kcal / mol, within Phenix/OPLS3e refinement compared to the lowest energy conformer to reduce local overfitting.

When comparing the standard Phenix restraints model and the OPLS3e force field (Figure S4 and S5), we find no significant difference in the structure quality-correlation curve between the two, owing to the additional restraints introduced by default in the real space refinement protocol including secondary structure, rotamer, NCS and Ramachandran restraints. Although refinements with Phenix/OPLS3e result in a curve starting at low correlations, refinement with the standard Phenix restraints model is more focused, i.e. even at very low weights the correlation remains rather stable as well as the MolProbity score and the Ramachandran Z-score, and the curve is less smooth compared to Phenix/OPLS3e refinement but does still show a critical weight term after which structure quality deteriorates rapidly. This might be due to tighter restraints where even low weights on the geometric terms result in large forces at deviations from ideality, but we did not investigate this further.

Having established a narrow acceptable weight range, we can evaluate the current deposited ligand complexes (see black dots in Figure 3, 4 and 5). When checking the location of the structure quality versus correlation locations of the deposited structures we find that most modelers have been conservative in their modeling and most structures are in the stable MolProbity -- high-cross correlation region. Intriguing, several cases are found very close to the break point between structure quality and real space correlation, e.g. EMD-0280, EMD-4536, and EMD-3953. Other models can be further refined, i.e. both their real space correlation as well as their structure quality can be improved, such as EMD-6997, EMD-0339 and EMD-9781. For 2 cases, EMD-4590 and EMD-0567, the deposited model is found after the break point, and could indicate slight overfitting of the structure. Including the Ramachandran Z-score in the analysis, however, highlights additional cases with signs of

overfitting. A recent investigation of the Ramachandran Z-score indicated that an absolute score below 2 is considered good, between 2 and 3 suspicious, and above 3 bad. Worryingly, for 7 out of 15 cases the initial Z-score is below -3, going down as far as -5.6 for EMD-6997, with all scores being negative except for EMD-9781 which has a Z-score of 2.9. Interestingly, EMD-3953 and EMD-4536 that appeared to be in the stable MolProbity region close to the break point, both show a Z-score below -4. In total, 8 out of 15 structures show signs of overfitting, with either MolProbity scores well beyond the break point or having absolute Ramachandran Z-scores above 3, which is more lenient compared to the Ramachandran Z-score break point as that falls in all cases well before 3. It should also be noted that although the MolProbity score and Ramachandran Z-score are complementary in their detection of overfitting, the Ramachandran Z-score seems the more sensitive, as in all but 1 case it shows signs of overfitting, while the MolProbity score would indicate overfitting for 3 cases.

Investigating the initial models locally by comparing the ligand strain energy and cross-correlation is more complicated, as the energy model used by the initial refinement for the ligand does not necessarily translate to the OPLS3e energies, and would result predictably in excessively high energies. However, in many cases the location of the initial ligand strain energy - cross-correlation point lies either on or is very close to the ligand quality-cross correlation curves of the Phenix/OPLS3e refined cases, e.g. for EMD-0280 and EMD-4590; even when the point does not overlap, the measured cross-correlation remains a decent proxy for overfitting by comparing its value and mapping this to the energy that corresponds to the Phenix/OPLS3e curve. Although the energy models are different and the energies would not necessarily translate, if the cross-correlation is well beyond what is found at a ligand strain energy of 10 kcal/mol as found with the Phenix/OPLS3e curves, this is likely due to local overfitting of the ligand in the map and the resulting correlations are overly optimistic of the goodness-of-fit. However, in 13 out of 15 cases the initial ligand cross correlation is very close to the reported cross correlation at the 10 kcal/mol boundary for the Phenix/OPLS3e refinement curves. Only in EMD-8942 and EMD-8943 could the ligand correlation be slightly improved further.

Although our analysis would indicate that the number of cases that show signs of overfitting is more than half of the 15 cases investigated here, it is of a positive note that the issue is easily remedied by a simple refinement at a lower weight factor for the data. The weight factor should be ideally chosen such that the metrics stay within the stable structure regime, or at least close to the break point for both the Ramachandran Z-score as well as the MolProbity score, while local signs of overfitting can be taken into account by monitoring the ligand strain energy.

The procedure to vary the weight factor between data and model has been investigated earlier by Monroe et al. (2017) in the context of MDFF and Rosetta, where they found that most models could be “refined further”, although their dataset was mainly aimed at lower resolution cases down to 20Å. In addition, besides the difference in software platform used, they did not explicitly show how the weight factor relates to the cross correlation or how to decide on the optimal weight factor at high resolution cryo-EM modeling.

Finally, to help with determining a near-optimal weight, we implemented an automatic weight scanning tool that runs several refinements at different weights and produces a plot similar to Figures 3 and 5 showing MolProbity score versus real space correlation as the global metrics and ligand energy versus correlation for local ligand metrics.

Conclusions

Here we have described our implementation of Phenix/OPLS3e, which incorporates the OPLS3e force field and VSGB2.1 solvation model into the popular refinement package Phenix, for both reciprocal and real space refinement. Our approach alleviates the need to handcraft an accurate restraints dictionary for ligands as gradients are provided by the force field and includes additional physics based forces, most notably electrostatics and a GBSA implicit solvent model. More specifically, for reciprocal space refinement a clear improvement in structure quality is observed starting at resolutions worse than 1.5Å, while R_{free} is generally slightly increased compared to the standard Phenix restraints model, results that are in line with previous observations (Bell et al, 2012; Moriarty et al., 2020). For real space refinement of cryo-EM structures, no significant improvement is noticed in refinement when comparing the structure quality-correlation plots between the two restraints models by scanning across weights, though global and ligand energies are significantly reduced when using Phenix/OPLS3e. We furthermore have shown that, although cryo-EM currently has no universally used cross validation metric, and thus no formal target to determine the weight ratio for data versus restraints, there is actually only a relatively narrow weight range where structure quality is acceptable while correlation is maximized. Insights into local overfitting can be obtained by inspecting ligand strain compared to the low energy conformer. The nature of the quality-correlation curve implies that a weight scan should preferably be performed to find the optimal point that balances structure quality and data fit, i.e. MolProbity score or Ramachandran Z-score and cross correlation, although we find that a weight value of 3 often provides a reasonable estimate. We've also shown that 8 out of 15 cryo-EM cases show signs of overfitting, with 7 having a Ramachandran Z-score lower than -3, and 2 cases, of which one unique, have a MolProbity score well beyond the break point. However, we propose a straightforward solution by performing a simple re-refinement at the appropriate data weight factor. Finally, the use of the OPLS3e/VSGB force field within Phenix is made user-friendly by a tool that includes preparing the input structure, performing the refinement and calculating several validation metrics, and which is available in the 2020–3 release of the Schrödinger software. This release of the software is compatible with Phenix version 1.16 and later.

STAR Methods

RESOURCE AVAILABILITY

Lead contact—Further information and requests for data will be fulfilled by the lead contact Kenneth Borrelli (kenneth.borrelli@schrodinger.com).

Materials availability—No experimental material was used in this study.

Data and code availability—The Phenix software suite is freely available for academic institutions and can be downloaded from <https://www.phenix-online.org>. The OPLS3e force field and VSGB model are part of the Schrödinger software suite, and free licenses are available via the SBGrid Consortium (<https://www.sbgrid.org>) for North American academic institutions. The newly created user-friendly phenix.py pipeline within the Schrödinger distribution does not require any additional licenses. The Maestro interface and the Protein Preparation Wizard are freely available and can be accessed again via the SBGrid Consortium or downloaded from <https://www.schrodinger.com> after registration.

All re-refined structures are freely available upon request and a table of the 2284 re-refined X-ray model PDBIDs and their global refinement statistics can be found in Data S1.

Experimental model and subject—For the X-ray crystallography benchmark all 2282 models and structure factors were downloaded from the PDB and their corresponding PDB IDs can be found in Data S1. Their aggregate analysis after re-refinement is reported in Figures 1, 2, S1, and S2. The cryo-EM benchmark models were downloaded from the PDB with the following IDs: 6cvm, 5a1a, 6a95, 6qm7, 6n4b, 6eti, 6j8h, 6nr3, 6dzz, 6oik, 6o0h, 6huk, 6qee, 6nr2, 6dzy. The experimental cryo-EM maps were downloaded from the EMDB with the following IDs: 7770, 2984, 6997, 4590, 0339, 3953, 9781, 0488, 8943, 20079, 0567, 0280, 4536, 0487, 8942. Individual re-refinement statistics are reported in Figures 3, 4, 5, S3, S4, and S5.

METHOD DETAILS

Overview—Macromolecular refinement with Phenix in general consists of three stages: an initial stage where all the input data is processed and prepared; the actual refinement stage which involves the concept of the macrocycle, an iterative sequence of distinct procedures that optimize different aspects of the model; and an end stage where the refinement is finalized and the structure and any other outputs are written to disk. An important procedure in a macrocycle is the optimization of the atom coordinates, which ultimately is a fine balancing act between the available experimental data and knowledge-based restraints. The target function T during coordinate optimization can be written as a simple linear function consisting of two terms

$$T(x, y, z) = w \cdot E_d + E_r$$

Where E_r and E_d are the (pseudo-)energies of the restraints and fit to the data respectively; w is a weight factor determining the impact of the restraints, and is typically determined automatically during refinement, using differing algorithms depending on the refinement engine, i.e. reciprocal space (`phenix.refine`) or real space refinement (`phenix.real_space_refine`). In the reciprocal space refinement engine, the total weight factor is the product of two terms, where the first term is determined by normalizing the restraints gradients against the data fit gradients, after performing a local minimization and a short molecular dynamics simulation (Adams et al., 1997); the second weight term is by default set to 0.5 or can be determined using a grid search that aims to optimize the metrics for model to data fit, while simultaneously maintaining overall geometric quality

within a reasonable range of values. In real space refinement, when no NCS restraints are present, the weight factor is determined by refining small pieces of the structure against the density at different weights. The final weight is chosen such that a reasonable stereochemistry is maintained. In the case of NCS restraints the whole structure is refined instead of random pieces (Afonine et al., 2018a).

The restraints energy E_r in turn can again be separated into a linear combination of terms given by

$$E_{rc} = E_{ff} + E_{other}$$

Where E_{ff} is the force field energy consisting of a linear combination of bond length, bond and torsion angle, planarity, chirality, parallelity, and nonbonded restraint terms. In the default Phenix restraints model the nonbonded restraints consist of solely a repulsive term to prevent atom clashes. The second E_{other} term is a collection of additional restraints such as non-crystallographic symmetry (NCS) restraints, deformable elastic network (DEN) restraints and others. Our aim is to use the energies and forces provided by the OPLS3e force field and VSGB2.1 energy model and exchange it for the default force field energy term E_{ff} , while still allowing the use of restraints included in E_{other} .

Implementation details of the OPLS3e/VSGB force field in Phenix—Since the Schrödinger software stack is written primarily in Python3, while Phenix currently supports Python-2.7, this makes a direct connection between the software entities challenging. Thus, we decided to create an external program that runs in the background during a refinement using either the `phenix.refine` or `phenix.real_space_refine` engines. This acts as an energy server, running parallel to the refinement process. The external process is initialized using the starting structure in Maestro format and additional options that impact the energy calculation, such as including crystal symmetry or energy flags to include or exclude certain terms. During all subsequent energy calculations, the refinement protocol writes coordinates to disk that are then read by the external server. The external energy server internally updates its coordinates and calculates the energy and gradients of the system, which are then written to disk and read by the refinement program. In the case of a multiconformer model, the energy and gradients are calculated for each consistent conformer, after which the energies and gradients are summed and divided by the number of conformers. The concept of a consistent conformer simply entails that atoms carrying a certain alternate location ID interact only with atoms with no or the same alternate location ID.

Our implementation is versatile and allows for inclusion of the full range of Phenix energy terms, such as NCS and DEN restraints, alongside the OPLS3e energy terms. In addition, the force field can be used for only a subset of the structure, by providing a selection, such as a ligand and its surrounding residues. In this mode, atoms governed by the OPLS3e force field will be fully aware of its environment, while energies and gradients are calculated appropriately for the subset. A more restricted version applies the force field to a ligand that is decoupled from its environment, i.e. interaction energies with the macromolecule are not

included, similar to the Phenix-AFFIT plugin. This is useful when higher resolution data is available for the macromolecule while the ligand densities remain ambiguous, or for larger complexes found typically in cryo-EM where refinement with the full force field can be lengthy. Crystal symmetry, which by default is not used, needs to be explicitly accounted for in the OPLS3e force field energy calculations when using crystallographic data by using the `schrodinger.use_symmetry=True` parameter. After obtaining energies and gradients, the default Phenix energy and gradient terms for the relevant selected atoms (default is all) are subtracted from the total energy and gradients, while the OPLS3e/VSGB force field energies and gradients are added with a weighting term (default is 10). In addition, the riding hydrogen model in the version of Phenix (1.16) used in this work was disabled for the force field selected part, as this increased the energy in the OPLS3e force field calculations. It should furthermore be noted that chemical components need to be consistent with the OPLS3e force field, meaning that the valency rules of atoms need to be adhered to, i.e. hydrogens need to be added to the system. Truncated residues are supported, though requires the addition of non-standard hydrogens.

Performing a refinement with Phenix/OPLS3e—Initially the structure to be refined needs to be processed with Schrödinger’s Protein Preparation Wizard (Sastry et al, 2013), which adds hydrogens and additional bonds between residues and ligands, among other things, so the structure is fully compliant with the OPLS3e force field specification. Although performing a refinement with the force field takes care of all structural restraints, restraint dictionaries describing the topology of ligands still need to be provided to Phenix for restrained atomic displacement (ADP) refinement and to prevent errors due to unknown atom types. Restraint file generation is provided by the `hetgrp_ffgen` utility, an internally used atom typing program, which generates restraints based on the OPLS2005 force field. After processing the input structure and generating the required restraint files, the structure can be refined through the command line. The OPLS3e force field can be requested by setting the parameters `schrodinger.use_schrodinger=True` and `schrodinger.maestro_file=<STRUCT_FN>` where `<STRUCT_FN>` is the processed input file in Maestro format.

The pipeline can be cumbersome to go through manually, as atom naming and residue naming conventions in Phenix are different compared to Schrödinger’s approach, thus there is a consolidated single entry point command line script within the Schrödinger distribution to provide a user-friendly experience. The minimal input solely consists of a PDB or Maestro structure file and a data file and in case of real space refinement the map resolution. The `phenix.refine` (Afonine et al., 2012) engine is used when a file containing reflections in either MTZ or mmCIF format is provided, and `phenix.real_space_refine` (Afonine et al., 2018a) when a map in CCP4 or MRC format file is provided. Currently providing a real space map for reciprocal space refinement is not supported, but is under development for the next release. The calling signature is condensed to the following

```
$SCHRODINGER/run -FROM psp phenix.py <STRUCT_FN> <EXP_DATA_FN> [-resolution
<RESOLUTION>] [options]
```

The resolution option providing the cryo-EM map resolution is required solely for real space refinement. All Phenix refinement parameters can be provided through the command via key-value pairs, or by providing an input parameter file, which is directly passed on to the refinement engine. Before and after refining the structure, validation scripts are run including `phenix.molprobity` (Williams et al., 2018) to calculate geometry statistics and R-factors, `phenix.real_space_correlation` (Liebschner et al., 2019) for real space correlations on a per residue basis when dealing with reciprocal space refinement and `phenix.model_map_cc` (Afonine et al., 2018b) for real space refinement plus total and per-atom based energies are calculated with Prime. All calculated energies and metrics are stored in the output Maestro structure file and are viewable in the Worktable in Maestro and are accessible as structure level and atom level properties. The user-friendly script is available in the Schrödinger 2020–3, and works ‘out of the box’ when both Phenix and the Schrödinger suite are installed and users have set the `SCHRODINGER` and `PHENIX` environment variables to the root directories of the Schrödinger and Phenix distributions.

Benchmarking the impact of the OPLS3e force field on macromolecular refinement—The Phenix/OPLS3e protocol was benchmarked on a diverse dataset of 2284 protein-ligand cases downloaded from the Protein DataBank (wwPDB consortium, 2019). The 2284 cases were selected based on the following criteria: the macromolecule is only protein; the sequences contain less than 30% sequence identity among entries; and the entry should have at least 1 ligand according to Schrödinger’s LigandFinder (total of 4928 ligands). To ascertain the impact of the Schrödinger force field, the cases were divided in 6 different resolution bins with a width of 0.5 starting at 1.0Å and ending at 3.5Å resolution, where each bin contains 11, 863, 823, 433, 56 and 98 number of structures, respectively. Starting structures were downloaded from the PDB database with their associated reflection files in MTZ format using `phenix.fetch_pdb` with the `--mtz` option. Each structure was prepared using Schrödinger’s Protein Preparation Wizard tool to add and optimize hydrogen positions (heavy atoms positions were fixed), and to define bonds and linkages between residues, e.g. disulfide bridges. The output structure is further processed by renaming residues and atom names to be consistent with the naming conventions in Phenix. Waters near special positions were removed for Phenix/OPLS3e as these resulted in non-bonded overlaps in the current Schrödinger energy calculation routines. Input restraint files were generated using `hetgrp_ffgen`. Reciprocal space refinement was performed using `phenix.refine` using as input the PDB and Maestro file describing the structure, the PDB MTZ file, and generated restraint files as input files with the following refinement options: `main.number_of_macrocycles=5, optimize_xyz_weight=True, optimize_adp_weight=True, strategy=individual_sites+individual_adp+occupancies weight_selection_criteria.bonds_rmsd=0.020 weight_selection_criteria.angles_rmsd=2.5 schrodinger.use_schrodinger=True schrodinger.use_symmetry=True schrodinger.maestro_file=<MAESTRO_FILE>`, where `<MAESTRO_FILE>` is the input

structure in Maestro format. The macrocycle thus follows the iterative optimization scheme: 1) locally optimize atom positions; 2) optimize individual ADPs; and 3) optimize the atom occupancies atom positions for residues with alternate conformations or atoms with occupancies less than 1 but greater than 0. The default `individual_sites_real_space` strategy component was disabled, as this typically raised the energy dramatically, presumably due to rotamer idealization. The weight selection criteria for the automatic restraints weight determination were increased compared to default values for refinement at lower resolution, as shown in previous work (Bell et al., 2012). R_{free} flags were generated for the downloaded MTZ files, using `phenix.reflection_file_converter` using default values. For the reference Phenix workflow, which we will refer to as standard Phenix, the input structure and restraint files were prepared using `phenix.ready_set` (Word et al., 1999; Moriarty et al., 2009) with the option `add_h_to_water=True`. The output model, ligand restraint, and metal/link edit files were used as input to `phenix.refine` using the same options as described above, except for the Schrödinger specific options and the weight selection criteria were left at their default values. All re-refined output models were analyzed using `phenix.molprobit` and `phenix.real_space_correlation`, and energies were calculated using Prime. For `phenix.molprobit` we used the following option `keep_hydrogens=True`. Calculations were performed using Phenix-1.16.

Re-refinement of high-resolution cryo-EM models—To determine the impact of the OPLS3e/VSGB force field in cryo-EM structure refinement, 15 protein-ligand complexes were retrieved from the PDB and EMDB (Lawson et al., 2016) containing pharmaceutically active compounds. These 15 are from a total of 19 high-resolution (< 4.5Å resolution) cryo-EM structures that contained a pharmaceutically relevant ligand, and that were originally used to benchmark the GemSpot pipeline, a robust ligand fitting procedure for cryo-EM structures (Robertson et al., 2020). Structures were manually prepared using the `Protein Preparation Wizard` in Maestro. For all cases default options were used in the `Preprocessing` step, i.e. bond orders were assigned, hydrogens were added, metal zero-order bonds created and ligand protonation states determined. Since in some cases a substantial part of the structure was missing side chains, they were added using Prime for all cases. Protein protonation states and hydrogen bond networks were optimized and a restrained energy minimization of hydrogens only was performed. For 6NR2 (EMD-0487) and 6NR3 (EMD-0488) large stretches of unknown (UNK) residues were removed. The refinement pipeline is similar as described above for reciprocal space refinement, with the following differences: cross-correlations were calculated with `phenix.model_map_cc`; Ramachandran Z-scores were calculated with `phenix.rama_z` (Sobolov et al., 2020); and `phenix.real_space_refine` was used as refinement engine with different input parameters: the refinement strategy was set to only the global minimization and ADP optimization protocol, the weight factor of the experimental data in the refinement target function and the exclusion of crystallographic symmetry (*refinement.run=minimization_global+adp, weight=<WEIGHT> and schrodinger.use_symmetry=False*). Since cryo-EM modeling currently lacks a proper broadly accepted cross-validation term, determining an acceptable weight factor in the target function between the restraints and the experimental data is not straightforward. Default

`phenix.real_space_refine` behavior is to perform a guided weight grid search, where pieces of the model are refined at different weights and monitoring the bond length and bond angle RMSD. Default values are set to 0.1 Å and 1.0° angle deviation from ideality, which is too tight for the OPLS3e force field (Bell et al., 2012). Therefore, we performed an explicit weight scan by performing a full refinement at different weights. For comparison we also re-refined the structures using solely Phenix tools as described above and identical refinement options. Calculations were performed using Phenix-1.16, except for `phenix.rama_z` for which Phenix-1.18 was used.

Limitations of current implementation—Our current implementation is lacking support for automatic water and ion placement and simulated annealing in torsion angle space. In addition, multiprocessor use for weight scanning within the refinement protocols is not available. Currently, only two conformers can be handled, as the `Protein Preparation Wizard` only takes into account the A and B conformer when adding hydrogens. Finally, as mentioned above, residues on special positions need to be removed to prevent high energies caused by clashes within the OPLS3e force field.

QUANTIFICATION AND STATISTICAL ANALYSIS

Refinement metrics were calculated using Phenix and energies with Prime. Specifically, MolProbity statistics were calculated with `phenix.molprobity`, real space correlations with `phenix.real_space_correlation` for X-ray crystallographic structures, and `phenix.model_map_cc` for cryo-EM structures, and the Ramachandran Z-scores were calculated with `phenix.rama_z` and `phenix.molprobity`. Bond length and bond angle RMSZ scores were calculated using a custom made Python script using the Phenix API. Phenix-1.16 was used for all calculations, except for the Ramachandran Z-scores, for which Phenix-1.18 was used.

Statistical analysis was performed using a Jupyter notebook and the Python `scipy` and `pandas` package, where the former was used for Welch's t-test using a cutoff p-value of 0.05, although our analysis and conclusions would stand at p-values of 0.01 and lower. The Python packages `seaborn` and `matplotlib` were used for visualization and plotting engine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the US National Institutes of Health (NIH) (grant P01GM063210) and the Phenix Industrial Consortium. This work was supported in part by the US Department of Energy under Contract No. DE-AC02-05CH1123.

References

Adams PD, Pannu NS, Read RJ and Brünger AT (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl. Acad. Sci. U. S. A* 94, 5018–5023. [PubMed: 9144182]

- Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC and Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* 66, 213–221. [PubMed: 20124702]
- Adams PD, Aertgeerts K, Bauer C, Bell JA, Berman HM, Bhat TN, Blaney JM, Bolton E, Bricogne G, Brown D, Burley SK, Case DA, Clark KL, Darden T, Emsley P, Feher VA, Feng Z, Groom CR, Harris SF, Hendle J, Holder T, Joachimiak A, Kleywegt GJ, Krojer T, Marcotrigiano J, Mark AE, Markley JL, Miller M, Minor W, Montelione GT, Murshudov G, Nakagawa A, Nakamura H, Nicholls A, Nicklaus M, Nolte RT, Padyana AK, Peishoff CE, Pieniazek S, Read RJ, Shao C, Sheriff S, Smart O, Soisson S, Spurlino J, Stouch T, Svobodova R, Tempel W, Terwilliger TC, Tronrud D, Velankar S, Ward SC, Warren GL, Westbrook JD, Williams P, Yang H and Young J (2016) Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop. *Structure* 24, 502–508. [PubMed: 27050687]
- Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH and Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* 68, 352–367. [PubMed: 22505256]
- Afonine PV, Poon BK, Read RJ, Sobolev OV, Terwilliger TC, Urzhumtsev A and Adams PD (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* 74, 531–544.
- Afonine PV, Klaholz BP, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, Adams PD, and Urzhumtsev A (2018) New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D* 74, 814–840.
- Alam A, Kowal J, Brode E, Roninson I, Locher KP (2019) Structural insight into substrate and inhibitor discrimination by human P-glycoprotein. *Science*, 363, 753–756 [PubMed: 30765569]
- Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S (2015) 2.2 Å resolution cryo-EM structure of β -galactosidase in complex with a cell-permeant inhibitor. *Science*, 348, 1147–1151 [PubMed: 25953817]
- Bartesaghi A, Aguerrebere C, Falconieri V, Banerjee S, Earl LA, Zhu X, Grigorieff N, Milne JLS, Sapiro G, Wu X, Subramaniam S (2018) Atomic Resolution Cryo-EM Structure of β -Galactosidase. *Structure*, 26, 848–856 [PubMed: 29754826]
- Bell JA, Ho KL and Farid R (2012) Significant reduction in errors associated with nonbonded contacts in protein crystal structures: automated all-atom refinement with PrimeX. *Acta Crystallogr. D* 68, 935–952. [PubMed: 22868759]
- Beshnova DA, Pereira J and Lamzin VS (2017) Estimation of the protein-ligand interaction energy for model building and validation. *Acta Crystallogr. D* 73, 195–202.
- Borbulevych OY, Plumley JA, Martin RI, Merz KM and Westerhoff LM (2014) Accurate macromolecular crystallographic refinement: incorporation of the linear scaling, semiempirical quantum-mechanics program DivCon into the PHENIX refinement package. *Acta Crystallogr. D* 70, 1233–1247. [PubMed: 24816093]
- Brunger Axel T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355, 472. [PubMed: 18481394]
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T and Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* 54, 905–921. [PubMed: 9757107]
- Cereto-Massagué A, Ojeda MJ, Joosten RP, Valls C, Mulero M, Salvado MJ, Arola-Arnal A, Arola L, Garcia-Vallvé S, and Pujadas G (2013) The good, the bad and the dubious: VHELIBS, a validation helper for ligands and binding sites. *J. Cheminformatics* 5, 36.
- Coleman JA, Yang D, Zhao Z, Wen PC, Yoshioka C, Tajkhorshid E, Gouaux E (2019) Serotonin transporter-ibogaine complexes illuminate mechanisms of inhibition and transport. *Nature*, 569, 141–145 [PubMed: 31019304]
- Deller MC and Rupp B (2015) Models of protein-ligand crystal structures: trust, but verify. *J. Comput. Mol. Des* 29, 817–836.

- DiMaio F, Zhang J, Chiu W, and Baker D (2013) Cryo-em model validation using independent map reconstructions. *Protein Sci* 22, 865–868. [PubMed: 23592445]
- Engh RA and Huber R (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* 47, 392–400.
- Engh RA and Huber R (2012) Structure quality and target parameters. In *Int. Tables Crystallogr* Rossmann MG & Arnold E ed. (Kluwer Academic Publishers), pp. 382–392.
- Groom CR, Bruno IJ, Lightfoot MP and Ward SC (2016) The Cambridge Structural Database. *Acta Crystallogr. B* B72, 171–179.
- Falkner B and Schröder GF (2013) Cross-validation in cryo-EM-based structural modeling. *Proc. Natl. Acad. Sci. U. S. A* 110, 8930–8935. [PubMed: 23674685]
- Fenn TD, Schnieders MJ, Mustyakimov M, Wu C, Langan P, Pande VS and Brunger AT (2011) Reintroducing electrostatics into macromolecular crystallographic refinement: application to neutron crystallography and DNA hydration. *Structure*, 19, 523–533. [PubMed: 21481775]
- Headd JJ, Echols N, Afonine PV, Grosse-Kunstleve RW, Chen VB, Moriarty NW, Richardson DC, Richardson JS and Adams PD (2012) Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr. D* 68, 381–390. [PubMed: 22505258]
- Jackson SM, Manolaridis I, Kowal J, Zechner M, Taylor NMI, Bause M, Bauer S, Bartholomaeus R, Bernhardt G, Koenig B, Buschauer A, Stahlberg H, Altmann KH, Locher KP (2018) Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nat. Struct. Mol. Biol.* 25, 333–340. [PubMed: 29610494]
- Jaskolski M, Gilski M, Dauter Z and Wlodawer A (2007) Stereochemical restraints revisited: how accurate are refinement targets and how much should protein structures be allowed to deviate from them? *Acta Crystallogr. D* 63, 611–620. [PubMed: 17452786]
- Janowski PA, Moriarty NW, Kelley BP, Case DA, York DM, Adams PD and Warren GL (2016) Improved ligand geometries in crystallographic refinement using AFITT in PHENIX. *Acta Crystallogr. D* 72, 1062–1072.
- Joosten RP, Long F, Murshudov GN and Perrakis A (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ*, 1, 213–220.
- Li J, Abel R, Zhu K, Cao Y, Zhao S and Friesner RA (2011) The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins* 79, 2794–2812. [PubMed: 21905107]
- Korostelev A, Fenley MO, and Chapman MS (2004) Impact of a Poisson-Boltzmann electrostatic restraint on protein structures refined at medium resolution. *Acta Crystallogr. D* 60, 1786–1794. [PubMed: 15388925]
- Kovalevskiy O, Nicholls RA, Long F, Carlon A, and Murshudov GN (2018) Overview of refinement procedures within REFMAC5: utilizing data from different sources. *Acta Crystallogr. D* 74, 215–227.
- Krishna Kumar K, Shalev-Benami M, Robertson MJ, Hu H, Banister SD, Hollingsworth SA, Latorraca NR, Kato HE, Hilger D, Maeda S, Weis WI, Farrens DL, Dror RO, Malhotra SV, Kobilka BK, Skiniotis G (2018) Structure of a Signaling Cannabinoid Receptor 1-G Protein Complex. *Cell*, 176, 448–458
- Lagerstedt JA, Jakobi A, Burnley T, Patwardhan A, Topf M and Winn M (2020) Comparing cryo-EM reconstructions and validating atomic model fit using difference maps. *J. Chem. Inf. Model* 60, 2552–2560. [PubMed: 32043355]
- Lawson CL, Patwardhan A, Baker ML, Hryc C, Sanz Garcia E, Hudson BP, Lagerstedt I, Ludtke SJ, Pintilie G, Sala R, Westbrook JD, Berman HM, Kleywegt GJ, Chiu W (2016) EMDatabank unified data resource for 3DEM. *Nucl. Acids Res* 44, D396–D403. [PubMed: 26578576]
- Li Q, Pellegrino J, Lee DJ, Tran AA, Chaires HA, Wang R, Park JE, Ji K, Chow D, Zhang N, Brilot AF, Biel JT, van Zundert G, Borrelli K, Shinabarger D, Wolfe C, Murray B, Jacobson MP, Mühlh E, Chesneau O, Fraser JF and Seiple IB (2020) Synthetic group A streptogramin antibiotics that overcome Vat resistance. *Nature* 586 145–150. [PubMed: 32968273]
- Liebeschuetz J, Hennemann J, Olsson T and Groom CR (2012) The good, the bad and the twisted: a survey of ligand geometry in protein crystal structures. *J. Comput. Mol. Des* 26, 169–183.

- Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung L-W, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger TC, Urzhumtsev AG, Videau LL, Williams CJ & Adams PD (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* 75, 861–877.
- Maeda S, Qu Q, Robertson MJ, Skiniotis G, Kobilka BK (2019) Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes. *Science*, 364, 552–557. [PubMed: 31073061]
- Masiulis S, Desai R, Uchanski T, Serna Martin I, Lavery D, Karia D, Malinauskas T, Zivanov J, Pardon E, Kotecha A, Steyaert J, Miller KW, Aricescu AR (2019) GABAA receptor signalling mechanisms revealed by structural pharmacology. *Nature*, 565, 454–459 [PubMed: 30602790]
- Monroe L and Terashi G and Kihara D (2017) Variability of protein structure models from electron microscopy. *Structure* 25, 592–602. [PubMed: 28262392]
- Moriarty NW, Grosse-Kunstleve RW, and Adams PD (2009) Electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D*, 2009, Vol. 65, pp. 1074–1080. [PubMed: 19770504]
- Moriarty NW, Tronrud DE, Adams PD and Karplus PA A new default restraint library for the protein backbone in Phenix: a conformation-dependent geometry goes mainstream. *Acta Crystallogr. D*, 2016, Vol. 72, pp. 176–179.
- Moriarty NW, Janowski PA, Swails JM, Nguyen H, Richardson JS, Case DA and Adams PD (2020) Improved chemistry restraints for crystallographic refinement by integrating the Amber force field into Phenix. *Acta Crystallogr. D* 76, 51–62.
- Moulinier L and Case DA and Simonson T (2003) Reintroducing electrostatics into protein X-ray structure refinement: bulk solvent treated as a dielectric continuum. *Acta Crystallogr. D* 59, 2094–2103. [PubMed: 14646067]
- Peach ML, Cachau RE and Nicklaus MC (2017) Conformational energy range of ligands in protein crystal structures: The difficult quest for accurate understanding. *J. Mol. Recognit* 30, E2618
- Ponder JW and Case DA (2003) Force fields for protein simulations. *Adv. Protein Chem* 66, 27–85. [PubMed: 14631816]
- Reynolds CH (2014) Protein-ligand cocrystal structures: we can do better. *ACS Med. Chem. Lett* 5, 727–729. [PubMed: 25050154]
- Robertson M.I.J., van Zundert GCP, Borrelli K, and Skiniotis G (2020) GemSpot: A Pipeline for Robust Modeling of Ligands into CryoEM Maps. *Structure* 28, 707–716. [PubMed: 32413291]
- Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, Dahlgren MK, Mondal S, Chen W, Wang L, Abel R, Friesner RA and Harder ED (2019) OPLS3e: Extending force field coverage for drug-like small molecules. *J. Chem. Theory Comput* 5, 1863–1874.
- Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des* 27, 221–234. [PubMed: 23579614]
- Schröder GF and Levitt M and Brunger AT (2014) Deformable elastic network refinement for low-resolution macromolecular crystallography. *Acta Crystallogr. D* 70, 2241–2255. [PubMed: 25195739]
- Shen H, Li Z, Jiang Y, Pan X, Wu J, Cristofori-Armstrong B, Smith JJ, Chin YKY, Lei J, Zhou Q, King GF, Yan N (2018) Structural basis for the modulation of voltage-gated sodium channels by animal toxins. *Science*, 362, eaau2596 [PubMed: 30049784]
- Shen H, Liu D, Wu K, Lei J, Yan N (2019) Structures of human Nav1.7 channel in complex with auxiliary subunits and animal toxins. *Science*, 22, 1303–1308
- Sitzmann M, Weidlich IE, Filippov IV, Liao C, Peach ML, Ihlenfeldt W-D, Karki RG, Borodina YV, Cachau RE and Nicklaus MC (2012) PDB ligand conformational energies calculated quantum-mechanically. *J. Chem. Inf. Model* 52, 739–756. [PubMed: 22303903]
- Smart OS, Horský V, Gore S, Svobodová Va eková R, Bendová V, Kleywegt GJ and Velankar S (2018) Validation of ligands in macromolecular structures determined by X-ray crystallography. *Acta Crystallogr. D* 74, 228–236.

- Sobolev OV, Afonine PV, Moriarty NW, Hekkelman ML, Joosten RP, Perrakis A, Adams PD (2020) A Global Ramachandran Score Identifies Protein Structures with Unlikely Stereochemistry. *Structure* 28, 1249–1258. [PubMed: 32857966]
- Steiner RA and Tucker JA (2017) Keep it together: restraints in crystallographic refinement of macromolecule-ligand complexes. *Acta Crystallogr. D* 73, 93–102.
- Tickle IJ, Laskowski RA, and Moss DS (1998) Rfree and the Rfree Ratio. I. Derivation of Expected Values of Cross-Validation Residuals Used in Macromolecular Least-Squares Refinement. *Acta Crystallogr. D* 54, 547–557. [PubMed: 9761849]
- Tronrud DE, Berkholtz DS, and Karplus PA (2010) Using a conformation-dependent stereochemical library improves crystallographic refinement of proteins. *Acta Crystallogr. D* 66, 834–843. [PubMed: 20606264]
- Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, & Murshudov GN (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta crystallogr. D* 60, 2184–2195. [PubMed: 15572771]
- Volkman N (2009) Confidence intervals for fitting of atomic models into low-resolution densities. *Acta Crystallogr. D* 65, 679–689. [PubMed: 19564688]
- Wang J and Moore PB (2017) On the interpretation of electron microscopic maps of biological macromolecules. *Protein Sci* 26, 122–129. [PubMed: 27706888]
- Wang L, Kruse H, Sobolev OV, Moriarty NW, Waller MP, Afonine PV, Biczysko M (2020) Real-space quantum-based refinement for cryo-EM: Q|R#3. *Acta Crystallogr D* 76, 1184–1191.
- Wei J, Leit S, Kuai J, Therrien E, Rafi S, Harwood HJ, DeLaBarre B, Tong L (2019) An allosteric mechanism for potent inhibition of human ATP-citrate lyase. *Nature*, 568, 566–570 [PubMed: 30944472]
- Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB, Snoeyink J, Adams PD, Lovell SC, Richardson JS and Richardson DC (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* 27, 293–315. [PubMed: 29067766]
- Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol* 285, 1735–1747. [PubMed: 9917408]
- wwPDB consortium. (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucl. Acids Res* 47, D520–D528. [PubMed: 30357364]
- Wyllie S, Brand S, Thomas M, De Rycker M, Chung CW, Pena I, Bingham RP, Bueren-Calabuig JA, Cantizani J, Cebrian D, Craggs PD, Ferguson L, Goswami P, Hobrath J, Howe J, Jeacock L, Ko EJ, Korczynska J, MacLean L, Manthri S, Martinez MS, Mata-Cantero L, Moniz S, Nühs A, Osuna-Cabello M, Pinto E, Riley J, Robinson S, Rowland P, Simeons FRC, Shishikura Y, Spinks D, Stojanovski L, Thomas J, Thompson S, Viayna Gaza E, Wall RJ, Zuccotto F, Horn D, Ferguson MAJ, Fairlamb AH, Fiandor JM, Martin J, Gray DW, Miles TJ, Gilbert IH, Read KD, Marco M, Wyatt PG (2019) Preclinical candidate for the treatment of visceral leishmaniasis that acts through proteasome inhibition. *Proc. Natl. Acad. Sci. U. S. A.*, 116, 9318–9323 [PubMed: 30962368]
- Yin Y, Le SC, Hsu AL, Borgnia MJ, Yang H, Lee SY (2019) Structural basis of cooling agent and lipid sensing by the cold-activated TRPM8 channel. (2019) *Science*, 363, eaav9334 [PubMed: 30733385]
- Zheng M, Moriarty NW, Xu YT, Reimers JR, Afonine PV and Waller MP (2017) Solving the scalability issue in quantum-based refinement: Q|R#1. *Acta Crystallogr. D* 73, 1020–1028.
- Zheng M, Biczysko M, Xu Y, Moriarty NW, Kruse H, Urzhumtsev A, Waller MP, Afonine PV. (2020) Including crystallographic symmetry in quantum-based refinement: Q|R#2. *Acta Crystallogr. D* 76, 41–50.

Highlights

- The OPLS3e force field and VSGB2.1 solvation model are incorporated into Phenix
- Structure and ligand quality for lower resolution X-ray models improves
- Global and local structure quality measures provide insight into overfitting
- Indications of cryo-EM model overfitting are found in more than half the cases

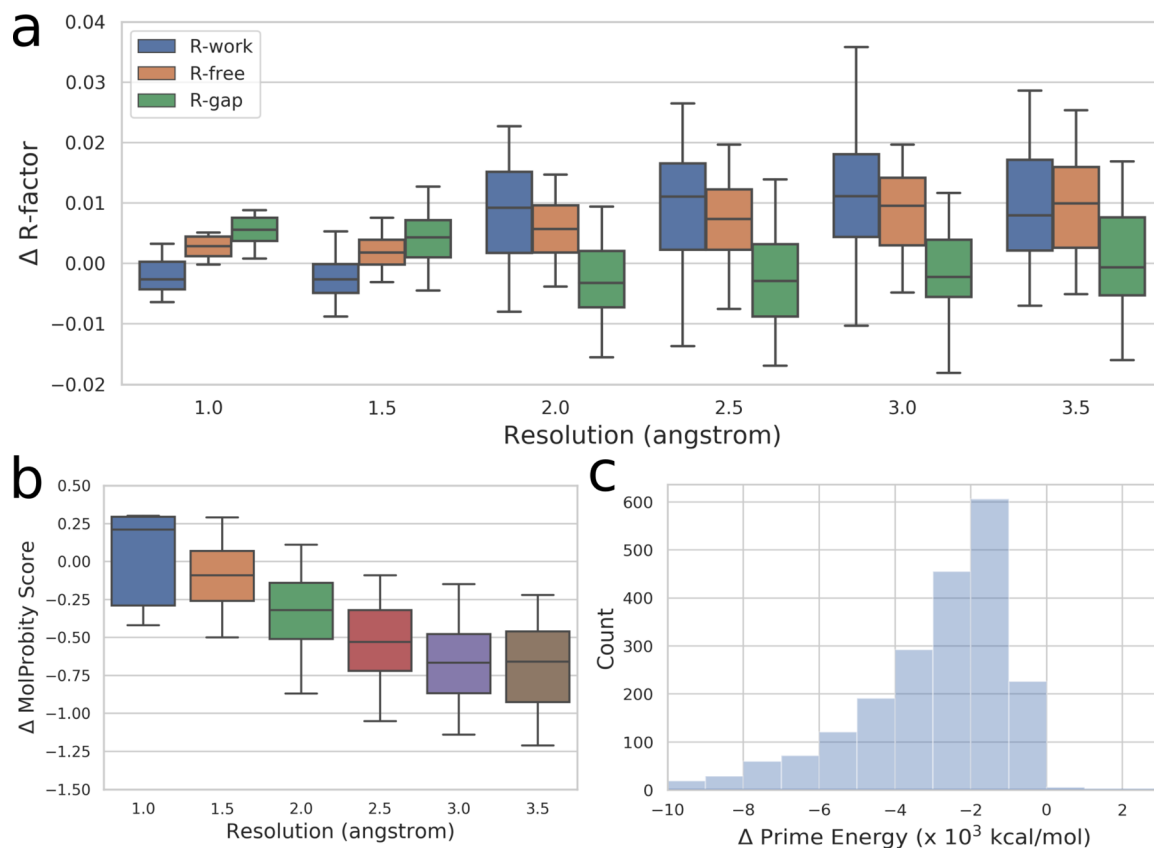


Figure 1. Global structure refinement metric differences.

Shown are the difference in R_{work} , R_{free} and R_{gap} (a) and MolProbity score (b) distributions between standard Phenix refinement and Phenix/OPLS3e refinement. Differences are calculated by subtracting standard Phenix refined structure metrics from Phenix/OPLS3e refined structure metrics. The whiskers in the box plots represent the 5th to the 95th percentile, the box the 25th to 75th percentile and the line in the box represents the median. (c) Energy difference distribution between Phenix/OPLS3e refined structure and starting structure. See also Data S1.

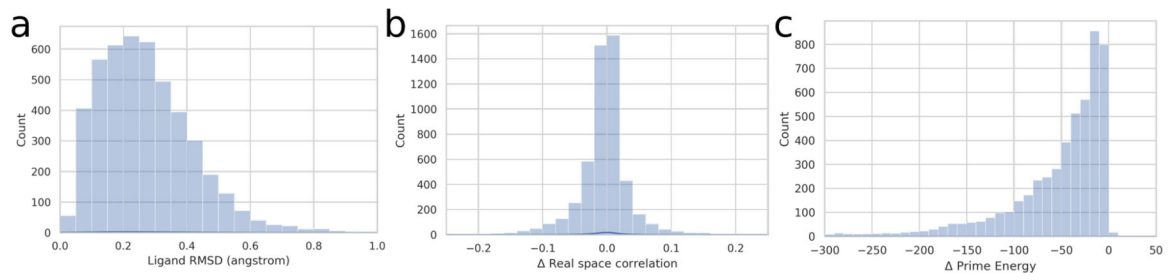


Figure 2. Ligand metrics before and after refinement with Phenix/OPLS3e.

Distribution difference plots before and after Phenix/OPLS3e refinement are shown for ligand RMSD (a), real space cross correlation (b), and Prime Energy (c) when starting out with the PDB deposited structure.

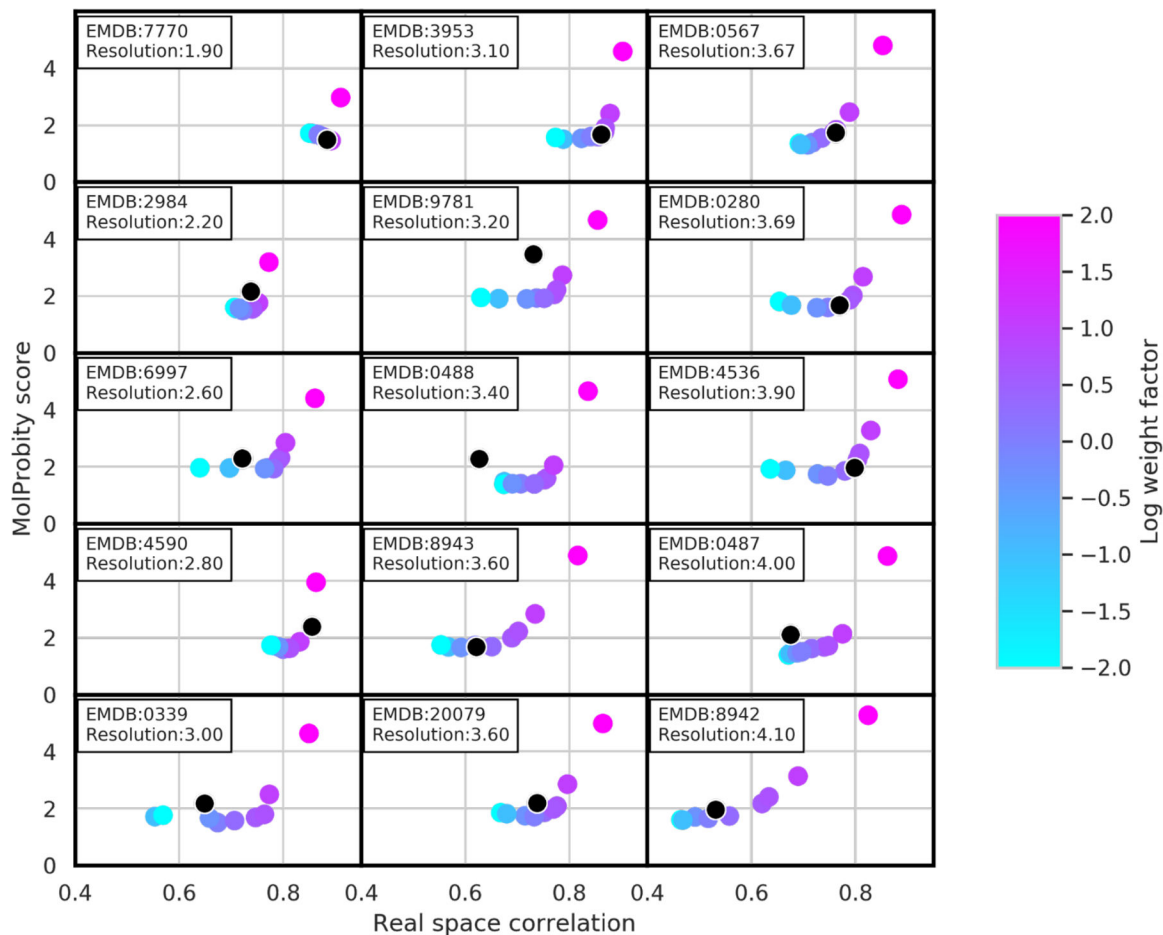


Figure 3. Phenix/OPLS3e real space refinement and the impact of data weight.

The MolProbity score versus real space correlation is shown after refinement at different weight factors when using the OPLS3e force field. The black dot represents the deposited structure. The color of each dot represents the weight parameter for the impact of the Coulomb potential map used during refinement with blue indicating lower weights, and pink indicating higher weights.

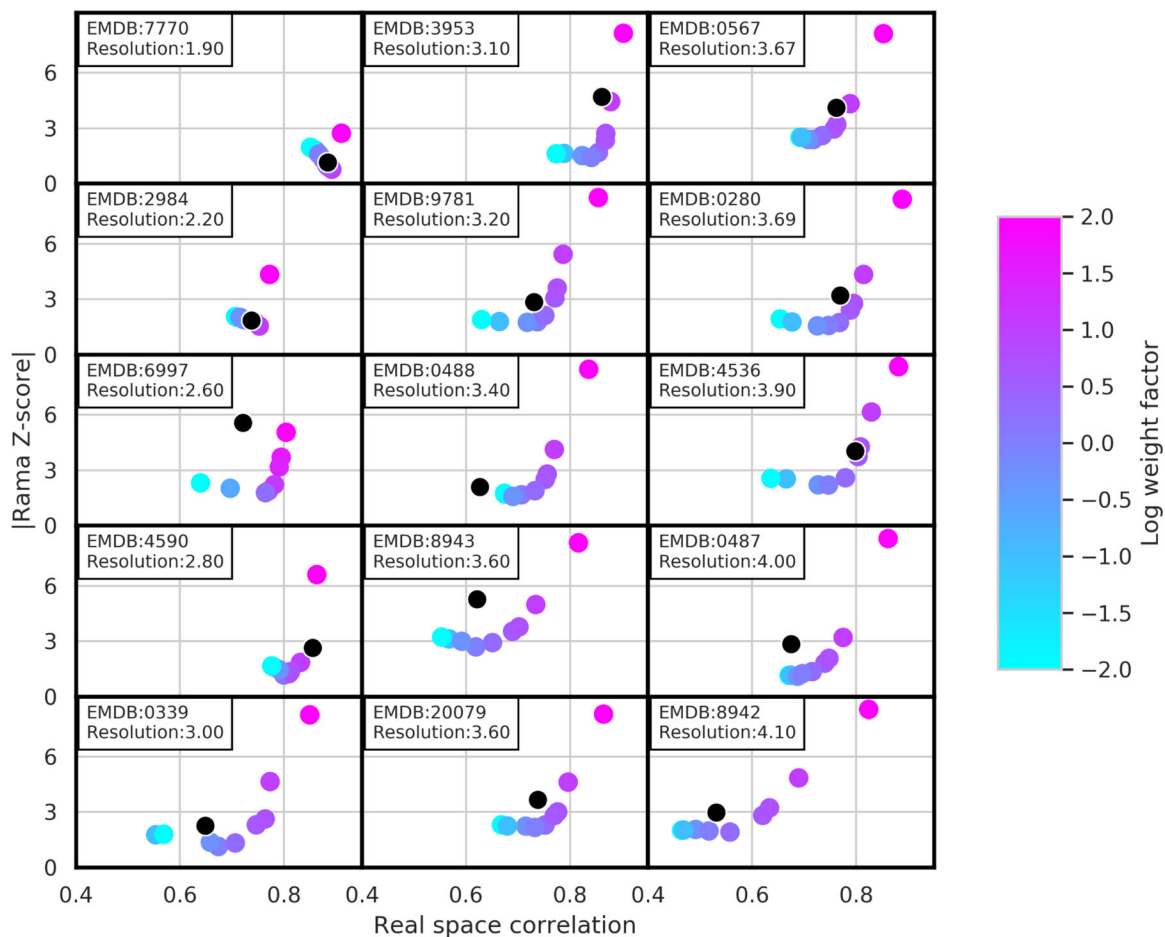


Figure 4. Phenix/OPLS3e real space refinement and the impact of data weight on the Ramachandran Z-score.

The absolute Ramachandran Z-score versus real space correlation is shown after refinement at different weight factors when using the OPLS3e force field. The black dot represents the deposited structure. The color of each dot represents the weight parameter for the impact of the Coulomb potential map used during refinement with blue indicating lower weights, and pink indicating higher weights. All values shown were negative, except for the initial (black dot) value of EMD-9781.

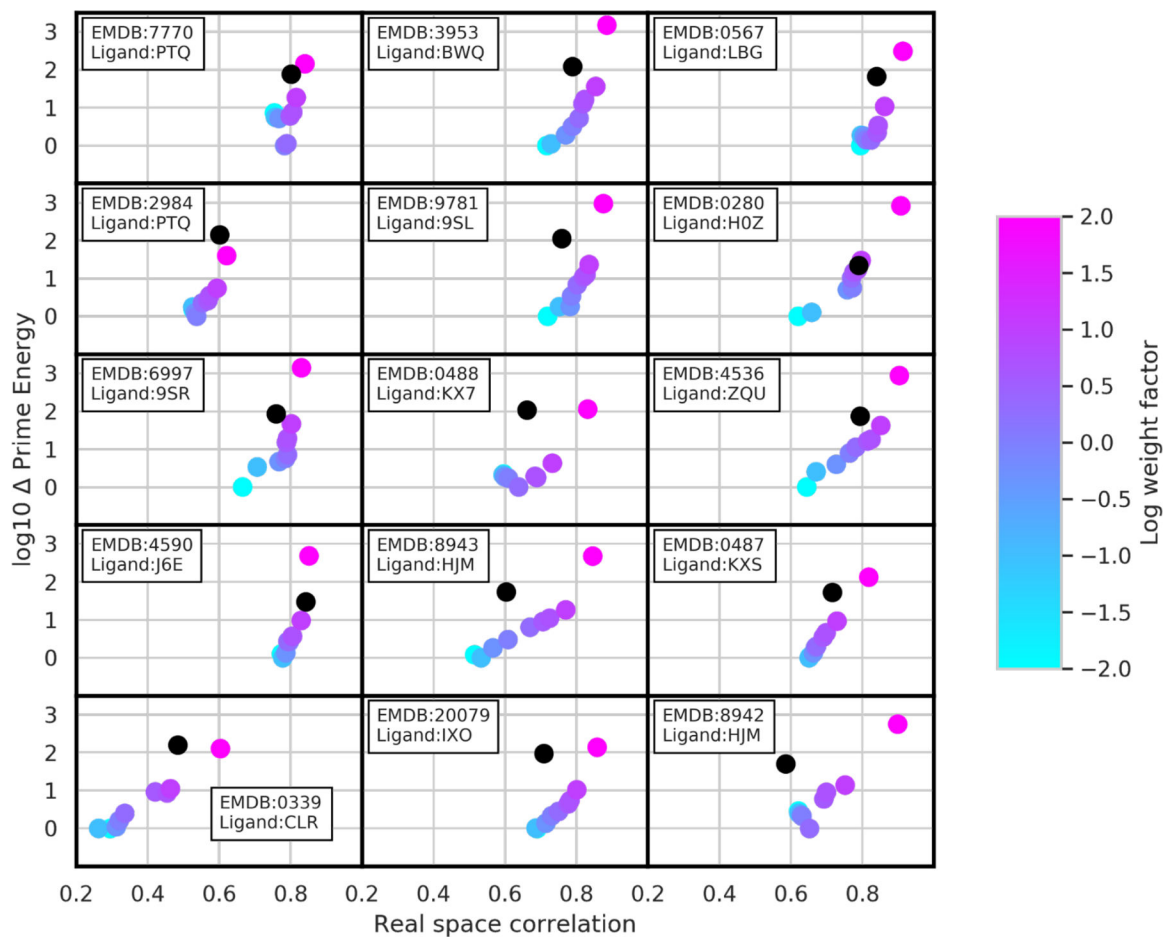


Figure 5. Ligand energies and correlations after Phenix/OPLS3e real space refinement
 Energy differences against the lowest energy conformer for 5 ligands, one for each cryo-EM structure, against the cross correlation are shown. The energy of the lowest energy conformer was set to 1 and the energy axis is on a log scale. The black dot represents the ligand as found in the deposited structure. The color of the dot represents the impact of the Coulomb potential map, with blue tints indicating a low weight and pink tints indicating a high weight.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Beta-galactosidase / 2-phenylethyl 1-thio-beta-D-galactopyranoside	Bartesaghi et al., 2018	PDB: 6cvm, EMD: 7770
Beta-galactosidase / 2-phenylethyl 1-thio-beta-D-galactopyranoside	Bartesaghi et al., 2015	PDB: 5a1a, EMD: 2984
Sodium channel protein PaFPC1 / Tetrodotoxin	Shen et al., 2018	PDB: 6a95, EMD: 6995
L. tarentolae proteasome 20S / GSK3494245	Wyllie et al., 2019	PDB: 6qm7, EMD: 4590
Cannabinoid Receptor 1-G Protein Complex / Fubinaca	Krishna Kumar et al., 2018	PDB: 6n4b, EMD: 0339
Broad substrate specificity ATP-binding cassette transporter ABCG2 / BWQ	Jackson et al., 2018	PDB: 6eti, EMD: 3953
Sodium channel Nav1.7 / Saxitoxin	Shen et al., 2019	PDB: 6j8h, EMD: 9781
TRPM8 ion channel / icilin	Yin et al. 2019	PDB: 6nr3, EMD: 0488
Serotonin transporter / ibogaine	Coleman et al., 2019	PDB: 6dzz, EMD: 8943
Muscarinic acetylcholine receptor 2-Go complex / Iperoxo	Maeda et al., 2019	PDB: 6oiik, EMD: 20079
ATP-citrate lyase / NDI-091143	Wei et al., 2019	PDB: 6o0h, EMD: 0567
GABA _A receptor / Biculine	Masiulis et al., 2019	PDB: 6huk, EMD: 0280
Human-mouse chimeric ABCB1 / Zosuquidar	Alam et al., 2019	PDB: 6qee, EMD: 4536
TRPM8 ion channel / menthol analog WS-12	Yin et al. 2019	PDB: 6nr2, EMD: 0487
Serotonin transporter / ibogaine	Coleman et al., 2019	PDB: 6dzy, EMD: 8942
Software and Algorithms		
Prime	Bell et al., 2012	https://www.schrodinger.com
Maestro	Schrödinger Release 2020–3: Maestro, Schrödinger, LLC, New York, NY, 2020.	https://www.schrodinger.com
Protein Preparation Wizard	Sastry et al., 2013	https://www.schrodinger.com
phenix.refine	Afonine et al., 2012	https://www.phenix-online.org
phenix.real_space_refine	Afonine et al., 2018a	https://www.phenix-online.org
phenix.rama_z	Sobolev et al., 2020	https://www.phenix-online.org
phenix.molprobity	Williams et al., 2018	https://www.phenix-online.org
phenix.real_space_correlation	Liebschner et al., 2019	https://www.phenix-online.org
phenix.model_map_cc	Afonine et al., 2018b	https://www.phenix-online.org
phenix.ready_set	Word et al., 1999; Moriarty et al., 2009	https://www.phenix-online.org