

ChemEM: Flexible Docking of Small Molecules in Cryo-EM Structures

Aaron Sweeney, Thomas Mulvaney, Mauro Maiorca, and Maya Topf*

Cite This: *J. Med. Chem.* 2024, 67, 199–212

Read Online

ACCESS |



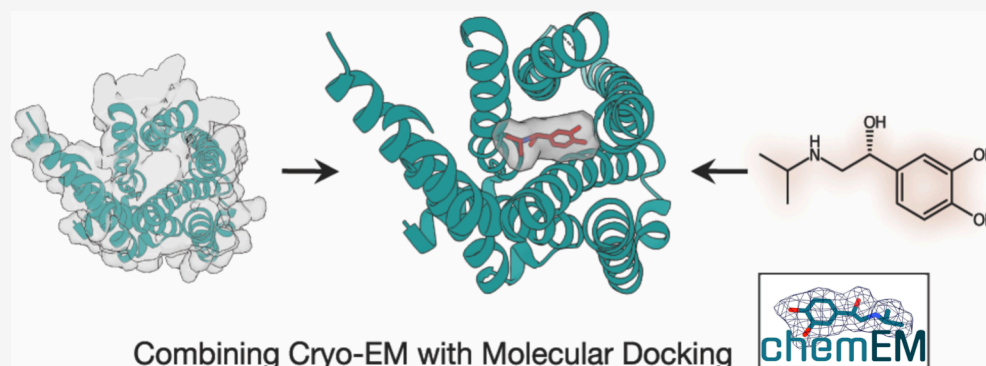
Metrics & More



Article Recommendations



Supporting Information



Combining Cryo-EM with Molecular Docking

ABSTRACT: Cryo-electron microscopy (cryo-EM), through resolution advancements, has become pivotal in structure-based drug discovery. However, most cryo-EM structures are solved at 3–4 Å resolution, posing challenges for small-molecule docking and structure-based virtual screening due to issues in the precise positioning of ligands and the surrounding side chains. We present ChemEM, a software package that employs cryo-EM data for the accurate docking of one or multiple ligands in a protein-binding site. Validated against a highly curated benchmark of high- and medium-resolution cryo-EM structures and the corresponding high-resolution controls, ChemEM displayed impressive performance, accurately placing ligands in all but one case, often surpassing cryo-EM PDB-deposited solutions. Even without including the cryo-EM density, the ChemEM scoring function outperformed the well-established AutoDock Vina score. Using ChemEM, we illustrate that valuable information can be extracted from maps at medium resolution and underline the utility of cryo-EM structures for drug discovery.

INTRODUCTION

Advancements in cryo-EM have increased the relevance of the technique in structure-based drug discovery. This is evidenced by an increase in the number of cryo-EM maps deposited in the Electron Microscopy Databank (EMDB)¹ at a resolution of 3 Å or better, up from 15 in 2015 to over 2000 in 2022. Advances in the resolution limits of cryo-EM have facilitated the use of more potential drug targets, including membrane proteins, in structure-based drug discovery.^{2–5} Improvements in the resolution limits of cryo-electron microscopy (cryo-EM) have broadened our understanding of drug target identification, design, and development, particularly in areas that proved challenging for other techniques, such as X-ray crystallography. A notable instance is the exploration of membrane proteins, such as the GABA_A receptors.⁵ These pentameric receptors, predominantly consisting of two alpha, two beta, and one gamma subunit, proved difficult to fully decipher using X-ray crystallography alone. Previous attempts yielded only a beta homopentamer⁶ or chimeric structures.⁷ However, cryo-EM has enabled detailed characterization of more physiologically relevant GABA_A receptor isoforms. More importantly, it has illuminated the interaction sites for drugs such as diazepam, which bind at the interface of the alpha and gamma subunits.

Although atomic resolution is achievable,⁸ the majority of maps are still deposited with a global resolution between 3 and 4 Å¹. Furthermore, most cryo-EM maps contain heterogeneous resolutions across the map. Even in high-resolution maps, small perturbations in the position of ligands, multiple conformations, or partial occupancy within the binding site may worsen the local resolution of the binding site relative to the global resolution. Therefore, it is necessary to develop methodologies for the accurate modeling of small molecules at both high and low resolutions.

Only recently have methodologies that specifically focus on small-molecule fitting in cryo-EM emerged.^{9,10} The success of such a process is dependent on the initial fit of the ligand, as having a ligand positioned close to the correct position increases the chance of generating a high-quality fit to the map. These methodologies have leveraged molecular docking for

Received: June 23, 2023

Revised: November 28, 2023

Accepted: December 8, 2023

Published: December 29, 2023



generating initial fits. Molecular docking software uses scoring functions to describe the interactions between a protein and a small molecule and identify physicochemically reasonable ligand conformations within a given protein binding site. One of the most common softwares, AutoDock Vina,¹¹ was shown to have among the best scoring functions for estimating affinity and identifying correct conformations within a set of decoys in the Critical Assessment of Scoring Functions 2016 report¹² (CASF-2016). In the commercial software, GemSpot,⁹ such a docking score is integrated with a goodness-of-fit term—the cross correlation coefficient (CCC). This allows only ligand conformations with a good fit to the map and favorable docking score to be selected for further refinement. While the CCC is the “gold standard” goodness-of-fit score for fitting atomic structures in cryo-EM, there is evidence that the Mutual Information (MI) score provides a comparable measure of fit at high resolution and is more robust to a decrease in resolution when fitting protein models to maps.¹³ However, data are lacking as to whether this can be applied to small-molecule fitting.

Thus, far, the goodness-of-fit has been assessed mostly between a density map simulated from the model and the experimental cryo-EM density map. However, methodologies where density-difference maps are used instead of the full map have been employed. For example, to fit the BTB-1 ligand bound to a human kinesin-8 motor domain at 4.8 Å resolution¹⁴ and the GSK-1 ligand bound to Eg-5 at 3.8 Å resolution,¹⁵ the molecular docking software GOLD,¹⁶ HADDOCK,¹⁷ and AutoDock Vina¹¹ were used along with difference maps for ligand fitting.

Recently, a technique for difference mapping based on local scaling of the cryo-EM map amplitudes was shown to be capable of identifying ligand density by subtracting an apoprotein density map from a map of the protein–ligand complex.¹⁸ However, there are no data that compare these procedures to using the full map. Additionally, an integrative automated workflow optimized for docking small molecules utilizing density difference maps has not yet been proposed.

In the following paper, we describe ChemEM—a software package for docking small molecules into medium-to-high-resolution cryo-EM structures. We first describe a novel molecular docking scoring function (ChemDock) shown to have a performance comparable to or better than that of the AutoDock Vina scoring function. We then show the usefulness of the MI score in fitting small molecules into cryo-EM difference maps compared with the commonly used CCC, especially for medium resolutions. Finally, we introduce ChemEM, a fully automated flexible docking software, which combines both ChemDock and the MI score to better identify and refine ligand conformation in binding sites of cryo-EM structures. ChemEM was validated using 32 experimental structures of proteins bound to small molecules at resolutions between 2.2 and 5.6 Å.

RESULTS

ChemEM Workflow. ChemEM docks small molecules to cryo-EM maps using a two-stage approach (Figure 1). In Stage 1, an approximate fit of a ligand into the protein structure is generated using a difference density map between the protein–ligand complex map and a simulated protein map.

An initial fit is achieved by molecular docking using an empirical scoring function integrated with the MI score. This ensures that the generated conformation is both physicochemi-

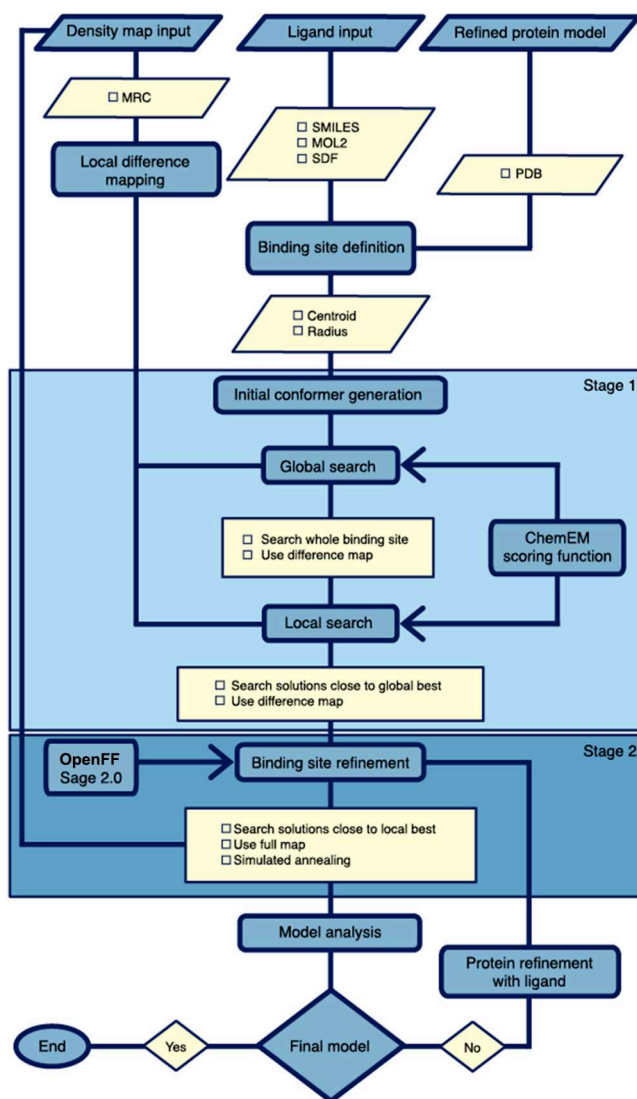


Figure 1. Schematic overview of the ChemEM methodology for docking small molecules, also in combination with cryo-EM maps. For running docking with ChemDock, ligand and protein inputs are needed. For running ChemEM, an additional density map input is required. Additionally, the binding site centroid and radius must be set by the user to identify protein residues that line the binding site. For ChemEM, a local density difference map is created using the user-supplied binding site definitions. For both ChemDock and ChemEM, stage 1 generates initial fits of small molecules within the binding site using either the ChemDock or ChemEM scoring functions. The global search is an Ant colony optimization (ACO) algorithm, while the local search is a Nelder–Mead minimization algorithm. Stage 2 refines the initial fits in the context of the binding site. Here, all binding site residue side chains are treated as fully flexible. The OpenFF Sage v2.0 force field is used for small-molecule parametrization, with protein parameters taken from the AMBER FF14SB force field. For ChemDock workflow, only energy minimization is applied. For the ChemEM workflow, the full density map term is added to the scoring function, and a simulated annealing search algorithm is used to identify final solutions.

cally acceptable and well-fitted to the difference map. In this stage, the difference map acts as a constraint for the docking algorithm, reducing the conformational space to be searched.

In Stage 2 (Figure 1), candidate conformations are refined into the full cryo-EM density map using a flexible fitting approach with the AMBER force field-14SB¹⁹ protein

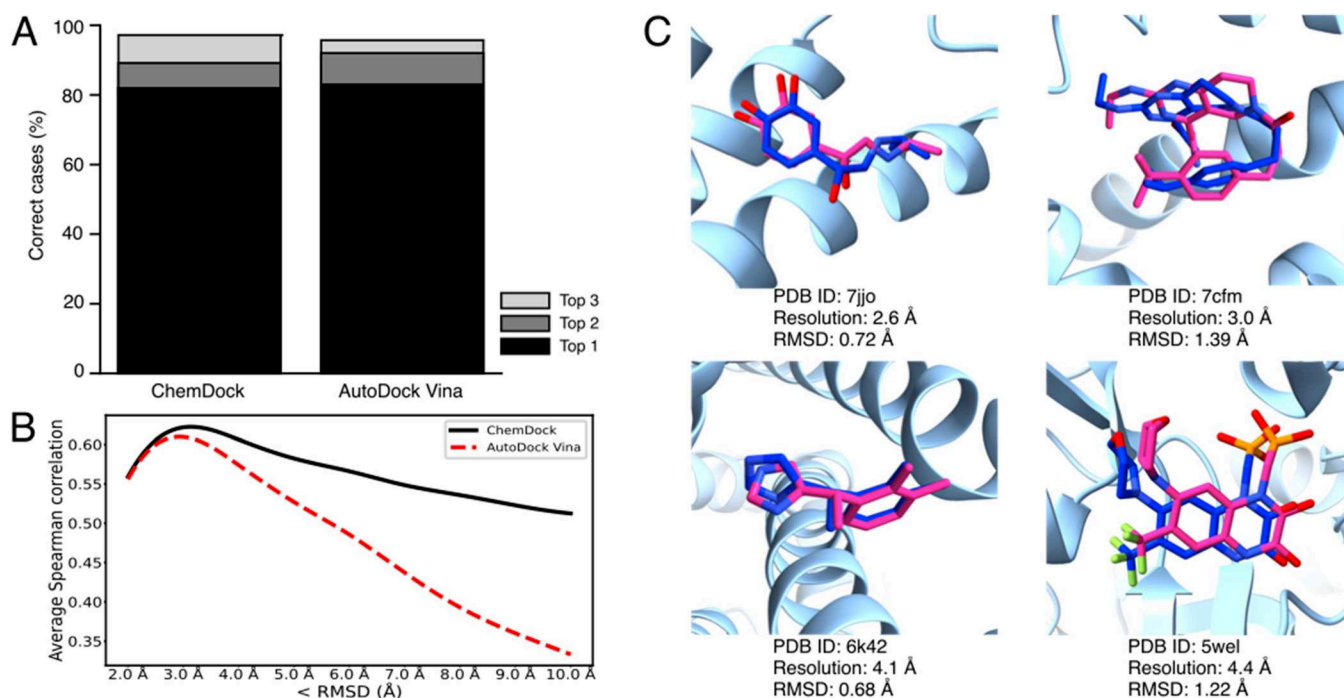


Figure 2. A comparison of the ChemDock score to the Autodock Vina docking score. (A) The percentage of cases where a conformation was ranked in the top 1 (black), 2 (dark gray), and 3 (light gray) positions by both scoring functions. (B) The binding funnel analysis for both the ChemDock (Black line) and AutoDock Vina (red dashed line) scores. The average Spearman correlation between the RMSD and either scoring function when only solutions <2, 3, 4, 5, 6, 7, 8, 9, or 10 Å RMSD were included in the RMSD-decoy benchmark. (C) ChemDock solutions (pink) from the high-resolution benchmark (PDB: 7JJO, 7CFM) and the low-resolution benchmark (PDB: 6K42, 5WEL). The PDB deposited solutions are also shown (dark blue) along with protein binding site elements (light blue). The resolutions and RMSD between solutions are also indicated.

parameters and the OpenForceField Sage 2.0.0²⁰ parameters for small molecules. This stage fine-tunes the initial fits generated in Stage 1, while simultaneously refining the fit of binding-site atoms.

Using the ChemEM Software. ChemEM is a software platform specifically designed to simplify and automate the process of fitting small molecules into cryoEM maps. To begin using the software, users must provide four key elements: a fitted protein structure in PDB format, a ligand file in SDF, Mol2, or SMILES string format, a cryoEM density map in MRC format, and the center point of the binding site to expedite the calculation of difference maps (Figure 1).

Following the provision of these inputs, the software's pipeline is initiated, requiring no further intervention from the user. However, in cases where certain structures necessitate more comprehensive assessment, ChemEM provides the flexibility for users to examine the preprocessed difference maps before progressing to the docking phase.

In complex scenarios, such as when two or more ligands are being fitted simultaneously to the same binding site, ChemEM allows users to manually assign disconnected densities to a specific ligand input (Figure S1). Furthermore, if high-quality difference maps cannot be generated using the default parameters, users have the option to adjust the number of protein atoms and the density of the map that are considered in the difference map calculations. This offers an effective method to optimize the map for fitting.

When the performance of the ChemEM software was benchmarked with the high-resolution data set, Stage 1 (Figure 1) showed an average processing time of 44.80 min (Figure S2). The average time of refinement per solution in Stage 2 was seen to be 13.97 min. Using the low-resolution

benchmark, the average processing time for Stage 1 of the algorithm was 52.62 min, and it took an average 5.95 min per solution for the Stage 2 refinement. For Stage 1, in both the high- and low-resolution cases, the time taken correlated with the number of rotatable bonds in the ligand (Figure S2).

Evaluation of the ChemDock Scoring Function. The density-independent part of the molecular docking scoring function (termed the ChemDock score) for the ChemEM scoring function included terms for hydrogen bonding, hydrophobic interactions, π - π stacks, and steric and repulsive terms (see Methods). The score was trained to estimate the pK_d of a given protein/ligand complex using 3,281 complexes taken from the PDBBind data set (Tables S1 and S2).

The correlation between the ChemDock or AutoDock Vina¹¹ scoring functions and binding affinity was assessed using a benchmark of 137 protein–ligand complexes, each associated with an experimentally determined binding affinity. The predicted binding affinities based on ChemEM had a Pearson correlation of 0.69 to the experimental binding affinities compared with 0.63 for AutoDock Vina.

We next looked at the ability of ChemDock to identify a *correct* conformation given a decoy set and rank decoys given their root-mean-square deviation (RMSD) from the *reference* structure (the PDB deposited structure). For each of the 137 protein–ligand complexes, up to 100 ligand decoys from CASF-2016 data set (with increasing RMSD values from the reference structure) were ranked by both ChemDock or AutoDock Vina scoring functions (Figure 2). ChemDock was able to rank a correct conformation (defined as ≤ 2.0 Å from the reference conformation) within the top 1, 2 and 3 ranked decoys in 81.88%, 89.13% and 97.1% of cases, respectively. The performance of ChemDock was comparable to that of

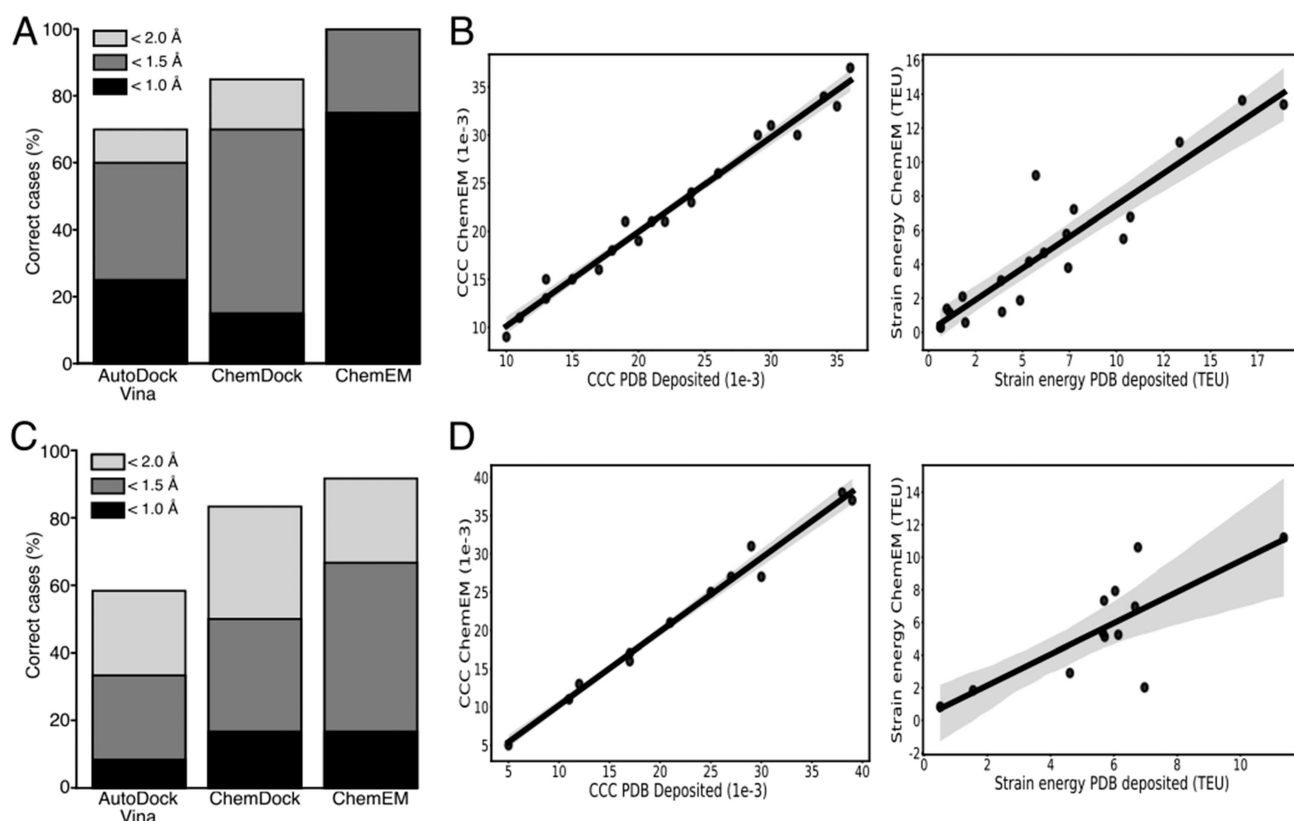


Figure 3. A statistical overview of docking the high- and low-resolution benchmarks with AutoDock Vina, ChemDock and ChemEM. (A) Percentage of cases from the high-resolution benchmark where docking/fitting produced a correct (<2 Å RMSD from the deposited solution), good (<1.5 Å), or excellent (<1.0 Å) solution. (B) Correlation between the CCC (left) and the strain energy (right) of the PDB deposited solutions with that of the final ChemEM refined solutions. (C) Percentage of cases from the low-resolution benchmark where docking and fitting produced a correct, good, or excellent solution. (D) Low-resolution benchmark correlations between PDB deposited and ChemEM refined solutions for the CCC (left) and strain energy (right).

Vina, where Vina ranked a correct conformation in the top third, second, or first ranks in 95.65, 92.03, and 83.33% cases, respectively (Figure 2A).

In order to further probe the ChemDock scoring function, the binding funnel analysis introduced in CASF-2016 report¹² (see Methods) was utilized. The analysis showed a good correlation between the ranking of conformations based on ChemDock and Vina scoring functions within a given RMSD threshold from the reference conformation, especially for conformations within <3.0 Å threshold (Figure 2B). However, the ChemDock correlation was more robust than that of Vina further away from the reference conformation (~ 6 – 10 Å).

Docking Small Molecules with the ChemDock Scoring Function. To run docking experiments, the ChemDock scoring function was integrated with an Ant-Colony Optimisation (ACO) algorithm.²¹ A “high-resolution” benchmark was constructed, consisting of 20 PDB protein–ligand complexes derived from cryo-EM experiments at 2.2–3.0 Å resolution (Table S3). All deposited ligands were removed from the structures and redocked starting from a random conformation in the same binding site using the ChemDock software. A correct conformation (<2 Å from the reference conformation) was generated 85% of the time, a good conformation (<1.5 Å) in 70% of cases, and excellent conformation (<1.0 Å) in 15% of cases (Figure 3A). When the same ligands were docked with Autodock Vina, a correct conformation was found in 70% of cases, good in 60%, and excellent in 25%. The quality of the final solutions was

analyzed using the conformational torsional strain energy²² (a measurement of the quality of conformations). Overall, the solutions generated with ChemDock were seen to be of better quality than those generated with Autodock Vina, with a mean minimum RMSD of the top solution (over the entire benchmark) and average torsional energy units (TEU) of 1.44 Å and 5.90 TEU compared to 1.93 Å and 6.53 TEU (Table S4).

To test the accuracy of the ChemDock software on a wider range of structures, docking experiments were repeated using 12 structures from lower-resolution cryo-EM maps (3.0–4.5 Å, Table S3). ChemDock generated a correct conformation in 83% of cases, while good and excellent solutions were found 50 and 16% of the time, respectively (Figure 3C). Autodock Vina generated a correct conformation in 58% of cases, with 33% being good and 8% excellent. The mean minimum RMSD from the deposited solution was 1.72 Å for ChemEM and 2.55 Å for Autodock Vina. However, the solutions generated by autodock Vina showed a better mean strain energy of 7.68 TEU compared to 8.32 TEU for ChemEM (Table S5).

Finally, we assessed the ability of ChemDock to rank generated ligand solutions. We found that the rank of the ChemDock solution with the lowest RMSD to the reference solution occurred on average within the top 15% of generated solutions for both high-resolution (Table S6) and low-resolution (Table S7) benchmarks.

Integrating ChemDock with Cryo-EM Density (ChemEM). We investigated the performance of two goodness-of-fit

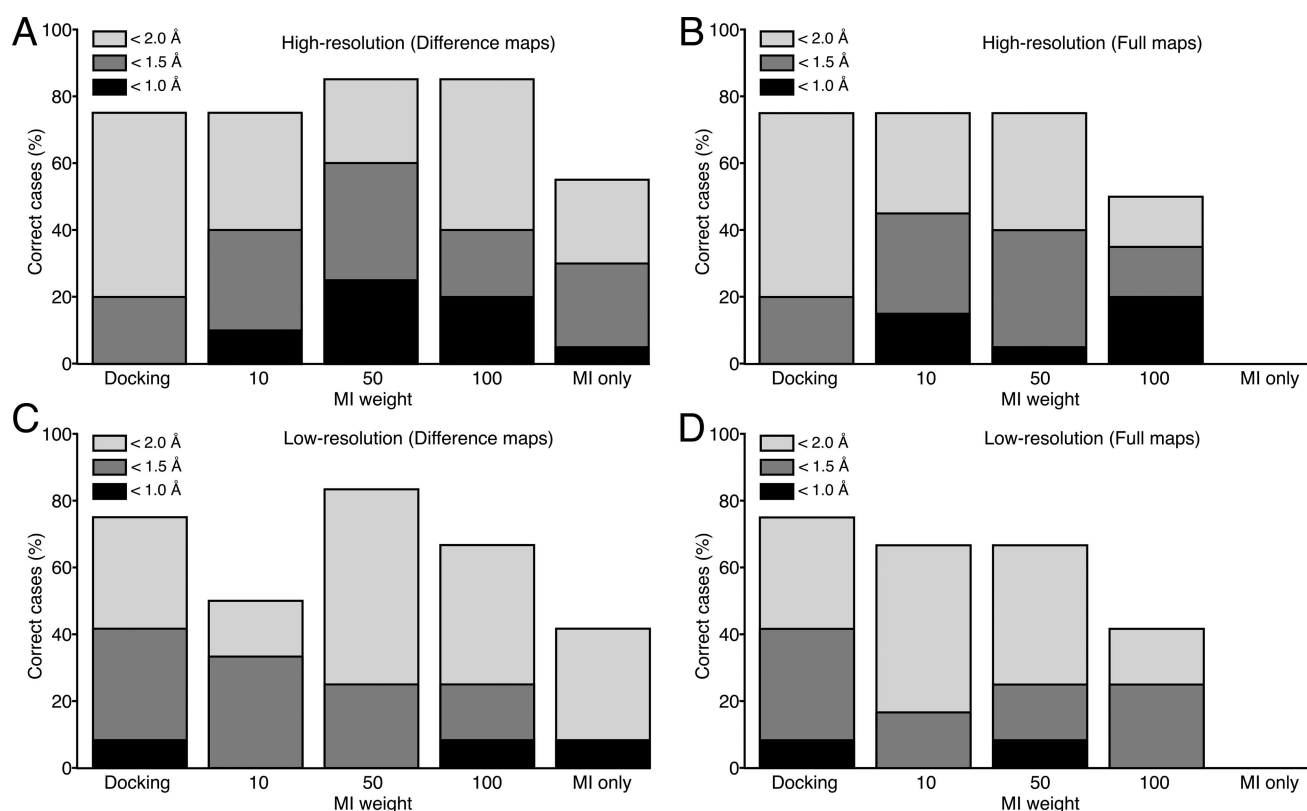


Figure 4. Results of fitting with the ACO and the ChemDock score (docking), MI alone, or the ChemEM score with MI score weights of 10, 50, and 100. (A) Number of cases where a correct solution was generated when the analysis was run for the high-resolution benchmark using density difference maps. (B) Correct cases when the high-resolution benchmark and full maps were used. (C) Number of cases where a correct solution was generated when the analysis was run for the low-resolution benchmark using density difference maps. (D) Correct cases determined using the full maps and low-resolution benchmark.

metrics, CCC and MI, for fitting small molecules by using a simulated benchmark. Density difference maps were simulated from the X-ray structure corresponding to 285 protein/ligand complexes from the CASF-2016 database¹² at a 2.5–8.5 Å resolution range (every 0.5 Å). For each case, the CASF-2016 decoy set was ranked by either MI or CCC.

At each resolution tested, for both MI and CCC, there was a strong negative correlation between the RMSD and the goodness-of-fit metric (Table S8); however, MI had a higher correlation. This difference was statistically significant in the 4.5–8.5 Å range. The Pearson correlation coefficient for the MI score remained relatively consistent for each resolution while decreasing for the CCC. This indicated that the MI score was more robust to changes in the resolutions of the difference density maps, inline with previous reports using the MI score to fit protein structures.¹³

Next, the ChemEM workflow was tested for fitting small molecules into the experimental cryo-EM maps mentioned above at both high- and low-resolutions. The first part of the workflow fits small molecules to the experimental difference density maps using the ACO algorithm.^{21,23} To identify an optimal weight to integrate the MI score with the molecular docking score, this step was performed without minimization.

The ChemDock score was calibrated to estimate pK_d , with values typically ranging from 0 to 10. To determine a representative range for the MI score, we calculated it for 3,204 solutions at various stages of the ChemEM algorithm. These solutions differed in RMSD from the ground truth structures. In practice, these MI score values generally lie

between 0 and 0.1 (Figure S3). As a result, we tested MI weights of 10, 50, and 100. These correspond to scaling the MI score to 10% (0.1 \times), 50% (0.5 \times), and 100% (1 \times) of the molecular docking score's magnitude, respectively.

Using the high-resolution benchmark showed the docking score alone identified a correct ligand conformation in 75% of cases, while the MI score alone identified a correct conformation in only 55% of cases (Figure 4A). The success rate of this initial fit rose to 85% when MI was integrated with the docking score at a weight of 50 (Figure 4A). This was accompanied by an increase in solution accuracy compared to using docking or MI scores alone, with 60% of cases generating a solution ≤ 1.5 Å and 25% within 1 Å of the deposited solutions (Figure 4A).

To ascertain whether the MI weight of 50 was applicable at resolutions >3.0 Å, the experiment was repeated using the lower resolution experimental benchmark. Once again a combination of the MI and docking score identified the most correct solutions (83%) compared to each score alone (75 and 42% correct cases, respectively; Figure 4C). However, only 25% were identified at an RMSD of ≤ 1.5 Å from the deposited ligand.

Assessing the rank of the ChemEM solution with the lowest RMSD to the reference solution showed that the solution with the lowest RMSD appeared in the top 5% of solutions on average for the high-resolution benchmark (Table S9) and the top 10% for the low-resolution benchmark (Table S10).

Finally, the ChemEM scoring function performed better using difference maps compared to full maps. For the high-

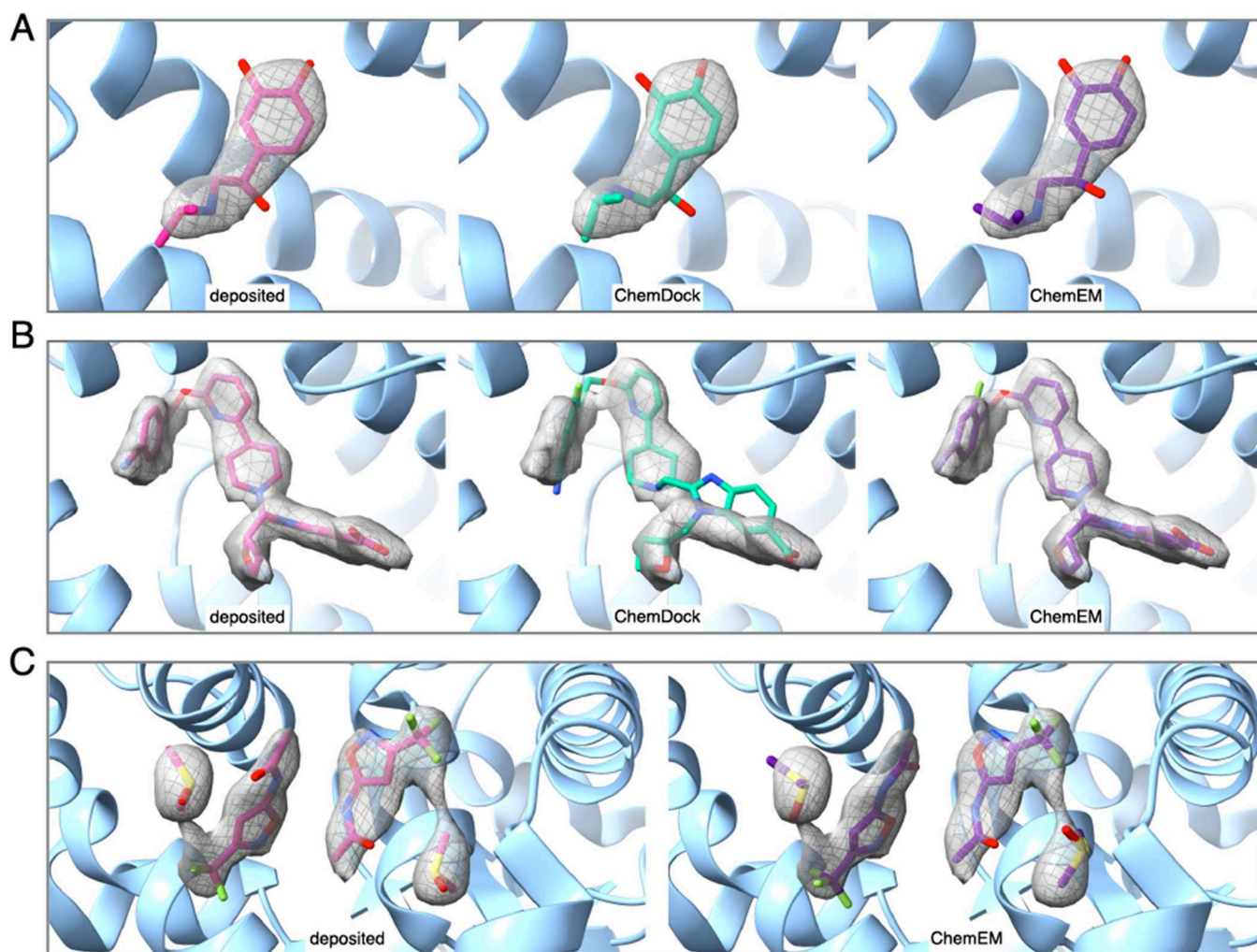


Figure 5. Specific examples of ChemDock (Docking) and ChemEM (refined) solutions for the high-resolution benchmark compared to the deposited solutions when (A) isoprenaline was fit to the β -Adrenergic receptor protein (PDB: 7JJO, EMD: 22357). (B) Small molecule PF06882961 bound to the human glucagon-like peptide-1 receptor (PDB: 6X1A, EMD: 21994) (C) Two fragment molecules and two molecules of dimethyl sulfoxide (DMSO) were fitted to the human pyruvate kinase protein (PDB: 6TTI, EMD: 10577).

resolution benchmark, no additional correct solutions were generated at any MI weight when using full maps compared to docking alone (Figure 4B). However, integrating MI did improve the solution quality, resulting in 45% good and 15% excellent solutions with an MI weight of 10 (Figure 4B). With lower resolutions, using full maps resulted in a lower number of correct solutions when integrating MI (compared to docking alone), with the best MI weights (10 and 50) resulting in only 66% of cases having a correct solution. Furthermore, the solutions were of a lower overall quality compared to that of docking alone (Figure 4D).

Flexible Fitting to Full Maps. For each case in the high- and low-resolution benchmarks, the best initial fits were generated using density difference maps and the ChemEM scoring function with an MI weight of 50. Following this molecular docking protocol, the solutions were further refined into the full map with a flexible fitting protocol (Methods, Figure 1).

In all 20 cases of the high-resolution benchmark, a good solution was generated, with an excellent solution found 75% of the time (Figure 3A). The CCC of the refined solutions (in the original full map) correlated well with that of the reference solutions, with a mean CCC of 0.023 ± 0.008 standard

deviation for the reference conformations and 0.022 ± 0.008 standard deviation for the refined solutions (Figure 3B). However, the quality of the conformations generated by ChemEM was notably better than that of the deposited solutions with mean strain energies of 4.86 and 6.47 TEU, respectively (Figure 3B).

The workflow was repeated by using the 12 low-resolution protein–ligand complexes. Following four rounds of flexible fitting, 92% of solutions were correct (Figure 3C). This was accompanied by an increase in the quality of solutions, with 67% good solutions and 17% excellent. The mean CCC was 0.022 ± 0.01 and 0.023 ± 0.01 for the refined and reference conformations, respectively (Figure 3D). The strain energies over the whole data set were inline with the published structures with a mean of 5.61 TEU compared to 5.64 and 5.99 TEU in the reference structures and high-resolution controls, respectively (Figure 3D and Table S5).

Finally, we assessed the sensitivity of the ChemEM score to varying levels of map sharpening using two distinct cases from the EMDB: a high-resolution example (PDB: 6X1A, EMD: 21994) and a low-resolution one (PDB: 6U8S, EMD: 20691), both of which had their half-maps deposited. We applied two widely used postprocessing methods to these maps: B-factor

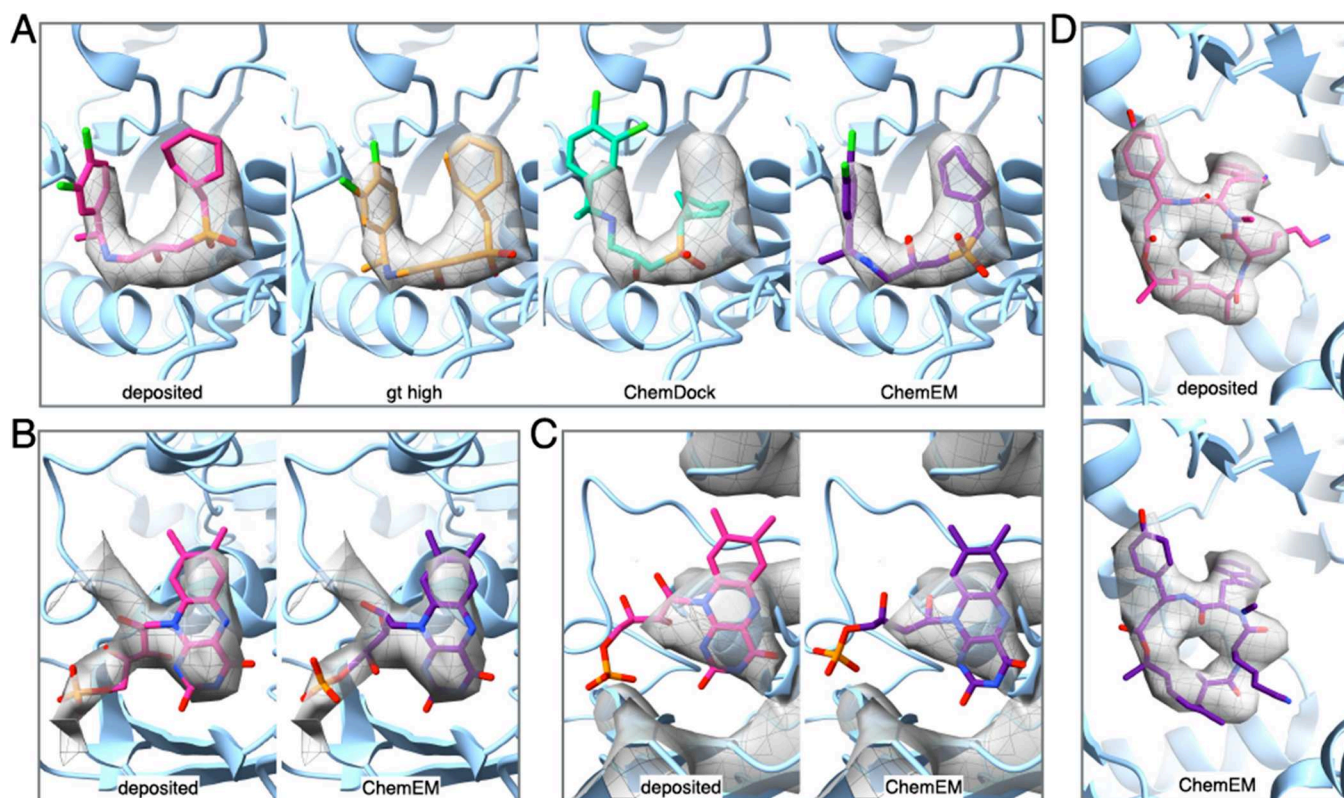


Figure 6. Specific examples of ChemDock and ChemEM solutions for the low-resolution benchmark compared to the deposited and high-resolution “ground truth” solutions (gt high) when the (A) molecule CGP 54626 was fit to Human GABA_B (PDB: 7CUM, EMD: 30472). (B) Molecule of Flavin mononucleotide was fit to bacterial respiratory complex I (PDB: 6ZIY, EMD: 11231) at a resolution of 4.25 Å. (C) Flavin mononucleotide molecule bound to mammalian respiratory complex I (PDB: 5LDX, EMD: 4041) at a resolution of 5.6 Å. (D) Jaspilakinolide was fit to actin (PDB: 6T24, EMD: 10366).

sharpening²⁴ in RELION²⁵ and sharpening-like feature enhancement with DeepEMhancer.²⁶ Subsequent recalculations of the ChemEM scores for solutions generated with the deposited maps revealed that for the high-resolution case, the recalculated scores displayed a Pearson correlation coefficient of 0.99 with the original scores for both postprocessing techniques (Figure S4). The low-resolution case demonstrated a slightly diminished, yet substantial, correlation of 0.91 for DeepEMhancer and 0.92 for B-factor sharpening (Figure S4). These results indicated that ChemEM may display some robustness against variations stemming from map sharpening or blurring techniques. However, while these results are promising, it is essential to note they represent a limited data set.

Analysis of Specific Examples. High-resolution benchmark (≤ 3.0)—As discussed above, the ChemEM software was able to find meaningful solutions for most of the 20 high-resolution structures that agreed well with the deposited ligands. In a few cases, using ChemEM led to only modest improvements in the RMSD from the reference conformation compared with docking alone. However, in most cases, it led to much-improved fits, with a mean CCC of 0.022 ± 0.008 for the best solutions, compared to 0.019 ± 0.007 using docking alone (Table S4). Additionally, the mean ligand strain energies decreased from 8.12 to 4.86 TEU (Table S4). For example, in the case of Isoprenaline bound to an avian β -Adrenergic receptor protein²⁷ (PDB: 7JJO, EMD: 22357), the docking solution had an RMSD of 0.72 Å compared to 0.64 Å when the MI score and flexible fitting protocol were included (Figure

5A). The similarity in quality of both solutions in the context of the cryo-EM density was confirmed by looking at the CCC and ligand map Q-scores²⁸ (a map-based metric for assessing the quality and accuracy of models fitted into cryo-EM maps), where the ChemDock solution had a CCC of 0.036 and a map Q-score of 0.722 compared to a CCC of 0.037 and map Q-score of 0.731 for the ChemEM solution.

However, in the case of PF06882961 bound to the human glucagon-like peptide-1 receptor²⁹ (PDB: 6X1A, EMD: 21994), docking with ChemDock failed to find a correct ligand conformation, with RMSD of 2.09 Å from the reference conformation for the best solution (Figure 5B), a CCC of 0.017, a map Q-score of 0.349, and a strain energy of 13.41 TEU, compared with a CCC, map Q-score, and strain energy of 0.032, 0.811, and 13.38 TEU for the reference structure. When the cryo-EM map was used for fitting, ChemEM was able to generate a correct solution with an RMSD, CCC, map Q-score and strain energy of 0.89 Å, 0.03, 0.768, and 11.18 TEU (Figure 5B).

Low-resolution benchmark (>3.0 – 4.5 Å)—For the 12 lower resolution complexes, the docking software with ChemDock score was also able to find meaningful solutions within the top-scored solutions that agreed well with the reference ligands. Similar to the high-resolution cases, using the full ChemEM protocol (MI and flexible fitting) generally led to improvement in the RMSD from the reference conformation. An example of the benefit of including the MI score was in the case of CGP 54626 bound to the GABA_B³⁰ receptor at 3.52 Å (PDB: 7CUM, EMD: 30472), where docking alone was not able to

yield a correct solution (Figure 6A). The full ChemEM protocol improved the fit significantly, resulting in a conformation close to the high-resolution reference conformation (1.38 Å) and markedly better fit to the density than the cryo-EM deposited structure (Figure 6A) (Qscore ChemEM: 0.526, Qscore deposited: 0.442).

The ChemEM workflow produced accurate fits, even at the lower end of the resolution range tested. One example was the case of Flavin mononucleotide bound to the bacterial respiratory complex I³¹ (PDB: 6ZIY, EMD: 11231) at 4.25 Å resolution. Following refinement, the CCC was 0.005, equal to that of the deposited solution, and a map Q-score of 0.427 was comparable to that of the deposited solution of 0.438. Additionally, the strain energy was better than the high-resolution control structure and comparable to the reference ligand (Figure 6B). The workflow was only systematically tested on lower resolution structures at 3.00–4.5 Å resolution range, as at worse resolutions, it was difficult to identify structures with ligands that had high-resolution control structures deposited in the PDB. However, one interesting case to test was a homolog of the above structure—the Flavin mononucleotide bound to the mammalian respiratory complex I³² at 5.6 Å resolution (PDB: 5LDX, EMD: 4041). At this resolution, it was still possible to fit the ligand in a correct conformation, although it was necessary to reduce the MI weight to 10. The final solution had an RMSD of 1.25 Å and correlated well with the conformation seen in the reference and high-resolution structures. The CCC and map Q-score of the ChemEM solution were 0.023 and 0.234, showing a similar fit to that map as the reference solution (CCC: 0.022, map Q-score: 0.241, Figure 6C). However, the ChemEM solution had a notably better strain energy than that of the reference solution (strain energy ChemEM: 8.06 TEU, strain energy reference: 14.72 TEU, Figure 6C). When the same ligand was fitted using only ChemDock, the RMSD of the best solution was 1.71 Å, and the CCC and TEU were 0.022 and 12.36, respectively, worse than the solution generated when the cryo-EM map was included.

As described above, the ChemEM methodology resulted in a solution similar to the reference conformations in all cases of the high-resolution benchmark and all but one of the low-resolution cases. This refers to the case of Jasplakinolide-bound actin³³ at 3.7 Å resolution (PDB: 6T24, EMD: 10366). Jasplakinolide is a large ligand with 47 heavy atoms, and although the aromatic moieties were accurately placed in the density, the large acyl-ring did not converge to the correct solution (Figure 6D).

Multiligand Fitting. One example within the high-resolution benchmark described the use of the human pyruvate kinase protein³⁴ (PDB: 6TTI, EMD: 10577) for fragment screening using cryo-EM. The binding site contained multiple ligands, with two “fragment” molecules and two molecules of dimethyl sulfoxide (DMSO).

All four ligands were fitted by using the ChemEM workflow (Figure 5C). The final conformations had an average RMSD of 1.38 Å from the reference structure, over all four ligands. The individual RMSDs of the fragment molecules were 0.38 and 1.17 Å. In comparison, the two DMSO molecules were less well fitted, with RMSDs of 1.88 and 2.09 Å.

The CCC across all four ligands was 0.016, slightly lower than that of the reference (CCC: 0.019). However, for the two fragment molecules, the CCCs were the same (0.011 and 0.009), and the map Q-scores were comparable (ChemEM:

0.745 and 0.686, reference ligands: 0.767 and 0.792). Furthermore, the torsional strain energies of both the refined fragments and reference fragments were comparable, with TEUs of 1.14 and 1.89, and 1.12 and 1.17, respectively. The fit of the DMSO molecules was less good with CCCs of 0.006 and 0.005, compared to 0.01 and 0.09 in the reference structure; this was confirmed by comparing map Q-scores with scores of 0.554 and 0.513 in the ChemEM solution compared to 0.725 and 0.757 in the reference structure. However, the two molecules were still placed in the correct region in the map.

DISCUSSION

There is currently a requirement for a diverse range of software to accurately and automatically fit small molecules to cryo-EM structures due to significant developments in this technique over the past decade, making it increasingly relevant to drug discovery. ChemEM is a novel method that meets these needs by combining small-molecule docking with density fitting by using difference maps. The method, which can be used with or without the cryo-EM data (as a “classical” docking method), was assessed on a benchmark of cryo-EM protein–ligand complexes at 2.2–4.5 Å resolution with corresponding X-ray structures. This resolution range is consistent with what is currently required in the field, with 22.5% of maps deposited within the EMDB in 2022 at resolutions better than 3.0 Å and 47.8% between 3 and 4 Å range. Additionally, ChemEM was able to produce accurate fits for an individual test case at a lower resolution (5.6 Å).

The ChemEM ChemDock scoring function was derived to take account of the physicochemical features of protein–ligand interactions. It was shown to outperform AutoDock Vina scoring function when estimating the relative affinity of compounds and had a comparable performance when identifying a correct ligand conformation from an incorrect one (Figure 2A). Importantly, ChemDock had improved performance in the binding funnel analysis over AutoDock Vina (Figure 2B). It has been suggested that docking scores that can perform well on this task are good candidates for “blind” docking scoring functions.¹²

When the ChemDock score was combined with a search algorithm, a correct conformation was generated 85 and 83% of the time for the benchmarks derived from high- and low-resolution structures, respectively, a marked improvement compared to AutoDock Vina (Figure 3). Additionally, when both scoring functions failed to reach a correct solution, the solutions generated by ChemDock were generally closer to the reference (Tables S4 and S5), which was useful when ChemDock was combined with the cryo-EM data.

To achieve this, the cryo-EM density difference map was used and the MI score was added to the ChemDock score (ChemEM). It has previously been shown that MI-based scores are generally more discriminatory than CCC-based scores for fitting proteins to cryo-EM maps.^{35,36} Our results support these findings, showing that when fitting small molecules to density maps at resolutions of 2.5–8.5 Å, MI has a higher discriminatory power than CCC (Table S8).

When we used difference maps with ChemEM in stage 1 of the algorithm (Figure 1), there were additional correct solutions generated compared with those generated when using full maps (Figure 4). These solutions were also generally closer in RMSD to the ground truth. This improved accuracy is likely due to difference maps emphasizing the most relevant

density for the search algorithm. Although we have previously used cryo-EM difference maps for ligand docking,^{14,15} a systematic comparison with using full-density maps has not been undertaken. Our study shows that the advantage of using difference maps at 2.2–3.0 Å resolution is reduced when fitting ligands in lower-resolution densities (3.0–4.5 Å), and this effect holds true with full maps. This is due to an increased search space, leading to more false positive solutions being identified.

At lower resolution, our investigation indicated that more onus should be placed on the physicochemical properties rather than the goodness-of-fit to derive a good solution. This was exemplified when the ChemEM score was used for fitting without refinement: in the high-resolution benchmark, a greater MI weight of 100 resulted in the same number of correct solutions identified with a weight of 50 (Figure 4). Whereas, for the lower resolution benchmark when the contribution of the MI score was increased to 100, the number of correctly identified solutions decreased (Figure 4). Naturally, this was a consequence of less information being present in the map. However, the fact that this effect starts to occur at resolutions close to 3.0 Å is somewhat surprising.

Once initial fits are identified with ChemEM (Stage 1, Figure 1), the addition of refinement in the density in combination with the OpenFF Sage v2.0.0 force field (Stage 2, Figure 1) improved the quality (TEU, RMSD < 1.5/1.0 Å) and success rate (RMSD < 2.0 Å) of solutions (Figure 3A,C). This improvement was most pronounced for the high-resolution benchmark (Table S4). However, even in the low-resolution benchmark, only one case failed to identify the correct solution (Figure 6D). In one low-resolution instance (PDB: 7CUM, EMD: 3047226) where the deposited cryo-EM structure appeared to be incorrect (Figure 6A), ChemEM identified a solution more aligned with the high-resolution control structure (PDB: 4MR7). This solution not only had closer RMSD proximity to the high-resolution control but also boasted a higher map Q-score compared to that of the deposited cryo-EM solution. Additionally, as seen for generating the initial fits with ChemEM, the best results after refinement for the lower resolution benchmark were achieved when the contribution of the map component within the score was lower in comparison to the high-resolution cases (Table S4). One important point to note is that when fitting into maps at resolutions worse than 3.0 Å, having less information present in the map will affect not only the accuracy of the ligand placement in the map but also of the protein side-chain placement. This in turn may affect the density-independent part of the scoring function as misplaced side chains may lead to improper bonds being scored between the protein and ligand. The flexible fitting stage (Stage 2, Figure 1) of the ChemEM workflow aims to minimize these errors by allowing the protein side chains to move during refinement.

Furthermore, adding the density not only increased the number of solutions where a correct case was generated but also improved their rank. This explains why there was a greater improvement in the ranking of the best solutions in the high-resolution benchmark (Tables S6 and S9) over the lower-resolution benchmark (Tables S7 and S10).

In addition to the resolution of the map, the quality of the generated solutions (as assessed by RMSD to the reference solution) is limited by the number of rotatable bonds contained within the ligand (Figure S5). As the number of rotatable bonds increases, the RMSD from the reference

increases. This effect was seen in both the high- and low-resolution benchmarks and was found to be independent of the resolution. It is most likely a consequence of the increased search space needed to evaluate solutions with a larger number of rotatable bonds. This issue is somewhat taken into consideration within the ChemEM workflow, as the number of iterations of the ACO algorithm will increase as a function of the number of rotatable bonds to account for the greater search space needed.

As explained above, our benchmark only included cases better than 4.5 Å resolution except for a single case at 5.6 Å resolution (Figure 6C), where it was possible to use ChemEM to dock a Flavin mononucleotide to the mammalian respiratory complex I, yielding a solution that fits the map better than the PDB-deposited solution.³² While fewer structures at this resolution contain small molecules, ~14% of the total EMDB deposited maps in 2022 were between 4 and 6 Å resolution. At this resolution range, a good fit could only be achieved by reducing the MI weight from 50 to 10 in the ChemEM score. However, including the map with a low weight was still helpful. This may indicate that at relatively low resolutions, there may still be some benefit to including the map in docking small molecules, even though more weight is needed to be placed on force field scores.

Finally, the ChemEM software was extended for the automated simultaneous fitting of multiple ligands (Figure 5), which was computationally feasible, owing to the use of difference maps. This feature was shown to work relatively well, with 3 of 4 ligands fitted correctly. Future work would extend this feature for the parallel fitting of other binding site components, for example, ions.

EXPERIMENTAL SECTION

Computational Data Sets. Data to train the ChemDock scoring function were taken from the PDBBind-v2020 refined data set.³⁷ This data set consists of 5316 protein–ligand atomic models derived from X-ray crystallography at resolutions better than 2.5 Å. PDB files that failed to load into the software or those that formed part of the CASF-2016 data set were removed from the training set, leaving 3281 complexes in the final training set.

The data set for testing the scoring function was taken from the CASF-2016 data set consisting of 285 protein–ligand atomic models from X-ray crystallography experiments, at resolutions better than 2.5 Å.¹² When comparing the ChemDock scoring function to the Autodock Vina scoring function any files that failed to load into the ChemDock software or the Open Drug Discovery Toolkit (ODDT) software³⁸ (used for Autodock Vina rescoring) were discarded from the data set. For the remaining ligands in the test set, the Tanimoto coefficient was calculated in comparison to those in the training set using RDKit. When ligands in the test set exhibited high similarity (>0.85 Tanimoto coefficient) to those in the training set, the protein similarity was assessed using Clustal Omega.³⁹ If the protein's percent identity exceeded 70%, the corresponding protein–ligand complex was excluded from the test set. The final data set contained 137 protein–ligand complexes.

The investigations of the CCC and MI goodness-of-fit scores used all 285 protein models in the CASF-2016 data set.¹² Additionally, for each case in the CASF-2016 data set, up to 100 ligand decoys were included. These decoys were ligand conformations obtained from molecular docking software that had RMSDs distributed from 0.0 to 10.0 Å from the deposited conformation.¹² Simulated maps were generated from the deposited protein and ligand models in USCF-Chimera⁴⁰ using the “molmap” command. For each complex, maps were generated at resolutions from 2.5 to 8.5 Å in 0.5 Å increments. Simulated density difference maps were generated between the full map and the deposited protein model (without the ligand) using a

local density difference mapping methodology¹⁸ implemented in TEMPY2.⁴¹

Two experimental data sets were generated, one containing high-resolution structures (≤ 3.0 Å, “high-resolution” benchmark) and a second containing lower resolution structures (3.0 Å < resolution ≤ 4.5 Å, “low-resolution” benchmark). For the high-resolution benchmark, 20 protein–ligand complexes from the Protein Data Bank⁴² (PDB) that had an associated Cryo-EM map in the Electron Microscopy Databank¹ (EMDB) were selected. Similarly, 12 protein–ligand complexes were selected for the low-resolution benchmark; however, for the low-resolution structures to be included a high-resolution control (≤ 3.0 Å, representing the ground truth structure for experiments conducted with this data set) structure of the same protein–ligand complex had to be available in the PDB. Experimental density difference maps were generated using the deposited maps and protein models for each case, using the same methodology used for the simulated maps.

The reason we chose not to systematically test cases at worse resolution was that we wanted to have a high-resolution control structure for each low-resolution case (3 – 4.5 Å) and could not find cases beyond this resolution range. However, one case above this resolution was chosen where a control structure was available (PDB: SLDX, EMD: 4041).³²

ChemDock Scoring Function. The density-independent docking scoring function was designed to predict the pK_d of a given protein–ligand complex. The score contained terms for atom–atom attraction/repulsion, hydrogen bonding, π – π stacking, hydrophobic interactions, and ligand intramolecular interactions (eq 1):

$$\begin{aligned} \text{ChemDock} = & w_1 \text{VG}_1 + w_2 \text{VG}_2 + w_3 \text{VS} + w_4 \text{HB}_{\text{dist}} + w_5 \text{HB}_{\text{angle}} \\ & + w_6 \text{HPI} + w_7 \text{Arm}_{\text{dist}} + w_8 \text{Arm}_{\text{angle}} + w_9 \text{ligand}_{\text{vg1}} \\ & + w_{10} \text{ligand}_{\text{vg2}} + w_{11} \text{ligand}_{\text{vs}} + \text{bias} \end{aligned} \quad (1)$$

Here, we keep the general term for scoring atom–atom steric interactions for Vina.¹¹ This score is the combination of two Gaussian functions (eq 1, VG_1 , VG_2) and a repulsion function (eq 1, VS). This function was also used to score the intramolecular interactions between nonconsecutive atoms of the ligand (eq 1, ligand-VG_1 , ligand-VG_2 , ligand-VS).

Hydrogen bonding was accounted for using two geometric parameters H-bond–Donor–Acceptor distance (eq 2, HB_{dist}) and H-bond–Donor–Hydrogen–Acceptor angle (eq 3, HB_{angle}). The distance score was a function of the van der Waals overlap of the hydrogen-bonded atoms.

$$\text{HB}_{\text{dist}}(d) = \begin{cases} -1 & \text{if } d \leq -0.7, \\ 0 & \text{if } d > 0, \\ \frac{d}{-0.7} - 1 & \text{else} \end{cases} \quad (2)$$

where $d = d_{ij} - r_i - r_j$; where, d_{ij} is the distance between atomic centers of protein atom i and ligand atom j , and r_i/r_j is the van der Waals radii of atom i or j , respectively.

The term used to describe the hydrogen bond angle was a linear interpolation between 0 and 1 for hydrogen bond angles between 90° and 180° (eq 3).

$$\text{HB}_{\text{angle}}(\theta) = \begin{cases} 0 & \text{if } \theta < 90, \\ \frac{\theta - 90}{90} & \text{else} \end{cases} \quad (3)$$

To score hydrophobic interactions the hydrophobic matching algorithm was used, first outlined in SCORE.⁴³ This scored the local environment for a given hydrophobic atom. Hydrophobic atoms were defined as any carbon atoms that were bound exclusively to any carbon atom or hydrogen.

The environment that a hydrophobic atom is placed in was scored as the sum of the $\log P$ (a measure of hydrophobicity) of all atoms within 6.0 Å of the hydrophobic atom, as in X-CSCORE⁴⁴ and SCORE.⁴³ Values for $\log P$ were taken from XlogP3;⁴⁵ however, unlike X-CSCORE and SCORE, the hydrophobic score (eq 4, HPI) was the sum of the $\log P$ values for each hydrophobic atom in the ligand (as

opposed to a binary 1 or 0 score if the hydrophobic atom is or is not within a local hydrophobic environment).

$$\text{HPI}(j) = \sum_{i=1}^n \log p(i) \quad (4)$$

where j is a given hydrophobic atom and n is the set of atoms within 6.0 Å of atom j .

To investigate the preferred geometry of π – π stacking interactions, the Tough-D1 data set⁴⁶ was used. This data set contained 3079 protein–ligand complexes where the ligand was stabilized by at least one aromatic ring. For all protein–ligand aromatic π – π stacking interactions, data regarding the angle between interacting aromatic rings and the distance between ring centers was extracted. The data were split into two sets, one set containing distance and angle data for π – π P-stacks (angle $> 45^\circ$) and one for π – π T-stacks (angle $< 45^\circ$). From this data, the aim was to estimate the probability of a π – π stacking interaction being true given the ring-plane angle and ring center–ring center distance values.

To do this, probability density functions were derived by fitting the data to 106 common distributions with the Python package Fitter. Each fit was scored by using the residual sum-of-squares criteria. The distributions that best describe the angle data for both P-stack and T-stack data sets were seen to be beta normal distributions, with sum of the square residuals of 0.0037 and 0.0023, respectively. For the distance data in the P-stack set, an exponential normal distribution was seen to fit the data best with the sum of the squared residuals being 0.29, while for the T-stack distance data, a skewed normal distribution was seen to fit the data best, where the sum of the squared residuals was seen to be 0.13.

For inclusion in the scoring function, the aromatic score was defined as follows:

$$\text{Arm}_{\text{angle}}(\theta, \alpha, \beta) = \frac{\Gamma(\alpha + \beta) \theta^{\alpha-1} (1 - \theta)^{\beta-1}}{\Gamma(\alpha) \Gamma(\beta)} \quad (5)$$

where θ is the angle between two planes, α and β are constants set based on the angle value, and Γ is the gamma function: $\Gamma(n) = (n - 1)!$

The score for the aromatic distance was dependent on the plane–plane angle. If the angle was lower than or equal to 45° , the interaction was scored using the equation:

$$\text{Arm}_{\text{dist}}(d) = \frac{1}{2/k} \exp\left(\frac{1}{2/k^2} - d/k\right) \text{erfc}\left(-\frac{d - 1/k}{\sqrt{2}}\right) \quad (6)$$

where d was the ring center–ring center distance, k is a constant, and erfc was the complementary error function. If the angle was greater than 45° , the aromatic distance is scored as

$$\text{Arm}_{\text{dist}}(d) = \frac{\exp\left(\frac{-(d - \mu)^2}{2\sigma^2}\right) \text{erfc}\left(\frac{-a(d - \mu)}{\sqrt{2}\sigma}\right)}{\sqrt{2\pi}\sigma} \quad (7)$$

Where d was the ring center–ring center distance, σ and μ were constants, and erfc is the complementary error function.

The final ChemDock scoring function was a summation of the weighted terms of all atom–atom pairs within the binding site (eq 1). Additionally, a bias term (bias) was included to bring the magnitude of the score in line with the expected pK_d values.

Fitting the Weights in ChemDock Scoring Function. Each term in the scoring function had to be fitted to a weight. This fitting was done using a linear regression to the experimentally determined pK_d values from the PDBBind-v2020 data set.³⁷

A 5-fold validation methodology was used for the regression. Briefly, the individual scoring terms were precalculated with the ChemEM software for each protein–ligand complex in the final PDBBind-v2020 refined data set.³⁷ The order of the cases in the data set was randomized and split into 5 approximately equal sets of each containing 609 complexes. The weights were fit five times using four of the sets, leaving a unique set each time for validating the parameters. The weights for the scores were determined by linear

regression to the experimentally determined pK_d values (Table S1). Following the linear regression, the weights for each scoring term were averaged from the values in each round (Table S2).

ChemEM Scoring Function. When fitting into density maps, the MI score was included in the calculation. When combined with the density independent part of the scoring function, the MI score was weighted by a constant, w_{MI} (eq 8).

$$\text{ChemEM} = (w_{MI}MI) + \text{ChemDock} \quad (8)$$

To identify the weight for the MI score in the ChemEM scoring function, docking of both the high- and low-resolution benchmarks was conducted in the absence of any further refinement. The weight that gave the best results at both high and low resolution was seen to be a weight that scaled the MI score to approximately 0.5 \times the magnitude of the density independent part of the scoring function (a value of 50, Figure 4). This value was used for all of the further experiments with the full ChemEM scoring function.

ChemDock Scoring Evaluation. To evaluate the ChemDock and AutoDock Vina scoring functions, the 285 protein–ligand complexes in the CASF-2016 benchmark set¹² were used. For scoring with ChemDock, the supplied protein PDB files were uploaded into the ChemEM software along with the native ligand and each ligand in the associated RMSD benchmark for each case (in MOL2 format). For each protein, all benchmark ligand decoys (RMSD benchmark) were scored individually using the ChemDock scoring function with no additional docking applied. To score protein–ligand complexes with AutoDock Vina, the Open Drug Discovery Toolkit (ODDT)³⁸ and an “in house” python script were used. The ODDT was used as AutoDock Vina does not currently support scoring without docking. Of the 285 cases, 132 failed to load into the ODDT leaving 153 for evaluation.

To evaluate the program’s ability to rank correct ligands, the RMSD benchmark was used. For both ChemDock and AutoDock Vina, the solutions in the RMSD benchmark were scored and ranked accordingly. The number of occurrences of either scoring function ranking a correct ligand (defined as <2.0 Å from the deposited conformation) at position 1, 2, or 3 was calculated.

For the binding funnel analysis, the RMSD benchmark was split into subsets containing all the ligands for that case that had an RMSD below a certain cutoff. The cutoffs were slowly raised in steps of 1.0 Å from 2.0 to 10.0 Å. At each cutoff, for each case, the subsets were ranked by either the AutoDock Vina or ChemDock score and their respective RMSDs, and the Spearman correlation was calculated.

To evaluate the correlation between the predicted and experimental values, a Pearson correlation coefficient was calculated between the ChemDock or AutoDock Vina scores for the experimental protein/ligand complex and experimental affinity values supplied with the CASF-2016 benchmark set.

All correlation coefficients were calculated using the SciPy python software and an “in-house” python script.

Molecular Docking Search Algorithm. ChemEM and ChemDock used a min–max Ant Colony Optimisation (ACO) Algorithm combined with the Nelder–Mead local minimizer, as in PLANTS.²¹ Briefly, for a given ligand a table of probabilities was constructed corresponding to the position of the ligand with 6-degrees of freedom (x, y, z rotation and translation) and n degrees of freedom for each rotatable bond. A ligand is constructed based on the values in the table and locally minimized with the Nelder–Mead algorithm. The probabilities evolve as the algorithm searches through various solutions with values close to high scoring conformations given a higher chance of being picked. The number of iterations of the search algorithm was proportional to the number of rotatable bonds within a ligand as in.²¹ When ChemDock was run for molecular docking, the output solutions of the ACO were energy minimized within the binding site using OpenMM.⁴⁷ During this step, all side-chains at the surface of the pocket were treated as flexible. Protein parameters were taken from the AMBER FF14SB¹⁹ with small-molecule parameters from the OpenFF Sage 2.0.0.²⁰ Minimization was conducted with the “minimizeEnergy” function in OpenMM with the tolerance set to 10 kJ/mol and allowed to run until convergence. For fitting small

molecules in cryo-EM density maps, the algorithm was sped up by segmenting the supplied density map around the center of the binding site, into a cubic map containing all voxels within a specified distance (30 Å by default). This segmented map was used to generate the difference maps for molecular docking (see above).

Docking solutions were further refined by using simulated annealing into the segmented full map. Once again, both the AMBER FF14SB¹⁹ and Sage OpenFF 2.0.0²⁰ were used. Additionally, the full map was added to the potential (eq 9) biasing atom movements toward areas of high density. The weight of the map potential was controlled by a constant k . Here, experiments were conducted using three values of k , 25, 50, and 75 (Table S11).

$$E^{\text{map}} = -kD(x, y, z) \quad (9)$$

where $D(x,y,z)$ is the normalized density value at position x,y,z within the full map.

Following minimization, the system’s temperature was gradually raised from 0 to 300 K in 1 K increments. At each increment, the simulation was equilibrated for 1000 steps.

Once the desired temperature was reached, the system was heated from 300 to 315 K by temperature steps of 1K. Again, at each temperature step, the system was allowed to advance 1000 steps (with a step size of 1 fs), before being cooled to 300 K in the same manner. This continued for a defined amount of cycles (4 by default) before yielding the final solutions.

For both minimization and simulated annealing atoms not within binding sites or forming part of the protein backbones were not minimized/refined. These atoms were kept in place using a harmonic pinning potential whose strength (k_{pin}), was 500 kJ/mol.

The final solutions were grouped by their RMSD (<2.0 Å), where each group was represented by the solution with the highest ChemEM/ChemDock score.

Molecular Docking Experiments and Small-Molecule Fitting. For molecular docking experiments, SMILES strings for the 32 ligands in the benchmark were extracted from the PDB chemical repository. Protein models were preprocessed by removing solvent and ligands. The binding site was defined from the native structure as the centroid of the deposited ligand, a radius of 12.5 Å was used for ChemDock experiments, and a box size of 12.5 Å was used for docking with AutoDock Vina. When using the ChemEM software, protein models were input as preprocessed PDB files, while ligands were supplied as a SMILES string. For AutoDock Vina, preprocessed protein files were prepared using the “DockPrep” command in UCSF Chimera. Ligands for AutoDock Vina were prepared by first generating a 3D conformation in RDKit using SMILE strings. These ligands were then prepared into PDBQT files using the Chimera “DockPrep”.

For multiligand docking experiments, the ChemEM software was extended to receive multiple inputs within a single binding site. Ligands were given as individual SMILES strings. Regarding the difference map inputs and center of binding sites, there are two options; the first was to give a difference map and binding site centroid as an input. The second option (used for experiments reported here) is to give multiple difference map inputs and binding site centroids. The search space for each ligand was then defined from the assigned difference density, with the center of the ligand binding site for a specific ligand being the center point of the assigned difference density. To fit multiple ligands into the binding site simultaneously, the scoring function was extended such that interligand interactions were accounted for in the same manner as protein–ligand interactions when using the ChemDock score.

To perform multiligand docking, first ligand density within the difference map was assigned to a given ligand and disconnected difference densities were extracted from the whole difference map using a map processing function contained in ChemEM (Figure S1A). Briefly, this command grouped regions of density together, where the boundaries were defined as voxels with any adjacent voxel below a given threshold (0.0 by default). The densities assumed to correspond to individual ligands were extracted from the whole map and all voxels in a given disconnected density were vectorised and the centroid for

the binding site for that ligand was determined by the centroid of the vectorised points (Figure S1B).

To evaluate the solutions given by the docking/fitting programs the RMSD to the reference structure was calculated using the symmetry-corrected RMSD implemented in the python package SpyRMSD.⁴⁸ For each experiment, the RMSDs given represented the mean minimum RMSD, and this was defined as the average of the solutions for each case with the lowest RMSD to the reference structure.

Additionally, the strain energy of ligand conformations was calculated using the method outlined in Gu et al.²² Calculation of the cross-correlation coefficient was conducted using the python package TEMPy2,⁴¹ using full maps downloaded from the electron microscopy data bank.¹

Computational Hardware. All analyses were conducted on a system equipped with an Intel(R) Xeon(R) Gold 6248R CPU (96 cores, 3.00 GHz base frequency), 196GB RAM, 512GB SSD, and an NVIDIA Quadro RTX 6000 GPU.

■ ASSOCIATED CONTENT

Data Availability Statement

The ChemEM software is available at: chemem.topf-group.com

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jmedchem.3c01134>.

Difference mapping procedure for multiple ligand fitting; performance of the ChemEM algorithm; histogram of MI scores calculated for 3,204 structures from various stages of the ChemEM algorithm; correlation of ChemEM scores with different sharpening methods versus using the deposited map; fivefold validation results from training the ChemDock scoring function; scoring function weights for each term in the ChemDock scoring function; rank data for ChemDock solutions with the best RMSD in the high- and low-resolution benchmarks; and average Pearson correlation coefficients between the MI/CCC of simulated difference maps and RMSD of the CASF-2016 decoy sets (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Maya Topf – Leibniz Institute of Virology (LIV), Hamburg 20251, Germany; Centre for Structural Systems Biology, Hamburg 22607, Germany; Universitätsklinikum Hamburg Eppendorf (UKE), Hamburg 20246, Germany; orcid.org/0000-0002-8185-1215; Email: maya.topf@leibniz-liv.de

Authors

Aaron Sweeney – Leibniz Institute of Virology (LIV), Hamburg 20251, Germany; Centre for Structural Systems Biology, Hamburg 22607, Germany; Universitätsklinikum Hamburg Eppendorf (UKE), Hamburg 20246, Germany
Thomas Mulvaney – Leibniz Institute of Virology (LIV), Hamburg 20251, Germany; Centre for Structural Systems Biology, Hamburg 22607, Germany; Universitätsklinikum Hamburg Eppendorf (UKE), Hamburg 20246, Germany
Mauro Maiorca – Leibniz Institute of Virology (LIV), Hamburg 20251, Germany; Centre for Structural Systems Biology, Hamburg 22607, Germany; Universitätsklinikum Hamburg Eppendorf (UKE), Hamburg 20246, Germany

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jmedchem.3c01134>

Funding

This research was supported by the cooperation of Leibniz Institute of Virology (as part of Leibniz ScienceCampus InterACT, funded by the BWFG Hamburg and the Leibniz Association) and Wellcome Trust grant 209250/Z/17/Z.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Ronja Markworth, Dr. Tristan Cragolini, and the Topf lab for helpful discussions.

■ ABBREVIATIONS

ACO, ant colony optimizer; AMBER FF14SB, assisted model building with energy refinement force field 2014 SB; CASF-2016, Critical Assessment of Scoring Functions 2016; CCC, cross correlation coefficient; Cryo-EM, cryogenic electron microscopy; DMSO, dimethyl sulfoxide; EM, electron microscopy; EMD, electron microscopy databank identification number; EMDb, electron microscopy databank; erfc, complementary error function; GABA_A, gamma-aminobutyric acid type A; GABA_B, gamma-aminobutyric acid type B; GOLD, genetic optimization for ligand docking; Gt high, high-resolution ground truth solutions; HADDOCK, high ambiguity driven biomolecular docking; HB_{angle}, scoring term for hydrogen bond angle geometry; HB_{dist}, scoring term for hydrogen bond distance geometry; Hbond, hydrogen bond; HPI, scoring term for hydrophobic interactions; Ligand-VG¹/Ligand-VG², scoring term using the AutoDock Vina Gaussian 1/2 to score ligand atom-ligand atom contacts; Ligand-VS, scoring term using the AutoDock Vina steric score to assess repulsion between ligand atom-ligand atom contacts; LogP, calculated logarithm of the partition coefficient of a molecule in a solution of octanol and water; MI, mutual information; ODDT, open drug discovery toolkit; OpenFF, open force field; OpenMM, open molecular mechanics; PDB, protein data bank; pK_d, the negative logarithm of the dissociation constant of a ligand–receptor complex; PLANTS, protein–ligand ant system molecular docking software; RELION, regularized likelihood optimization; RMSD, root mean square deviation; SDF, structure data file; SMILES, simplified molecular input line entry system; STD, standard deviations; TEMPy2, toolkit for the analysis of macromolecular electron microscopy maps in Python version 2; TEU, torsional energy units represent deviation from preferred ligand torsion angles; VG¹/VG², scoring term using the AutoDock Vina Gaussian 1/2 to score ligand atom-protein atom interactions; VS, scoring term using the AutoDock steric term to score ligand atom-ligand atom repulsion; w_{MI}, the weight used to scale the MI component of the ChemEM score

■ REFERENCES

- (1) Lawson, C. L.; Patwardhan, A.; Baker, M. L.; Hryc, C.; Garcia, E. S.; Hudson, B. P.; Lagerstedt, I.; Ludtke, S. J.; Pintilie, G.; Sala, R.; Westbrook, J. D.; Berman, H. M.; Kleywegt, G. J.; Chiu, W. EMDatabank Unified Data Resource for 3DEM. *Nucleic Acids Res.* **2016**, *44* (D1), D396–D403.
- (2) Wang, N.; Jiang, X.; Zhang, S.; Zhu, A.; Yuan, Y.; Xu, H.; Lei, J.; Yan, C. Structural Basis of Human Monocarboxylate Transporter 1 Inhibition by Anti-Cancer Drug Candidates. *Cell* **2021**, *184* (2), 370–383.e13.
- (3) Conde, J.; Pumroy, R. A.; Baker, C.; Rodrigues, T.; Guerreiro, A.; Sousa, B. B.; Marques, M. C.; de Almeida, B. P.; Lee, S.; Leites, E.

- P.; Picard, D.; Samanta, A.; Vaz, S. H.; Sieglitz, F.; Langini, M.; Remke, M.; Roque, R.; Weiss, T.; Weller, M.; Liu, Y.; Han, S.; Corzana, F.; Morais, V. A.; Faria, C. C.; Carvalho, T.; Filippakopoulos, P.; Snijder, B.; Barbosa-Morais, N. L.; Moiseenkova-Bell, V. Y.; Bernardes, G. J. L. Allosteric Antagonist Modulation of TRPV2 by Piperlongumine Impairs Glioblastoma Progression. *ACS Cent. Sci.* **2021**, *7* (5), 868–881.
- (4) Griffith, D. A.; Edmonds, D. J.; Fortin, J.-P.; Kalgutkar, A. S.; Kuzmiski, J. B.; Loria, P. M.; Saxena, A. R.; Bagley, S. W.; Buckeridge, C.; Curto, J. M.; Derksen, D. R.; Dias, J. M.; Griffor, M. C.; Han, S.; Jackson, V. M.; Landis, M. S.; Lettiere, D.; Limberakis, C.; Liu, Y.; Mathiowetz, A. M.; Patel, J. C.; Piotrowski, D. W.; Price, D. A.; Ruggeri, R. B.; Tess, D. A. A Small-Molecule Oral Agonist of the Human Glucagon-like Peptide-1 Receptor. *J. Med. Chem.* **2022**, *65* (12), 8208–8226.
- (5) Kim, J. J.; Gharpure, A.; Teng, J.; Zhuang, Y.; Howard, R. J.; Zhu, S.; Novello, C. M.; Walsh, R. M., Jr; Lindahl, E.; Hibbs, R. E. Shared Structural Mechanisms of General Anaesthetics and Benzodiazepines. *Nature* **2020**, *585* (7824), 303–308.
- (6) Miller, P. S.; Aricescu, A. R. Crystal Structure of a Human GABAA Receptor. *Nature* **2014**, *512* (7514), 270–275.
- (7) Laverty, D.; Thomas, P.; Field, M.; Andersen, O. J.; Gold, M. G.; Biggin, P. C.; Gielen, M.; Smart, T. G. Crystal Structures of a GABAA-Receptor Chimera Reveal New Endogenous Neurosteroid-Binding Sites. *Nat. Struct. Mol. Biol.* **2017**, *24* (11), 977–985.
- (8) Nakane, T.; Kotecha, A.; Sente, A.; McMullan, G.; Masiulis, S.; Brown, P. M. G. E.; Grigoras, I. T.; Malinauskaitė, L.; Malinauskas, T.; Miehl, J.; Uchanski, T.; Yu, L.; Karia, D.; Pechnikova, E. V.; de Jong, E.; Keizer, J.; Bischoff, M.; McCormack, J.; Tiemeijer, P.; Hardwick, S. W.; Chirgadze, D. Y.; Murshudov, G.; Aricescu, A. R.; Scheres, S. H. W. Single-Particle Cryo-EM at Atomic Resolution. *Nature* **2020**, *587* (7832), 152–156.
- (9) Robertson, M. J.; van Zundert, G. C. P.; Borrelli, K.; Skiniotis, G. GemSpot: A Pipeline for Robust Modeling of Ligands into Cryo-EM Maps. *Structure* **2020**, *28* (6), 707–716.e3.
- (10) Muenks, A.; Zepeda, S.; Zhou, G.; Veelsler, D.; DiMaio, F. Automatic and Accurate Ligand Structure Determination Guided by Cryo-Electron Microscopy Maps. *Nat. Commun.* **2023**, *14* (1), 1–10.
- (11) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461.
- (12) Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model.* **2019**, *59* (2), 895–913.
- (13) Vasishtan, D.; Topf, M. Scoring Functions for cryoEM Density Fitting. *J. Struct. Biol.* **2011**, *174* (2), 333–343.
- (14) Locke, J.; Joseph, A. P.; Peña, A.; Möckel, M. M.; Mayer, T. U.; Topf, M.; Moores, C. A. Structural Basis of Human Kinesin-8 Function and Inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (45), E9539–E9548.
- (15) Peña, A.; Sweeney, A.; Cook, A. D.; Locke, J.; Topf, M.; Moores, C. A. Structure of Microtubule-Trapped Human Kinesin-5 and Its Mechanism of Inhibition Revealed Using Cryoelectron Microscopy. *Structure* **2020**, *28* (4), 450–457.e5.
- (16) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible docking11Edited by F E. Cohen. *J. Mol. Biol.* **1997**, *267* (3), 727–748.
- (17) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastriitis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428* (4), 720–725.
- (18) Joseph, A. P.; Lagerstedt, I.; Jakobi, A.; Burnley, T.; Patwardhan, A.; Topf, M.; Winn, M. Comparing Cryo-EM Reconstructions and Validating Atomic Model Fit Using Difference Maps. *J. Chem. Inf. Model.* **2020**, *60* (5), 2552–2560.
- (19) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713.
- (20) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17* (10), 6262–6280.
- (21) Korb, O.; Stützel, T.; Exner, T. E. An Ant Colony Optimization Approach to Flexible Protein–ligand Docking. *Swarm Intelligence* **2007**, *1* (2), 115–134.
- (22) Gu, S.; Smith, M. S.; Yang, Y.; Irwin, J. J.; Shoichet, B. K. Ligand Strain Energy in Large Library Docking. *J. Chem. Inf. Model.* **2021**, *61* (9), 4331–4341.
- (23) Korb, O.; Stützel, T.; Exner, T. E. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In *Ant Colony Optimization and Swarm Intelligence*; Springer: Berlin Heidelberg, 2006; pp 247–258.
- (24) Rosenthal, P. B.; Henderson, R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-Particle Electron Cryomicroscopy. *J. Mol. Biol.* **2003**, *333* (4), 721–745.
- (25) Kimanius, D.; Dong, L.; Sharov, G.; Nakane, T.; Scheres, S. H. W. New Tools for Automated Cryo-EM Single-Particle Analysis in RELION-4.0. *Biochem. J.* **2021**, *478* (24), 4169–4185.
- (26) Sanchez-Garcia, R.; Gomez-Blanco, J.; Cuervo, A.; Carazo, J. M.; Sorzano, C. O. S.; Vargas, J. DeepEMhancer: A Deep Learning Solution for Cryo-EM Volume Post-Processing. *Commun. Biol.* **2021**, *4* (1), 1–8.
- (27) Su, M.; Zhu, L.; Zhang, Y.; Paknejad, N.; Dey, R.; Huang, J.; Lee, M.-Y.; Williams, D.; Jordan, K. D.; Eng, E. T.; Ernst, O. P.; Meyerson, J. R.; Hite, R. K.; Walz, T.; Liu, W.; Huang, X.-Y. Structural Basis of the Activation of Heterotrimeric Gs-Protein by Isoproterenol-Bound β 1-Adrenergic Receptor. *Mol. Cell* **2020**, *80* (1), 59–71.e4.
- (28) Pintilie, G.; Zhang, K.; Su, Z.; Li, S.; Schmid, M. F.; Chiu, W. Measurement of Atom Resolvability in Cryo-EM Maps with Q-Scores. *Nat. Methods* **2020**, *17* (3), 328–334.
- (29) Zhang, X.; Belousoff, M. J.; Zhao, P.; Kooistra, A. J.; Truong, T. T.; Ang, S. Y.; Underwood, C. R.; Egebjerg, T.; Šenel, P.; Stewart, G. D.; Liang, Y.-L.; Glukhova, A.; Venugopal, H.; Christopoulos, A.; Furness, S. G. B.; Miller, L. J.; Reedtz-Runge, S.; Langmead, C. J.; Gloriam, D. E.; Danev, R.; Sexton, P. M.; Wootton, D. Differential GLP-1R Binding and Activation by Peptide and Non-Peptide Agonists. *Mol. Cell* **2020**, *80* (3), 485–500.e7.
- (30) Kim, Y.; Jeong, E.; Jeong, J.-H.; Kim, Y.; Cho, Y. Structural Basis for Activation of the Heterodimeric GABAB Receptor. *J. Mol. Biol.* **2020**, *432* (22), 5966–5984.
- (31) Gutiérrez-Fernández, J.; Kaszuba, K.; Minhas, G. S.; Baradaran, R.; Tambalo, M.; Gallagher, D. T.; Sazanov, L. A. Key Role of Quinone in the Mechanism of Respiratory Complex I. *Nat. Commun.* **2020**, *11* (1), 4135.
- (32) Zhu, J.; Vinothkumar, K. R.; Hirst, J. Structure of Mammalian Respiratory Complex I. *Nature* **2016**, *536* (7616), 354–358.
- (33) Pospich, S.; Merino, F.; Raunser, S. Structural Effects and Functional Implications of Phalloidin and Jasplakinolide Binding to Actin Filaments. *Structure* **2020**, *28* (4), 437–449.e5.
- (34) Saur, M.; Hartshorn, M. J.; Dong, J.; Reeks, J.; Bunkoczi, G.; Jhoti, H.; Williams, P. A. Fragment-Based Drug Discovery Using Cryo-EM. *Drug Discovery Today* **2020**, *25* (3), 485–490.
- (35) Farabella, I.; Vasishtan, D.; Joseph, A. P.; Pandurangan, A. P.; Sahota, H.; Topf, M. TEMPY: A Python Library for Assessment of Three-Dimensional Electron Microscopy Density Fits. *J. Appl. Crystallogr.* **2015**, *48* (Pt 4), 1314–1323.
- (36) Joseph, A. P.; Lagerstedt, I.; Patwardhan, A.; Topf, M.; Winn, M. Improved Metrics for Comparing Structures of Macromolecular

Assemblies Determined by 3D Electron-Microscopy. *J. Struct. Biol.* **2017**, 199 (1), 12–26.

(37) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein–Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, 47 (12), 2977–2980.

(38) Wójcikowski, M.; Zielenkiewicz, P.; Siedlecki, P. Open Drug Discovery Toolkit (ODDT): A New Open-Source Player in the Drug Discovery Field. *J. Cheminform.* **2015**, 7, 26.

(39) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, 7, 539.

(40) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, 25 (13), 1605–1612.

(41) Cragnolini, T.; Sahota, H.; Joseph, A. P.; Sweeney, A.; Malhotra, S.; Vasishtan, D.; Topf, M. TEMPy2: A Python Library with Improved 3D Electron Microscopy Density-Fitting and Validation Workflows. *Acta Crystallogr. D Struct. Biol.* **2021**, 77 (Pt 1), 41–47.

(42) Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. The Worldwide Protein Data Bank (wwPDB): Ensuring a Single, Uniform Archive of PDB Data. *Nucleic Acids Res.* **2007**, 35, D301–D303.

(43) Wang, R.; Liu, L.; Lai, L.; Tang, Y. SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex. *Molecular modeling annual* **1998**, 4 (12), 379–394.

(44) Wang, R.; Lai, L.; Wang, S. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J. Comput. Aided Mol. Des.* **2002**, 16 (1), 11–26.

(45) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of Octanol-Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J. Chem. Inf. Model.* **2007**, 47 (6), 2140–2148.

(46) Brylinski, M. Aromatic Interactions at the Ligand-Protein Interface: Implications for the Development of Docking Scoring Functions. *Chem. Biol. Drug Des.* **2018**, 91 (2), 380–390.

(47) Eastman, P.; Pande, V. S. OpenMM: A Hardware Independent Framework for Molecular Simulations. *Comput. Sci. Eng.* **2015**, 12 (4), 34–39.

(48) Meli, R.; Biggin, P. C. Spyrmsd: Symmetry-Corrected RMSD Calculations in Python. *J. Cheminform.* **2020**, 12 (1), 49.