



Published in final edited form as:

*J Med Chem.* 2018 December 27; 61(24): 11183–11198. doi:10.1021/acs.jmedchem.8b01292.

## ***qFit-ligand* reveals widespread conformational heterogeneity of drug-like molecules in X-ray electron density maps**

Gydo C.P. van Zundert<sup>\*,1</sup>, Brandi M. Hudson<sup>\*,2</sup>, Saulo H.P. de Oliveira<sup>5</sup>, Daniel A. Keedy<sup>2</sup>, Rasmus Fonseca<sup>3</sup>, Amelie Heliou<sup>4</sup>, Pooja Suresh<sup>2</sup>, Kenneth Borrelli<sup>1</sup>, Tyler Day<sup>1</sup>, James S. Fraser<sup>2</sup>, Henry van den Bedem<sup>#,2,5</sup>

<sup>1</sup>Schrödinger, New York, NY 10036, USA

<sup>2</sup>Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA 94158, USA

<sup>3</sup>Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA 94305, USA.

<sup>4</sup>LIX, Ecole Polytechnique, CNRS, Inria, Université Paris-Saclay, 91128, Palaiseau, France.

<sup>5</sup>SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA 94025 USA

### **Abstract**

Proteins and ligands sample a conformational ensemble that governs molecular recognition, activity, and dissociation. In structure-based drug design, access to this conformational ensemble is critical to understand the balance between entropy and enthalpy in lead optimization. However, ligand conformational heterogeneity is currently severely underreported in crystal structures in the Protein Data Bank, owing in part to a lack of automated and unbiased procedures to model an ensemble of protein-ligand states into X-ray data. Here, we designed a computational method, *qFit-ligand*, to automatically resolve conformationally averaged ligand heterogeneity in crystal structures, and applied it to a large set of protein receptor-ligand complexes. In an analysis of the cancer related BRD4 domain, we found that up to 29% of protein crystal structures bound with drug-like molecules present evidence of unmodeled, averaged, relatively isoenergetic conformations in ligand-receptor interactions. In many retrospective cases, these alternate conformations were adventitiously exploited to guide compound design, resulting in improved

<sup>#</sup>Corresponding Author vdbedem@stanford.edu (HvdB).

<sup>\*</sup>GCPvZ, BMH contributed equally

Author Contributions

GCPvZ, RF, AH, HvdB developed and implemented core algorithms; GCPvZ, BMH, SHPdO, KB, TD, JSF, HvdB analyzed data; DAK, PS contributed code and data analysis; JSF and HvdB conceived the study. GCPvZ, BMH, JSF, and HvdB wrote the manuscript. All authors edited the manuscript.

JSF is a paid consultant and has a financial stake in Relay Therapeutics. GCPvZ, KB, TD are employees of Schrodinger.

Supporting Information

The supporting information describes additional statistics for our data sets and results, detailed flowcharts of the *qFit-ligand* algorithm, and molecular visualizations of its multi-conformer models. Supplementary tables S1-S5 list details of our benchmark and twilight data sets, and results from prospective application of *qFit-ligand* to BRD containing proteins.

PDB ID Codes

6DMG (remodeled 4FV3); 6DMH (remodeled 4EV4); 6DMI (remodeled 5EZX); 6DMJ (remodeled 5CFW); 6DMK (remodeled 4NR5); 6DML (remodeled 4BW3).

potency or selectivity. Combining *qFit-ligand* with high-throughput screening or multi-temperature crystallography could therefore augment the structure-based drug design toolbox.

## Introduction

Ligands and their protein receptors sample an ensemble of conformations in solution. The energetic contribution of conformational entropy plays a critical role in receptor-ligand molecular recognition,<sup>1,2</sup> but the ensemble of conformations that determines the free energy of binding, activity, and dissociation often remains poorly characterized. The majority of 3-dimensional protein structures are single, static models obtained from X-ray crystallography by averaging over the unique conformations in the unit cells. Crystallographic Atomic Displacement Parameters (ADP) quantify harmonic displacements from average atomic positions, but are adversely affected when unmodeled discrete alternate conformations overlap. By their nature, such static, harmonic models cannot rationalize molecular attributes that rely on dynamic, anharmonic displacements of atoms.<sup>3,4</sup> Revealing discrete conformations<sup>5–7</sup> that are more fully representative of the receptor-ligand conformational ensemble from X-ray electron density maps would overcome this limitation, and create new opportunities to address open questions in chemical and structural biology. For example, such models might help to provide a structural basis for on-pathway conformational intermediates in substrate binding or release detected by NMR.<sup>8,9</sup>

Additionally, an incomplete picture of receptor-ligand structural dynamics impedes structure-based drug design. While overall binding affinities measured in solution report on a receptor-ligand ensemble, the structure-activity relationships are often informed by static models for further optimization, with some exceptions that incorporate flexibility<sup>10</sup>. During small molecule optimization, even minor chemical changes can lead to apparently altered binding modes that are unforeseen due to the limitations of conventional X-ray structural models, frustrating design<sup>11,12</sup> (Figure 1A). One hypothesis to explain how subtle modifications cause a switch to a second binding pose is that the unmodified ligand samples the second pose at low, but potentially detectable, occupancy (Figure 1B). Examples where subtle changes in chemical structure of the ligand led to different binding modes are abundant. Fragment optimization of CDK8 inhibitors revealed that small modifications led to a new binding mode, which was exploited to develop potent and selective inhibitors.<sup>13</sup> Dramatically different binding modes as a result of minor changes in the chemical structure of the ligand are also illustrated by Hsp90 and PTR1 inhibitors. In the course of structure-based optimization of Hsp90 inhibitors, Casale and coworkers observed a flipped binding mode of the ligand in the crystal structure, leading them to change the direction of design towards a low nanomolar compound (Figure 1C).<sup>14</sup> In PTR1, compounds that presumably could not be accommodated by the binding site ultimately led to a boost in affinity, owing to altered binding modes.<sup>14,15</sup> In human lipoprotein-associated phospholipase A2, two related ligands (Figure 1D) that explored distinct subpockets were merged into a more potent ligand (Figure 1E).<sup>16</sup> These selected examples highlight both the perils of design based solely on an initial ligand pose, but also how the fortuitous discovery of alternate poses can create new opportunities for design. It is possible that some of these “new” binding poses pre-exist as alternative conformations at slightly higher energies than the conformation modeled into the

conventional X-ray structure. Many of those are infrequently sampled due to high energy differences and fall below the detection level in the maps. However, some of these conformations may be evident in electron density maps, but at levels that are frequently ignored in modeling.

Indeed, in the phospholipase example above, the presence of difference electron density suggested that multiple conformations might have been sampled in at least one of the smaller ligands (Figure 1F). This view posits that degeneracies in ligand binding modes can also be accessed by small modifications of the ligand chemical structure that shift the receptor-ligand equilibrium ensemble. This hypothesis is supported by additional anecdotal examples, such as long time scale molecular dynamics (MD) simulations of trypsin, in which both the ligand and receptor adopt many stable configurations.<sup>17</sup> Experimentally, multiple conformations can be present in X-ray electron density, as in HDAC6 where many related ligands were modeled in multiple conformations in concert with distinct conformations of the receptor.<sup>18</sup> In other cases where the alternate conformations have been experimentally revealed, they have been exploited to improve affinity. For example, X-ray crystallography revealed alternate conformations for singly-substituted EphB4 ligands that inspired the creation of bis-substituted ligands with increased potency (Figure 1G).<sup>19</sup> NMR measurements of conformational heterogeneity for ligands generated against the antibacterial target LpxC uncovered a larger cryptic envelope that was filled by larger, more potent ligands.<sup>20</sup> Collectively these examples suggest that multiple ligand poses are likely energetically accessible for many proteins. Fully realizing the potential of this phenomenon in structure-based design, exemplified by EphB4 and LpxC inhibitors, requires reliable characterization of pre-existing conformational heterogeneity (Figure 1B) of ligand-receptor complexes.

While protein conformational heterogeneity has been automatically and systematically characterized in X-ray crystallography data,<sup>7,21–24</sup> ligand conformational heterogeneity is less explored. Various software algorithms can identify and build ligands into electron density maps without human intervention;<sup>25–30</sup> however, these approaches typically provide several top scoring ligand conformations at unit occupancy. While in principle the user can select multiple candidate conformations for the final model, none of these approaches consider an ensemble of alternate conformations from the outset. Moreover, they may build unrealistic conformations that incorrectly fit into the ensemble-averaged electron density<sup>31</sup>. Despite the importance of low occupancy alternative conformations in biochemical function and ligand design, statistical measures for monitoring the fit of ensemble models to data are less developed than for single conformer models.

Here, we present a new, automated approach based on *qFit*,<sup>22,23</sup> called *qFit-ligand*, to create parsimonious multiconformer ligand models in crystallographic electron densities. We first surveyed the PDB to investigate ligand heterogeneity in current crystal structures, and selected a diverse, curated benchmark set of pharmaceutically relevant protein targets with alternate ligand conformations across a wide range of resolution and occupancy (Table S1). We found that *qFit-ligand* can detect alternate conformations at occupancies down to 20%, even at relatively modest resolutions of 2.0Å. We then applied our method prospectively to all cases of the *Drug Design Data Resource (D3R, [drugdesigndata.org](http://drugdesigndata.org))*, a subset of the

*Twilight database*,<sup>32</sup> and all PDB entries for the bromodomain-containing protein 4 (BRD4), revealing unmodeled alternate conformations in 29% of the cases for the latter. To evaluate the quality of our multiconformer ligands, we calculated R-factors and ligand energies relative to a single conformer ligand model. Our results indicate that *qFit-ligand* is a powerful, efficient, and user-friendly tool to model and discover alternate ligand conformations.

## Results

### Creating a benchmark set of true positive ligand alternative conformations from the PDB

To estimate the prevalence of multiconformer ligands in crystal structures, we surveyed all 130,054 PDB entries as of June 2017 that contained non-covalently bound ligands with more than 15 non-hydrogen atoms in their X-ray crystal structure. This resulted in 44,620 PDB entries totaling 133,724 ligands. Of those, 2,611 ligands, or less than two percent (1,078 unique ligand codes), distributed over 1,845 PDB entries consisted of two or more alternate conformations (Materials and Methods, Figure S1). Many of these molecules are common crystallographic additives (PEG, cholesterol, etc.) or metabolites (ATP, NADPH, etc.). We therefore manually curated a true positive benchmark set of receptors of pharmaceutical interest containing multiconformer, drug-like molecules. Cases where ligands adopted entirely different binding modes, such as flipped ligands, were discarded. This resulted in 90 crystal structures that could be stably refined against the deposited structure factors and CIF restraints files (Table S1).

We apportioned the conformational heterogeneity of ligands in our benchmark set into four categories (Figure 2A): terminal end flips, where only terminal atoms are flipped/rotated; ring flips where a ring system is flipped, usually by 180 degrees; branching ligands, where a side-chain or branch of a ligand has an alternate conformation; and displaced ligands, where all atoms are at least slightly displaced in combination with differences in their internal degrees of freedom. The benchmark set heavily overrepresented 50/50 occupancy splits (Figure 2B), reflecting a historical tendency against refining occupancies in favor of refining ADPs only. Inaccurate 50/50 occupancy splits do not inform on which conformation represents the dominant binding mode, which is often critical information for design decisions in lead optimization. Re-refinement with PHENIX<sup>33</sup> substantially broadened the distribution (Figure 2B). Unlike the occupancies, the RMSD between alternate conformations was similar after re-refinement (Figure 2C). Interestingly, we observed no correlation between occupancy shift and difference in mean B-factor (Figure 2D). This result is in contrast to earlier reports<sup>34,35</sup>, likely owing to improvements in refinement software.

### Developing the qFit-ligand algorithm and calibration against synthetic data

We designed the *qFit-ligand* algorithm to iteratively explore a vast conformational space to determine a parsimonious ensemble of up to five occupancy-weighted conformations that, collectively, optimally fits the electron density (Figure 3A, Materials and Methods, Figure S2). Briefly, *qFit-ligand* takes as input a refined, single conformation receptor-ligand structure in PDB format, and a  $2mF_o - DF_c$  density map. It first determines rotatable bonds and rigid groups of atoms within the ligand (Materials and Methods, Figure S2). Starting

from each rigid group, *qFit-ligand* performs a local, six-dimensional translational and rotational search in the rigid group's neighborhood, selecting up to five occupancy-weighted candidate positions that, collectively, minimize the real-space density residual of the rigid group. In subsequent steps, *qFit-ligand* iteratively grows the rigid group by exhaustively sampling increments of several torsion angles simultaneously, while avoiding collisions with the receptor. At each step, it selects up to five occupancy-weighted conformations by again minimizing the density residual. This is repeated until all torsion angles are determined and the full ligand is built up. The maximum number of conformations generated by *qFit-ligand* at this stage is five times the number of rigid groups. The final occupancy-weighted, multiconformer ensemble is selected from this pool by combining cross-correlation, geometric, and density residual measures. The ligand multiconformer ensemble is then combined with the receptor and refined with phenix.refine.

We calibrated *qFit-ligand* on synthetic data calculated from the benchmark set at varying resolutions and occupancies (Materials and Methods, Figure S3). To create starting models we deleted the alternate 'B' conformation of the ligand, set the occupancy of the 'A' conformation to 1.0, and re-refined against the deposited structure factors to reposition the 'A' conformation as the only modeled conformation.

We first determined optimal sampling parameters for *qFit-ligand*, balancing accuracy of the results and computational demands, measured by the minimum normalized RMSD (nRMSD, Materials and Methods) between *qFit-ligand* generated conformations and the benchmark B conformation. RMSDs between the A and B conformer range from 0.22 to 4.86 Å across the benchmark set. nRMSD controls for this large variation by normalizing the RMSD between *qFit-ligand* generated conformations and the B conformer by the RMSD between A and B. nRMSD values range from 0 to 1, with values closer to 0 indicating better performance. To compare performance we report the median nRMSD of the benchmark set for different resolutions and occupancies. Our analysis suggested that sampling two torsion angles simultaneously at 6° intervals gave the best result across resolutions and occupancies, while limiting computational costs (Figure S4).

Benchmarking *qFit-ligand* showed performance differences across category, resolution, and true occupancy. We evaluated the performance of *qFit-ligand* for each category of conformational heterogeneity (Figure 3B, 3C). At a resolution of 1.60 Å and equal occupancies for the A and B conformer, the median nRMSD is 0.11 for terminal end flips, 0.18 for ring flips, 0.37 for branched ligands and 0.69 for displaced ligands (Figure 3B). Unsurprisingly, *qFit-ligand* performance decreased with increasing complexity of conformational heterogeneity.

The resolution dependence is more complex, however, with better performance of *qFit-ligand* at an intermediate resolution of 1.60 Å (Figure 3C, 3E). We attribute this to more sharply defined density peaks at high resolutions relative to intermediate resolutions. Undersampling of conformations during ligand building can result in failure to accurately hit the density peaks at high resolution, thereby leading to suboptimal scores of electron density-based measures.

The performance of *qFit-ligand* is sensitive to the occupancy of the alternate conformation, indicated by an increasing median nRMSD for lower occupancies of the B conformation (Figure 3D, 3E). At occupancies of 0.2 and below, the alternate conformer was rarely detected at any resolution. While *qFit-ligand* samples conformations close to the alternate conformer, evidenced by favorable (low) nRMSDs before final rescoring (Figure S4), selecting them at low occupancies would increase the false positive rate (data not shown).

Finally, note that the *qFit-ligand* multi-conformer ensemble *collectively* explains the density best. In virtually all cases, density profile contributions from all ligand conformers are required to match the experimental profile. Overlapping ensemble members evaluated in isolation generally return suboptimal density fitting statistics.

### qFit-ligand re-identifies low energy alternative conformations in experimental data

Next, we applied our method to the experimental benchmark data set for cases with resolution better than 2.20 Å (73 cases), with only the 'A' conformation retained in the re-refined starting model (Figure 4). *qFit-ligand* performance with real data followed the trend we observed with simulated data: localized conformational disorder like terminal end flips and ring flips were determined with higher accuracy than branched or displaced disorder. In the case of terminal ends, only one or two atoms report on the distance between alternate conformations, but the nRMSD is dominated by small coordinate shifts distributed over the entire ligand. Despite a median nRMSD of 0.56 for terminal ends, the median nRMSD for only the 'reporting' atoms is less than 0.30 (Figure 4A; all RMSDs reported in Figure S5). Consequently, 13 out of 17 terminal end flip *qFit-ligand* results were sufficiently accurate to recognize the benchmark alternate conformer. Similarly, *qFit-ligand* determined ring rotations to within a median nRMSD of 0.27. However, the median nRMSD for branched cases is 0.78 and for displaced cases is 0.87. If we conservatively designate a *qFit-ligand* result with more than one conformation and nRMSD > 0.6 as a false positive (Figure S6), the false positive rates for each category are 29% (terminal flip), 14% (ring flip), 67% (branching), and 85% (displaced) (*see Discussion section regarding false positives*).

The distribution of  $R_{\text{free}}$  values of fully automated refined *qFit-ligand* models were nearly identical to that of the single conformer ligand models (*qFit-ligand*  $\bar{x}$ =0.2043, single conformer  $\bar{x}$ =0.2048; p-value=0.46, two-sided t-test), and statistically indistinguishable from the deposited, manually curated multiconformers ( $\bar{x}$ =0.2038; p-value=0.51) (Figure 4B, Figure S7, Table S2). The *qFit-ligand* multiconformer models improved (worsen)  $R_{\text{free}}$  in 53% (47%) of cases compared to single conformer ligand models.

The  $R_{\text{free}}$  value alone is generally not sufficiently sensitive to detect if a multi-conformer model is, collectively, supported by the electron-density. Therefore, as an independent validation in addition to the RSCC in the *qFit-ligand* algorithm, we calculated the Electron Density support for Individual Atoms (EDIAM) score<sup>36</sup>. The EDIAM score quantifies how well a group of atoms is supported by the electron density (Materials and Methods). EDIAM scores indicated that the *qFit-ligand* models improve agreement with the electron density in 63% of cases compared to a single conformer model (Figure 4D), and overall were within 0.2 of those of single conformer models. In cases where deposited models were well-supported by the electron density (EDIAM > ~0.8), *qFit-ligand* models generally increased



EDIAm compared to single conformer models--in some cases even improving the deposited model (Figure 4D,E, 3oik). Unsurprisingly, EDIAm scores correlated with the categories of conformational disorder; branched or displaced ligands had lower EDIAm scores. Nonetheless, even in those challenging cases, automated *qFit-ligand* multi-conformers EDIAm scores are on par with those of single-conformer models, while alerting to the presence of conformational heterogeneity (Figure 4E, right-hand panels; Suppl Fig S15). We emphasize that  $R_{\text{free}}$  and EDIAm values from fully automated modeling generally improve with manual refinement.

To further evaluate the quality of *qFit-ligand* multiconformer models, we examined internal ligand energies, ignoring interactions with the receptor. Although *qFit-ligand* does not observe restraints in modeling, the quality of *qFit-ligand* multiconformer models as measured by their internal energy was excellent. We found a median conformationally-averaged excess energy of only 0.35 kcal/mol (Figure S16, Table S2, Supplementary Material and Methods), i.e., *qFit-ligand* multiconformer models were, on average, nearly indistinguishable from that of the re-refined single 'A' conformation in our benchmark set. Interestingly, automatically building a *qFit-ligand* multiconformer model in some cases substantially reduced the ligand energy compared to the single 'A' conformation. For example, for acyliminobenzimidazole inhibitor 36 in complex with human anaplastic lymphoma kinase,<sup>37</sup> a series of concerted dihedral angle changes resulted in conformations that better fit the density, and reduced the energy by nearly 7 kcal/mol (Figure S8). This suggests that a ligand can accumulate strain energy when it is forced into an averaged conformation to fit the density. On the other hand, the distribution of ligand excess energies suggests that ligands also access higher energy conformations, within a few kcal/mol from the single conformation (Figure 4D). Indeed, ligands generally may not bind in the lowest energy conformation, or even adopt a local minimum.<sup>38</sup> Favorable non-covalent interactions with the receptor, buried hydrophobic surfaces, or desolvation of ordered waters in the binding pocket can overcome penalties of strained conformations.<sup>39,40</sup> Thus, alternate conformations, even at elevated ligand energies, may reduce the free energy of the receptor-ligand complex.

### **qFit-ligand discovers new alternate conformations in the D3R and the Twilight Databases.**

For prospective discovery, we first applied *qFit-ligand* to the 145 crystal structures in the D3R dataset, a high-quality collection of manually curated protein-ligand crystal structures ranging in resolution from 1.26 to 2.75 Å, designed for validation and improvement of methods in computer-aided drug design. Of the ten crystal structures in the D3R dataset with alternate ligand conformations, *qFit-ligand* recovered seven to within a median nRMSD of 0.24. Four of these overlap with our benchmark (PDB ID 4FV3, Figure 5A; and PDB IDs 4EK6, 4EK8, 4Y6D). We ranked all *qFit-ligand* multiconformer ligands using the Fisher z-transformation, a cross-correlation based metric which measures if alternate conformations are supported by the electron density (Materials and Methods). Three of the top four ranked ligands already had a modeled alternate conformation, and six out of seven recovered multiconformer ligands ranked within the top 20, indicating that the Fisher-z transformation is an effective ranking measure. A new alternate ligand conformation was uncovered by *qFit-ligand* in the crystal structure of the E166A mutant of *Serratia fonticola* carbapenemase

(PDB ID 4EV4; new PDB ID 6DMH) from the D3R dataset (Figure 5B). The terminal propanol functional group of a bound meropenem intermediate adopts a previously undetected conformation. Thus, *qFit-ligand* recovered 70% of ligand alternate conformations and even revealed a new alternate conformation in highly scrutinized experimental data (*qFit-ligand* multiconformer models of prospective cases can be found in Zenodo; DOI: [10.5281/zenodo.1256262](https://doi.org/10.5281/zenodo.1256262)).

We also applied *qFit-ligand* to a subset of the Twilight data set. The Twilight data set represents ligand structures in the PDB poorly supported by the electron density map, potentially indicating conformational disorder in the map, or incorrectly modeled ligands. We applied *qFit-ligand* to ligands in the Twilight database with 15 to 36 non-hydrogen atoms, resolutions better than 2.0Å, and a correlation coefficient higher than 0.6, resulting in 2,379 cases over 1,168 PDB entries, which we ranked by Fisher z-score to identify ‘hits’ of unmodeled, alternate conformations (Table S3). We proceeded by manually inspecting the top 10% of cases.

In many cases, the electron density near the ligand was severely disordered, consistent with the intent of the Twilight database to flag questionable models of ligands. While *qFit-ligand* suggested alternate conformations, their validity could not unambiguously be confirmed using electron density measures. Nonetheless, in some instances significantly improved receptor-ligand interactions signified plausible alternate ligand conformations (Figures S9, S10, S11).

However, for several crystal structures *qFit-ligand* unambiguously detected alternate ligand binding conformations. For example, in the crystal structure of BACE-1, *qFit-ligand* finds three conformations (occupancies of 0.32/0.36/0.32) of inhibitor 5T5 (PDB ID 5EZS, new PDB ID 6DMI) (Figure 5C). These conformations show the potential to engineer a ligand that can accommodate strong hydrogen bonding interactions to R189 and W137. In the single conformer model, one of the difluoromethoxy fluorines of the ligand and the amino group of R189 interact weakly through a 3.0Å hydrogen bond.<sup>41</sup> Our multiconformer model shows that the ligand can adopt a position where strong hydrogen bonding between these groups occur (i.e., shorter contact distances), although at the expense of sacrificing a favorable hydrogen bonding interaction to W137. In addition, one of the *qFit-ligand* conformations turns a non-ideal hydrogen bond geometry in the crystal structure between the difluoromethoxy oxygen and amino group of W137 into an ideal geometry. Strikingly, the difluoromethoxy group of BACE-1 inhibitor 5T6 (PDB ID 5EZZ) has a binding conformation that exploits this same hydrogen bonding interaction with W137 sampled by 5T5 at low occupancy (Figure 5D).<sup>42</sup> These insights could be used to create a new ligand with branching substituents that can simultaneously form strong hydrogen bonding interactions with R189 and W137 in hopes to increase binding affinity.

### **qFit-ligand identifies widespread conformational heterogeneity in BRD4 ligand**

In another notable example from the Twilight data set, *qFit-ligand* identified a minor, unmodeled population of inhibitor compound BDOIA383 bound to bromodomain-containing protein 4 (BRD4), a BET (bromodomain and extra terminal domain) BRD (Figure 6A, PDB ID 5CFW, new PDB ID 6DMJ). BRDs are small, epigenetic ‘readers’,



which recognize and bind histone acetylated lysine (AcK).<sup>43</sup> BRDs can thereby epigenetically control gene transcription, and have recently emerged as important drug targets.<sup>44</sup> The human genome contains 61 BRDs, distributed over 46 diverse proteins.<sup>43</sup> Several potent small-molecule inhibitors for BRD4 have been structurally characterized, but designing modulators selective between BRD4 and CREB binding protein (CBP) has proved challenging.<sup>45–47</sup>

The BRD4-BDOIA383 example highlights how alternate conformations can expose molecular surfaces away from the primary recognition site that could be exploited for selectivity. Inspection of the major population of BDOIA383 bound to BRD4 revealed that it is stabilized by a crystal contact between the BDOIA383 morpholine oxygen and the K91 carbonyl (Figure 6A). By contrast, the minor conformation does not engage in crystal contacts, and buries nearly 6% more ligand solvent accessible surface area (3115 Å<sup>2</sup>) than the major conformation (3302 Å<sup>2</sup>) (Figure 6B). Additionally, the minor conformation interacts with D145 at the N-terminus of helix  $\alpha$ C (Figure 6A,B). Interestingly, compound BDOIA383 in complex with CBP BRD revealed a rotation of the isoxazole-benzimidazole bond by 180°, exposing the phenethyl group to substituted R1173 at the  $\alpha$ C N-terminus<sup>46</sup> (PDB ID 5CGP), occupying the space of the *qFit-ligand* minor morpholine conformation in BRD4. Subsequent ligand modifications strengthened these interactions, leading to increasingly selective CBP modulators.

Flipped binding modes between BRD4 and CBP had been observed earlier with a bound isoxazolyl-benzimidazole ligand, also leading to improved selectivity.<sup>47</sup> Strikingly, in this case too *qFit-ligand* identified a minor conformation in the CBP complex structurally close to the BRD4 bound conformation (Figure 6C). Because selectivity is often achieved by leveraging unique structural ligand-receptor interactions, identifying alternate ligand conformations may help profile auspicious ligand-receptor secondary molecular recognition sites.

Application of *qFit-ligand* to all 126 BRD2–4 crystal structures in the PDB (deposited in Zenodo; DOI: [10.5281/zenodo.1256262](https://doi.org/10.5281/zenodo.1256262)) suggests that differential binding modes and ligand heterogeneity are remarkably ubiquitous (Materials and Methods). Visual inspection and manual curation revealed 12 new binding conformations detected with high confidence (Figure S12), and an additional 24 with possible alternate conformations (Table S5, Figure S13). The alternate conformation detected by *qFit-ligand* in the crystal structure of BRD4 with ligand 9BM<sup>48</sup> (phenyl ring flip) is the same conformation observed in the crystal structure of BRD2 with ligand S5B<sup>49</sup>, further supporting the idea that ligands can sample a second, minor pose which becomes dominant after small chemical modifications to their structure. These observations and the pervasiveness of ligand heterogeneity in BRDs suggest that the potential of alternate ligand conformations for structure-based drug design is significantly undervalued.

## Discussion

Ligand conformational heterogeneity is widespread in X-ray data deposited in the PDB, but underreported in the absence of automated and reliable computational methods. *qFit-ligand*

is a new method to model a parsimonious multiconformer ligand in crystallographic electron densities. We formulated the challenge of identifying an ensemble that collectively best agrees with the electron density, from up to tens of thousands of candidate conformations, as a stepwise combinatorial optimization problem. *qFit-ligand* relies on exhaustive sampling, iteratively restrained by the electron density, to fully cover the conformational space of each ligand. Our method was designed to provide a measure of statistical confidence for model selection in cases of ligand conformational heterogeneity, separating the contributions between harmonic displacement and multiple conformations.

We highlighted three examples of previously unmodeled alternate ligand conformations obtained from prospective application of *qFit-ligand* to the D3R and Twilight data sets. (Table S4 lists all prospective discoveries.) Strikingly, even the highly curated D3R data set revealed a previously unmodeled alternate conformation. Targeted application to a single bromodomain receptor suggests that as many as 29% of receptor-ligand crystal structures could have alternate ligand conformations. This is a likely conservative estimate of the accessible ligand conformational landscape in protein crystal structures. The portrait of rigid receptors and ligands is exacerbated by the common practice of collecting X-ray crystallographic data at cryogenic temperature (100 K). Although cryocooling increases the precision of structure determination by reducing thermal motions, cooling affects the conformational distribution of more than 35% of side chains in proteins<sup>50</sup> and has been found to alter ligand binding and abolish transient binding sites observed at room temperature.<sup>51</sup> Applying *qFit-ligand* to room temperature X-ray crystallography data, which can shift the equilibrium of receptor-ligand conformational ensembles, may reveal additional minor ligand binding poses that are typically masked in cryogenic data.

*qFit-ligand* conformational strain energies after refinement were nearly indistinguishable from those of the manually curated benchmark set, signifying its multiconformer models are chemically accurate. Our analysis showed that *qFit-ligand* conformations were nearly isoenergetic. Less often we identified a ligand for which the *qFit-ligand* multiconformer model significantly reduced strain energy, suggesting that conformational averaging of the single conformer model had led to a poorly modeled ligand to fit the density. Changes in the free energy of the ligand-receptor complex will give a more complete picture of the effect of multi-conformer ligands, but require a model of the receptor response to binding, which is beyond the scope of the current study. Ironically, multiconformer ligands are commonly filtered out of major test sets for development of docking and conformational sampling approaches, but may be a more ‘trustworthy’ representation of the underlying data<sup>52</sup>. Multiconformer ligand models can therefore address important challenges in ligand validation and deposition in the Protein Data Bank (PDB).<sup>53,54</sup>

Several aspects of *qFit-ligand* could be improved in the future. First and foremost, modeling the *joint* conformational ensemble of the binding site residues and ligand will be a major step forward towards a better understanding of receptor response to ligand binding. The potential of this approach is underlined by our finding that 12% of structures in the twilight data set already present alternate receptor conformations within 5 Å of the ligand. While ultimately an accurate thermodynamic distribution for receptor-ligand conformations is desired, our approach determines a parsimonious ensemble of conformations that is

supported by the data alone. These conformations can provide a starting point for MD simulation to obtain a better estimation for a thermodynamic distribution. Second, while in principle our conformational sampling approach could be combined with a sophisticated force field,<sup>55</sup> relying on energy restraints in the discovery stage can result in increased computational cost and can risk excluding promising candidate conformations owing to imperfect sampling and steep potential energy gradients. Rather, we advocate including energy restraints in refinement.<sup>31</sup> Other conformational search methods with force fields, such as Schrodinger's ConfGen<sup>56</sup> and MacroModel,<sup>57</sup> OpenEye's Omega,<sup>58</sup> MOE,<sup>59</sup> and many other freely available tools<sup>60</sup> could independently validate results in the absence of data.

Third, the false-positive rates can likely be decreased. Several techniques are available to guard against overfitting. For example, the BIC criteria weighs increased model complexity (number of conformations) to a better statistical fit. Terminal ends and ring flips had low false positive rates, whereas those for the branched and displaced categories were elevated. *qFit-ligand* recovered terminal end and ring flips to within an nRMSD of 0.27. Disordered parts of branched ligands were often solvent-exposed, and therefore more difficult to recover. Accounting for partial occupancy of ligand, receptor conformers, and solvent will be required to reveal details at the solvent interface. Displaced ligands are not 'anchored' in the binding pocket, and are among the most challenging to recover. An nRMSD of 0.5 or below often sufficed for further refinement, but even nRMSDs up to 0.75 sometimes required only minor manual adjustments, depending on the absolute RMSD between states. For example, ring flips involving rotationally symmetric atom species cannot be uniquely assigned based on the electron density alone, leading to inflated RMSD measures, which are easily adjusted (Materials and Methods, PDB ID 3P4V). All benchmark *qFit-ligand* results were obtained with the same parameters. In practice, specific problems will dictate tailored settings. For example, selecting finer sampling steps or larger volumes for rigid body searches could give better results at the expense of increased computational time (Figure S17). Nonetheless, on our benchmark set, the fully automated *qFit-ligand* multiconformer models were supported by real-space validation measures, and their  $R_{\text{free}}$  values were statistically indistinguishable from the manually curated set. While these measures cannot perfectly distinguish false positives, structural models consistent with the reflection data create a pool of testable hypothesis, which can be evaluated in drug discovery using ligand structure activity relationships or protein residue mutations.

Fourth, the partial occupancy of an "unbound"/apo state could be explicitly considered. Weak, overlapping densities originating from partially occupied receptor and ligand conformations in the binding pocket are often difficult to tease apart, owing to a vast number of possible ligand conformations to be evaluated, even in sterically constrained binding cavities. Partially occupied water molecules and crystallographic additives often further confound modeling efforts. In these challenging cases, difference densities from alternate states are often incorrectly resolved by waters. Promising new approaches such as PanDDa<sup>61</sup> can reveal the electron density of partially occupied states; however, it requires a large number of 'ground state' crystal structures to reliably compute their contribution to the partially occupied state. Our method is highly complementary and PanDDa maps could even be used as input to *qFit-ligand*. Synergy of these approaches holds the promise to enable

efficient, accurate and unbiased discovery of alternate fragment and ligand binding poses. This is increasingly important in view of the ability to structurally screen hundreds of candidate ligands within hours on modern synchrotrons. The combination of these methods may help remove temptation to fill all difference density with waters, while avoiding the overly optimistic modeling of partial occupancy ligands that are highlighted by the Twilight database.

Even with these improvements, it is important to be sensitive to concerns of overfitting when fitting additional conformations into electron density maps. We have guarded against this possibility in multiple ways: 1) we identify conformations similar to “expert” manual modeling in experimental test sets; 2) in synthetic datasets the Fisher Z score identifies a statistically significant parsimonious set that correspond in number to the underlying “true” number of conformations; 3) traditional global metrics (Rfree) and new local metrics (EDIA) do not identify major concerns. Furthermore, on the prospective set 1) new interactions are identified that make biochemical sense; 2) we identified similar conformations in ligands that differ by only a few atoms; 3) visual inspection confirms that Fo-Fc difference density is reduced. This last point underscores how visual inspection is important for eliminating false positives and, as shown here, for manually improving the fit. Furthermore, these checks are especially important for verifying the correct stereochemistry. Ultimately, biochemical tests with mutations or altered ligands are the best forward validation.

Finally, as the particle size and resolution limitations of single-particle cryo-electron microscopy (cryo-EM) continue to improve, that technology will have a major impact on drug discovery.<sup>62</sup> High-throughput and automation approaches,<sup>63</sup> combined with the size of the complexes in cryo-EM structure determination, will soon turn careful modeling of protein and ligand structural heterogeneity into a major bottleneck. *qFit-ligand* can provide an efficient, automated modeling approach at the amino-acid length-scale, as EM maps are immutable during modeling and refinement. Beyond applications to drug discovery, as time-resolved serial crystallography is rapidly becoming routine at X-ray free electron lasers and even synchrotrons,<sup>64</sup> *qFit-ligand* can help resolve minor populations of structural protein-ligand intermediates in light-driven pump-probe<sup>65</sup> or structural enzymology ‘mix-and-inject’ experiments.<sup>66,67</sup>

## Conclusions

Revealing the full receptor-ligand conformational ensemble can help drug design by exploiting the balance between entropy and enthalpy in compound design<sup>68,69</sup> and by characterizing the effect of pre-rigidifying ligands on affinity.<sup>70</sup> Equally important, it can help the rational design of ligand selectivity by exposing accessible molecular surfaces unique to their intended targets.<sup>71</sup> In the future, full integration of *qFit-ligand* with *qFit* could reveal the structural reorganization of binding pockets and allosteric signal propagation in the receptor upon ligand binding.<sup>21,72</sup> Our *qFit-ligand* open source software, available from [https://github.com/ExcitedStates/qfit\\_ligand](https://github.com/ExcitedStates/qfit_ligand), provides promising, new starting points for ligand optimization and structure based drug discovery. However,

communication between structural biologists, computational chemists, and medicinal chemists remains a requisite for successful, rational design.

## Materials and Methods

### Survey of the Protein Data Bank and benchmark creation

All structure coordinate files were downloaded from the PDB. Structures determined by X-ray crystallography were checked for *HETATM* entries (ligands) containing at least two different *altloc* identifiers with identical chemical composition. We aim to characterize conformational heterogeneity of ligands that are 1) drug-like, and 2) have multiple internal degrees of freedom. Ligands with fewer than 15 non-hydrogen atoms were therefore discarded, as were covalently linked ligands. This resulted in a list of 2,611 ligands divided over 1,845 PDB files. The list was further pruned to exclude ligand flips, i.e. alternate conformations that do not have a common cluster of atoms in space, and alternate conformations consisting exclusively of ring puckers as our algorithm was not designed to sample these types of conformational changes. We then selected receptors and ligands of pharmaceutical interest, resulting in a final benchmark set of 90 cases that refined against the deposited structure factors and CIF files. As the quality of X-ray data affects model building and refinement, we verified that low-resolution X-ray data cut-off and data completeness were within acceptable ranges (Table S1). Each case was re-refined using phenix.refine v1.11 with the following parameters

**For PDBs better than 1.5Å:** optimize\_xyz\_weight=true optimize\_adp\_weight=true optimize\_mask=True main.number\_of\_macro\_cycles=10 adp.individual.anisotropic="not water and not element H" adp.individual.isotropic="water or element H"

**For PDBs worse than 1.5Å:** optimize\_xyz\_weight=true optimize\_adp\_weight=true optimize\_mask=True main.number\_of\_macro\_cycles=10 adp.individual.isotropic=all

Single conformer ligand models were created by removing the ligand's 'B' conformation from the original deposited PDB model and resetting the occupancy of the 'A' conformation to 1. The single conformer models were re-refined as above.

### The *qFit-ligand* approach

The *qFit-ligand* algorithm takes as input the initial structure of a refined single conformer ligand modeled in the electron density, a real-space map ( $2mF_o-DF_c$ ) in *ccp4* format and its resolution, and, optionally, the receptor and other ligand and solvent atoms for clash avoidance. The *qFit-ligand* algorithm starts by scaling the map electron density values to approximate absolute scale, using only the density under the footprint of the receptor. Next, the algorithm calculates an  $F_c$  map corresponding to the receptor and any other atoms in the crystal not part of the ligand. This map is subtracted from the experimental density. Map values below the mean are set to zero to prevent building into spurious density. We call the resulting map the *target map*.

Next, *qFit-ligand* determines rotatable covalent bonds of the ligand and rigid groups of atoms. Covalent bonds are determined based on proximity; i.e., if the distance between two

atoms is smaller than their combined covalent radius plus 0.5 Å, the atoms are considered covalently bonded. It partitions the ligand into rigid groups of atoms (i.e., groups of atoms without internal DoFs) as follows: It constructs a graph  $G = (V_m, E_m)$  such that the vertices  $V_m$  represent atoms, and edges  $E_m$  represent covalent bonds. Next, members of  $E_m$  that 1) are in a ring system, and 2) edges  $(u, v)$  in  $E_m$  where  $u$  or  $v$  has degree one are edge-contracted. Hybridization states (partial double bonds) are ignored in the current implementation. A covalent bond is rotatable if it is not part of the same ring system. After edge-contraction, each edge  $E_m$  ultimately corresponds to a rotating bond (one degree of freedom) and each vertex  $V_m$  corresponds to a collection of atoms that form a rigid body. Thus, rigid groups of atoms do not have internal, rotatable covalent bonds (Figure S2).

Analogous to existing crystallographic model building procedures<sup>73</sup>, *qFit-ligand* iteratively generates and ranks a large number of candidate ensembles by scoring against the electron density and retaining a small ensemble (Figure 3A and S2). It uses modern statistical routines for subset selection in regression models by MIQP<sup>74,75</sup>. *qFit-ligand* builds up candidate ligand conformers starting from each rigid group of atoms in turn, similar to the approach implemented in FlexX/S<sup>76</sup> (Figure S2). To prevent a combinatorial explosion, *qFit-ligand* iteratively samples and selects candidate conformers for successive rigid groups. The first sampling iteration consists of a local rigid body search of the starting group within a box with an edge size of 0.4 Å at a 0.1 Å interval and 10 randomly generated orientations at a maximum rotation angle of 10 degrees (1,250 conformations). An MIQP step (details below) selects up to five occupancy-weighted candidate positions that, collectively, minimize the real-space density residual of the rigid group. For each selected conformation, subsequent iterations each sample  $N$  (user-defined, default 2) torsion angles with a user-defined (default 6°) step size to add  $N$  successive rigid groups simultaneously. Clash detection is divided into internal clashes, i.e., clashes within a single ligand conformer, and external clashes, i.e., clashes between a ligand conformer and the receptor. Clashing conformations, both internal and with the receptor (using an efficient,  $O(1)$ , spatial hashing algorithm), are detected and removed from the set before selection with MIQP. Bond lengths and angles are kept fixed during the whole procedure.

After each sampling iteration, the optimal occupancies of generated conformations are determined as follows. Each conformation is transformed into a density given by

$$\rho(r) = 8/r \int_{s_{min}}^{s_{max}} f(s) \exp(-Bs^2) \sin(4\pi rs) ds$$

where  $r$  is the position compared to the atom position,  $f(s)$  is the atomic scattering factor as a function of the momentum vector of the atom,  $B$  is the isotropic B-factor, and  $s_{min}$  and  $s_{max}$  are the minimum and maximum momentum vector length. For computational efficiency a lookup table is created for each atom at an interval of 0.01 Å.



A combined mask is calculated by forming the union of all individual conformation masks, using a resolution dependent radius, where  $r = 0.5 + R / 3$ . Density values under the footprint of the resulting mask are extracted and used as input for Quadratic Programming (QP) and Mixed Integer Quadratic Programming (MIQP) to select up to 5 conformations that best represent the density data locally. QP and MIQP solvers guarantee the global optimal occupancy of each conformation by minimizing the real-space residual given by

$$\begin{aligned} \min_w \quad & \| \rho^o - \sum_i w_i \rho_i^c \|_2 \\ \text{s.t.} \quad & w_i \geq t_{d_{min}}, i = 1, \dots, n \\ & 0 \leq \sum_i w_i \leq 1. \end{aligned} \quad (1)$$

where  $\rho^o$  is the target map,  $\rho_i$  is the calculated density of conformer  $i$ ,  $w_i$  is the weight, or occupancy, of conformer  $i$ , and  $0 \leq t \leq 1$  is a threshold on the occupancy. If  $t = 0$ , the constraints enforce non-negativity for the occupancies, and the program is a QP. The combined occupancy cannot exceed unity. Selecting  $t > 0$  introduces sparsity or threshold constraints, which turns the program (1) into a MIQP. To reduce computational complexity, relevant conformations are pre-selected using QP and used as input for a subsequent round of MIQP.<sup>23</sup>

After the ligand has been fully built up starting from each cluster, all resulting conformations are pooled together for a final round of rescoring. The individual Pearson product-moment cross-correlation (RSCC) is calculated for each individual conformation. Conformations with RSCC score less than 0.9 times the highest correlation, and redundant conformations for which all pairwise atom distances are within those of previously accepted conformations, are discarded. Finally, a parsimonious multiconformer model is created as follows. An MIQP step at an occupancy threshold of 0.20 is performed on the remaining pool of conformations. Visual inspection of conformers at minimum occupancy of 0.2 generally indicated density contours exceeded noise levels<sup>77</sup>. For very high resolution datasets, this threshold can be reduced. The resulting selected conformations are ranked by RSCC value, and starting from the conformation carrying the highest RSCC, additional conformations are added if the RSCC increases under the combined footprint and else discarded. This is performed iteratively until self-consistent, i.e. all conformations increase the RSCC under the combined footprint. The output of *qFit-ligand* consists of all conformations before the final rescoring round and the sparse, occupancy-weighted multiconformer ligand model.

*qFit-ligand* is implemented in Python 2.7 and relies on the open-source NumPy, SciPy and CVXOPT<sup>78</sup> packages and the freely available Community Edition of IBM ILOG CPLEX, with added modules from the mmLib Python toolkit.<sup>79</sup> *qFit-ligand* is released under the MIT license and can be downloaded free of charge from [https://github.com/ExcitedStates/qfit\\_ligand](https://github.com/ExcitedStates/qfit_ligand), where additional documentation and installation instructions can be found.

## Benchmarking *qFit-ligand*

Simulated structure factors were generated from the re-refined benchmark set using *phenix.fmodel* at 1.0, 1.3, 1.6, 2.0 and 2.5 Å resolution. Occupancy of the B-conformer was varied from 0.5 to 0.0 occupancy in 0.1 decrements, requiring that the combined occupancy of the multiconformer ligand summed up to unity. A random error of 10% was added to the amplitudes, a fraction of 0.1 was used for  $R_{\text{free}}$  flags, and we selected  $k_{\text{sol}} = 0.4$  and  $b_{\text{sol}} = 45 \text{ Å}^2$ . The B-factor of each ligand atom was adjusted by 10 times the difference in real and simulated resolution, i.e. the B-factor was inflated for lower resolution simulated data and sharpened for high resolution data. The resulting *mtz* files were converted into *ccp4* density files using *phenix.mtz2map*.

For each resolution and occupancy combination, *qFit-ligand* was run on the full benchmark set using different sampling parameters: sampling 1 degree of freedom (DoF) using a torsion sampling interval of 1 degree, and sampling 2 DoFs simultaneously using a torsion sampling interval of 6 degrees. To obtain a uniform metric for evaluating the performance of *qFit-ligand*, we therefore analyzed the results by calculating a normalized RMSD, calculated in the receptor frame of reference<sup>80,81</sup>:

$$\text{nRMSD} = \min_{C_i} \frac{\text{RMSD}(B, C_i)}{\text{RMSD}(A, B)}$$

between the output structures  $C_i$  of *qFit-ligand* and the ligand  $B$  conformer used during the structure factor generation, normalized by the RMSD between the A and B conformer. Conformers  $C_i$  for which  $\text{nRMSD} < 0.5$  are more similar to the B conformer than the A conformer. The nRMSD measure puts equal weight on each benchmark case by normalization, while providing uniform thresholds for determining success (Figure S14). We note that the RMSD (and therefore also the nRMSD) measure has important limitations, affecting the results. For example, since RMSD is typically calculated between unique points, it fails to account for symmetric transformations within a ligand, e.g. a 180° flipped acidic group or aromatic ring, thus introducing additional error. Unfortunately, to our knowledge there is no readily available method that addresses this issue. While a low nRMSD guarantees a good solution, a relatively high nRMSD can still suggest an alternate conformation that requires some manual adjustments, including symmetric transformations.

We subsequently evaluated the performance of *qFit-ligand* for each category in the benchmark set, at simulated resolutions and occupancies, after refinement with PHENIX v1.11 using the same protocol as stated above for re-refinement. The final selection stage was optimized heuristically by finding a balance between maximizing sensitivity and minimizing false-positives against the benchmark.

## D3R and Twilight database investigation

We downloaded the D3R (<https://drugdesigndata.org/about/datasets>) and Twilight (<http://www.ruppweb.org/twilight/newligands-2016.tsv.bz2>) data bases deposited in 2016. For all cases, structure coordinates and  $2mF_o - DF_c$  maps were downloaded from the PDB. For the

Twilight cases, we discarded structures with a reported resolution worse than 2.0 Å and ligands consisting of less than 15 and more than 35 non-hydrogen atoms and a cross-correlation score less than 0.6. The resulting *qFit-ligand* multiconformer models were ranked on the Fisher z-transformation score. The Fisher z-transformation is a mapping of the RSCC  $r$  so that its distribution is approximately normal. It was previously applied to significance testing for cryo-EM rigid body fitting<sup>82–84</sup> and is given by

$$z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)$$

where  $r$  is the RSCC. The standard error of  $z$  is given by

$$\sigma = \frac{1}{\sqrt{N-3}}$$

where  $N$  represents the number of independent observations, approximated by  $N = MV/(c \times R)$ , where  $MV$  represents the molecular volume of the ligand in Å<sup>3</sup>,  $R$  the resolution in Å, and  $c = 1$  Å<sup>2</sup> a constant<sup>83</sup>. Starting with the conformer with the highest RSCC, additional conformations are added. Our approach is based on hypothesis testing of the normally distributed, random variable. For each additional conformer, the  $z$ -score is recalculated for both the starting conformation and the combined multiconformer under the footprint of the latter. The resulting difference in  $z$ -score is divided by its standard error to provide a resolution and size corrected measure of significance for the increase in cross-correlation, where higher is better. The test statistic is

where  $z_{best}$  is the recalculated  $z$ -score of the single conformer with the highest RSCC, and  $z_i^{\Sigma}$  is the  $z$ -score of the  $i$ -th combined multi-conformer. The highest  $z$ -score increase found during this iterative process is reported in Table S3.

### EDIAm and OPIA Calculation

$$EDIA_m(U) = \sum_{U_X \in U} \left( \text{occ}(U_X) \times \left[ \left( \frac{1}{|U_X|} \sum_{a \in U_X} (\text{EDIA}(a) - 0.1)^{-1/2} \right)^{-2} + 0.1 \right] \right)$$

where  $U_X$  is a conformer of the ligand  $U$ ,  $a$  is an atom in  $U_X$  with an EDIA score of  $\text{EDIA}(a)$ , and  $\text{occ}(U_X)$  is the occupancy of that conformer. The OPIA of multiconformer mode was calculated as an occupancy-weighted average of the OPIA of each conformer.

### Ligand Energy

To assess ligand conformational energies, we carried out constrained minimizations (flat bottom width of 0.2 Å) of all ligands (from single conformer and *qFit-ligand* multiconformer models) using Jaguar<sup>85</sup> with the M06–2X functional<sup>86,87</sup> and 6–31G(d,p) basis set, except for bromine atoms which used the LAV2P basis set. Gas phase energies of ligand conformations generated by *qFit-ligand* (which were subsequently refined) were compared

to ligand conformations in the single conformer model (i.e., 0 kcal/mol). The relative energies of each alternative conformation in the *qFit-ligand* multiconformer model were multiplied by their respective occupancies, then summed to arrive at the occupancy-weighted ligand energy.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgment

We gratefully acknowledge support from Relay Therapeutics. D.A.K. was supported by an A.P. Giannini Postdoctoral Fellowship. R.F. is supported by a Novo Nordisk Foundation and Stanford Bio-X Program fellowship NNF15OC0015268. J.S.F. is supported by a Searle Scholar Award from the Kinship Foundation, a Pew Scholar Award from the Pew Charitable Trusts, a Packard Fellowship from the David and Lucile Packard Foundation, NIH GM124149, UC Office of the President Laboratory Fees Research Program LFR-17-476732, and NSF STC-1231306. S.H.P.d.O., J.S.F., and H.v.d.B. are supported by NIH GM123159.

## Abbreviations Used

<b>BRD4</b>	bromodomain-containing protein 4
<b>CBP</b>	CREB binding protein
<b>D3R</b>	Drug Design Data Resource
<b>EDIAm</b>	Electron Density support for Individual Atoms RSCC, Pearson product-moment cross-correlation

## References

- (1). Tzeng S-R; Kalodimos CG Protein Activity Regulation by Conformational Entropy. *Nature* 2012, 19, 236–240.
- (2). Caro JA; Harpole KW; Kasinath V; Lim J; Granja J; Valentine KG; Sharp KA; Wand AJ Entropy in Molecular Recognition by Proteins. *Proc. Natl. Acad. Sci. U. S. A* 2017, 114 (25), 6563–6568. [PubMed: 28584100]
- (3). Mobley DL; Dill KA Binding of Small-Molecule Ligands to Proteins: “What You See” Is Not Always “What You Get.” *Structure* 2009, 17 (4), 489–498. [PubMed: 19368882]
- (4). van den Bedem H; Fraser JS Integrative, Dynamic Structural Biology at Atomic Resolution—it’s about Time. *Nat. Methods* 2015, 12 (4), 307–318. [PubMed: 25825836]
- (5). Smith JL; Hendrickson WA; Honzatko RB; Sheriff S. Structural Heterogeneity in Protein Crystals. *Biochemistry* 1986, 25 (18), 5018–5027. [PubMed: 3768328]
- (6). Lang PT; Ng H-L; Fraser JS; Corn JE; Echols N; Sales M; Holton JM; Alber T. Automated Electron-Density Sampling Reveals Widespread Conformational Polymorphism in Proteins. *Protein Sci* 2010, 19 (7), 1420–1431. [PubMed: 20499387]
- (7). Burnley BT; Afonine PV; Adams PD; Gros P. Modelling Dynamics in Protein Crystal Structures by Ensemble Refinement. *Elife* 2012, 1, 1–29.
- (8). Carroll MJ; Mauldin RV; Gromova AV; Singleton SF; Collins EJ; Lee AL Evidence for Dynamics in Proteins as a Mechanism for Ligand Dissociation. *Nat. Chem. Biol* 2012, 8 (3), 246–252. [PubMed: 22246400]
- (9). Carroll MJ; Gromova AV; Miller KR; Tang H; Wang XS; Tripathy A; Singleton SF; Collins EJ; Lee AL Direct Detection of Structurally Resolved Dynamics in a Multiconformation Receptor-Ligand Complex. *J. Am. Chem. Soc* 2011, 133 (16), 6422–6428. [PubMed: 21469679]

- (10). Martinek TA; Otvös F; Dervarics M; Tóth G; Fülöp F. Ligand-Based Prediction of Active Conformation by 3D-QSAR Flexibility Descriptors and Their Application in 3+3D-QSAR Models. *J. Med. Chem* 2005, 48 (9), 3239–3250. [PubMed: 15857130]
- (11). Kuhnert M; Köster H; Bartholomäus R; Park AY; Shahim A; Heine A; Steuber H; Klebe G; Diederich WE Tracing Binding Modes in Hit-to-Lead Optimization: Chameleon-like Poses of Aspartic Protease Inhibitors. *Angew. Chem. Int. Ed Engl* 2015, 54 (9), 2849–2853. [PubMed: 25630461]
- (12). Kozakov D; Hall DR; Jehle S; Jehle S; Luo L; Ochiana SO; Jones EV; Pollastri M; Allen KN; Whitty A; Vajda S. Ligand Deconstruction: Why Some Fragment Binding Positions Are Conserved and Others Are Not. *Proc. Natl. Acad. Sci. U. S. A* 2015, 112 (20), E2585–E2594.
- (13). Han X; Jiang M; Zhou C; Zhou Z; Xu Z; Wang L; Mayweg AV; Niu R; Jin T-G; Yang S. Discovery of Potent and Selective CDK8 Inhibitors through FBDD Approach. *Bioorg. Med. Chem. Lett* 2017, 27 (18), 4488–4492. [PubMed: 28802632]
- (14). Casale E; Amboldi N; Brasca MG; Caronni D; Colombo N; Dalvit C; Felder ER; Fogliatto G; Galvani A; Isacchi A; Polucci P; Riceputi L; Sola F; Visco C; Zuccotto F; Casuscelli F. Fragment-Based Hit Discovery and Structure-Based Optimization of Aminotriazoloquinazolines as Novel Hsp90 Inhibitors. *Bioorg. Med. Chem* 2014, 22 (15), 4135–4150. [PubMed: 24980703]
- (15). Mpamhanga CP; Spinks D; Tulloch LB; Shanks EJ; Robinson DA; Collie IT; Fairlamb AH; Wyatt PG; Fearson JA; Hunter WN; Gilbert IH; Brenk R. One Scaffold, Three Binding Modes: Novel and Selective Pteridine Reductase 1 Inhibitors Derived from Fragment Hits Discovered by Virtual Screening. *J. Med. Chem* 2009, 52 (14), 4454–4465. [PubMed: 19527033]
- (16). Woolford AJ-A; Pero JE; Aravapalli S; Berdini V; Coyle JE; Day PJ; Dodson AM; Grondin P; Holding FP; Lee LYW; Li P; Manas ES; Marino J Jr; Martin ACL; McClelland BW; McMenamin RL; Murray CW; Neipp CE; Page LW; Patel VK; Potvain F; Rich S; Rivero RA; Smith K; Somers DO; Trotter L; Velagaleti R; Williams G; Xie R. Exploitation of a Novel Binding Pocket in Human Lipoprotein-Associated Phospholipase A2 (Lp-PLA2) Discovered through X-Ray Fragment Screening. *J. Med. Chem* 2016, 59 (11), 5356–5367. [PubMed: 27167608]
- (17). Plattner N; Noé F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat. Commun* 2015, 6, 7653. [PubMed: 26134632]
- (18). Porter NJ; Mahendran A; Breslow R; Christianson DW Unusual Zinc-Binding Mode of HDAC6-Selective Hydroxamate Inhibitors. *Proc. Natl. Acad. Sci. U. S. A* 2017, 114 (51), 13459–13464.
- (19). Bardelle C; Coleman T; Cross D; Davenport S; Kettle JG; Ko EJ; Leach AG; Mortlock A; Read J; Roberts NJ; Robins P; Williams EJ Inhibitors of the Tyrosine Kinase EphB4. Part 2: Structure-Based Discovery and Optimisation of 3,5-Bis Substituted Anilinopyrimidines. *Bioorg. Med. Chem. Lett* 2008, 18 (21), 5717–5721. [PubMed: 18851911]
- (20). Lee C-J; Liang X; Wu Q; Najeeb J; Zhao J; Gopalaswamy R; Titecat M; Sebbane F; Lemaitre N; Toone EJ; Zhou P. Drug Design from the Cryptic Inhibitor Envelope. *Nat. Commun* 2016, 7, 10638.
- (21). van den Bedem H; Bhabha G; Yang K; Wright PE; Fraser JS Automated Identification of Functional Dynamic Contact Networks from X-Ray Crystallography. *Nat Meth* 2013, 10 (August), 896–901.
- (22). van den Bedem H; Dhanik A; Latombe J-C; Deacon AM Modeling Discrete Heterogeneity in X-Ray Diffraction Data by Fitting Multi-Conformers. *Acta Crystallogr* 2009, D65 (10), 1107–1117.
- (23). Keedy DA; Fraser JS; van den Bedem H. Exposing Hidden Alternative Backbone Conformations in X-Ray Crystallography Using qFit. *PLoS Comput. Biol* 2015, 11 (10), e1004507.
- (24). Levin EJ; Kondrashov DA; Wesenberg GE; Phillips GN Jr. Ensemble Refinement of Protein Crystal Structures. *Structure* 2007, 15 (9), 1040–1052. [PubMed: 17850744]
- (25). Diller DJ; Pohl E; Redinbo MR; Hovey BT; Hol WG A Rapid Method for Positioning Small Flexible Molecules, Nucleic Acids, and Large Protein Fragments in Experimental Electron Density Maps. *Proteins* 1999, 36 (4), 512–525. [PubMed: 10450093]
- (26). Oldfield TJ X-LIGAND: An Application for the Automated Addition of Flexible Ligands into Electron Density. *Acta Crystallogr. D Biol. Crystallogr* 2001, 57 (Pt 5), 696–705. [PubMed: 11320310]

- (27). Zwart PH; Langer GG; Lamzin VS Modelling Bound Ligands in Protein Crystal Structures. *Acta Crystallogr. D Biol. Crystallogr* 2004, 60 (Pt 12 Pt 1), 2230–2239. [PubMed: 15572776]
- (28). Aishima J; Russel DS; Guibas LJ; Adams PD; Brunger AT Automated Crystallographic Ligand Building Using the Medial Axis Transform of an Electron-Density Isosurface. *Acta Crystallogr. D Biol. Crystallogr* 2005, 61 (Pt 10), 1354–1363. [PubMed: 16204887]
- (29). Terwilliger TC; Klei H; Adams PD; Moriarty NW; Cohn JD Automated Ligand Fitting by Core-Fragment Fitting and Extension into Density. *Acta Crystallogr. D Biol. Crystallogr* 2006, 62 (Pt 8), 915–922. [PubMed: 16855309]
- (30). Evrard GX; Langer GG; Perrakis A; Lamzin VS Assessment of Automatic Ligand Building in ARP/wARP. *Acta Crystallogr. D Biol. Crystallogr* 2007, 63 (Pt 1), 108–117. [PubMed: 17164533]
- (31). Janowski PA; Moriarty NW; Kelley BP; Case DA; York DM; Adams PD; Warren GL Improved Ligand Geometries in Crystallographic Refinement Using AFITT in PHENIX. *Acta Crystallogr. D Struct Biol* 2016, 72 (Pt 9), 1062–1072. [PubMed: 27599738]
- (32). Weichenberger CX; Pozharski E; Rupp B. Visualizing Ligand Molecules in Twilight Electron Density. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun* 2013, 69 (Pt 2), 195–200.
- (33). 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; Zwart PH PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr* 2010, D66, 213–221.
- (34). Bhat TN Correlation between Occupancy and Temperature Factors of Solvent Molecules in Crystal Structures of Proteins. *Acta Crystallogr. A* 1989, 45 ( Pt 1), 145–146. [PubMed: 2619964]
- (35). Carugo O. Correlation between Occupancy and B Factor of Water Molecules in Protein Crystal Structures. *Protein Eng. Des. Sel* 1999, 12 (12), 1021–1024.
- (36). Meyder A; Nittinger E; Lange G; Klein R; Rarey M. Estimating Electron Density Support for Individual Atoms and Molecular Fragments in X-Ray Structures. *J. Chem. Inf. Model* 2017, 57 (10), 2437–2447. [PubMed: 28981269]
- (37). Lewis RT; Bode CM; Choquette DM; Potashman M; Romero K; Stellwagen JC; Teffera Y; Moore E; Whittington DA; Chen H; Epstein LF; Emkey R; Andrews PS; Yu VL; Saffran DC; Xu M; Drew A; Merkel P; Szilvassy S; Brake RL The Discovery and Optimization of a Novel Class of Potent, Selective, and Orally Bioavailable Anaplastic Lymphoma Kinase (ALK) Inhibitors with Potential Utility for the Treatment of Cancer. *J. Med. Chem* 2012, 55 (14), 6523–6540. [PubMed: 22734674]
- (38). Perola E; Charifson PS Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem* 2004, 47 (10), 2499–2510. [PubMed: 15115393]
- (39). Young T; Abel R; Kim B; Berne BJ; Friesner RA Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A* 2007, 104 (3), 808–813. [PubMed: 17204562]
- (40). Klebe G. Applying Thermodynamic Profiling in Lead Finding and Optimization. *Nat. Rev. Drug Discov* 2015, 14 (2), 95–110. [PubMed: 25614222]
- (41). Howard JAK; Hoy VJ; O'Hagan D; Smith GT How Good Is Fluorine as a Hydrogen Bond Acceptor? *Tetrahedron* 1996, 52 (38), 12613–12622.
- (42). Kuhn B; Guba W; Hert J; Banner D; Bissantz C; Ceccarelli S; Haap W; Körner M; Kuglstat A; Lerner C; Mattei P; Neidhart W; Pinard E; Rudolph MG; Schulz-Gasch T; Woltering T; Stahl M. A Real-World Perspective on Molecular Design. *J. Med. Chem* 2016, 59 (9), 4087–4102. [PubMed: 26878596]
- (43). Filippakopoulos P; Picaud S; Mangos M; Keates T; Lambert J-P; Barsyte-Lovejoy D; Felletar I; Volkmer R; Müller S; Pawson T; Gingras A-C; Arrowsmith CH; Knapp S. Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* 2012, 149 (1), 214–231. [PubMed: 22464331]
- (44). Arrowsmith CH; Bountra C; Fish PV; Lee K; Schapira M. Epigenetic Protein Families: A New Frontier for Drug Discovery. *Nat. Rev. Drug Discov* 2012, 11 (5), 384–400. [PubMed: 22498752]

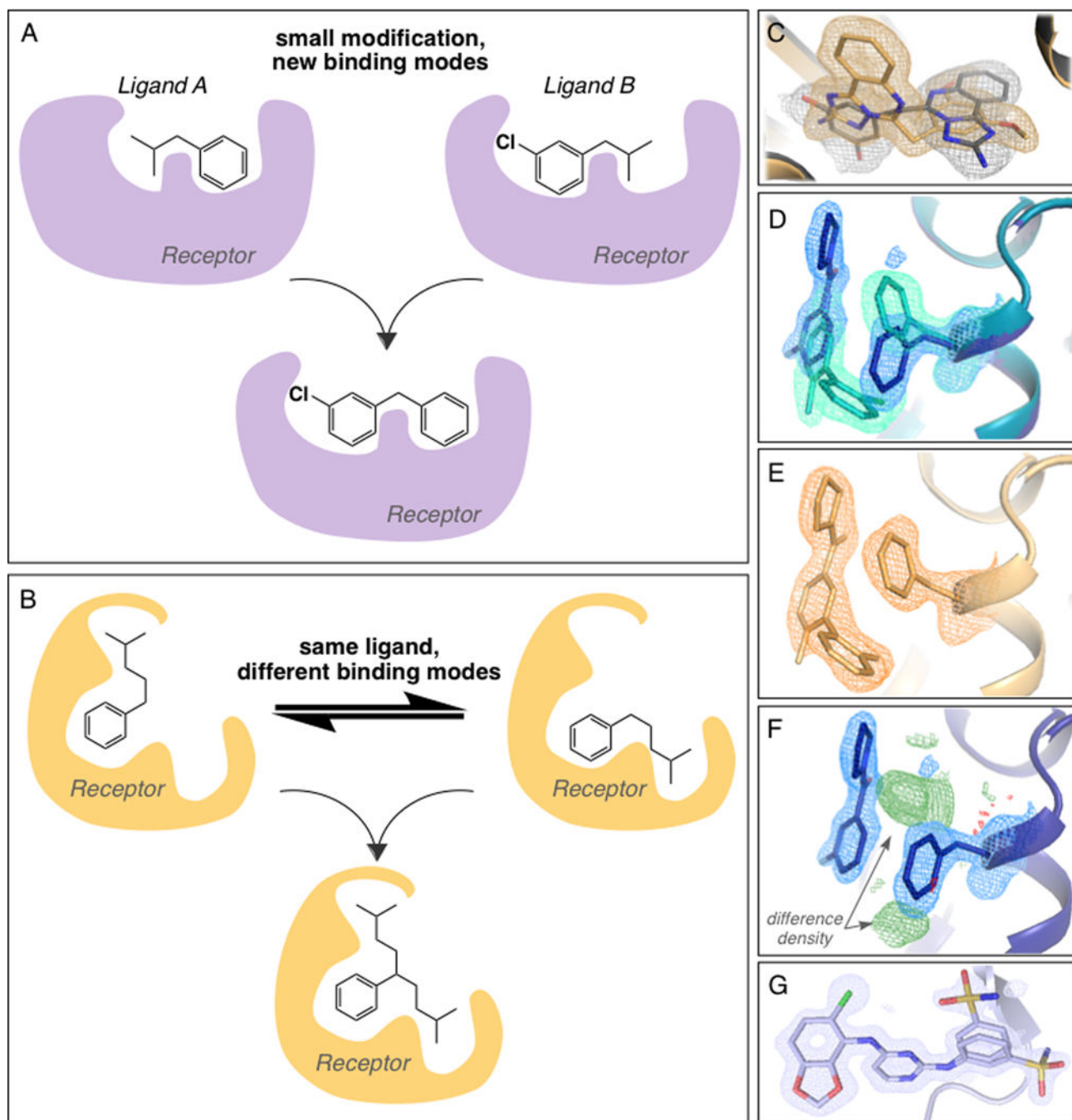


- (45). Rooney TPC; Filippakopoulos P; Fedorov O; Picaud S; Cortopassi WA; Hay DA; Martin S; Tumber A; Rogers CM; Philpott M; Wang M; Thompson AL; Heightman TD; Pryde DC; Cook A; Paton RS; Müller S; Knapp S; Brennan PE; Conway SJ A Series of Potent CREBBP Bromodomain Ligands Reveals an Induced-Fit Pocket Stabilized by a Cation- $\pi$  Interaction. *Angew. Chem. Int. Ed Engl* 2014, 53 (24), 6126–6130. [PubMed: 24821300]
- (46). Chekler ELP; Pellegrino JA; Lanz TA; Denny RA; Flick AC; Coe J; Langille J; Basak A; Liu S; Stock IA; Sahasrabudhe P; Bonin PD; Lee K; Pletcher MT; Jones LH Transcriptional Profiling of a Selective CREB Binding Protein Bromodomain Inhibitor Highlights Therapeutic Opportunities. *Chem. Biol* 2015, 22 (12), 1588–1596. [PubMed: 26670081]
- (47). Hay DA; Fedorov O; Martin S; Singleton DC; Tallant C; Wells C; Picaud S; Philpott M; Monteiro OP; Rogers CM; Conway SJ; Rooney TPC; Tumber A; Yapp C; Filippakopoulos P; Bunnage ME; Müller S; Knapp S; Schofield CJ; Brennan PE Discovery and Optimization of Small-Molecule Ligands for the CBP/p300 Bromodomains. *J. Am. Chem. Soc* 2014, 136 (26), 9308–9319. [PubMed: 24946055]
- (48). Mirguet O; Lamotte Y; Chung C-W; Bamborough P; Delannée D; Bouillot A; Gellibert F; Krysa G; Lewis A; Witherington J; Huet P; Dudit Y; Trotter L; Nicodeme E. Naphthyridines as Novel BET Family Bromodomain Inhibitors. *ChemMedChem* 2014, 9 (3), 580–589. [PubMed: 24000170]
- (49). Seal J; Lamotte Y; Donche F; Bouillot A; Mirguet O; Gellibert F; Nicodeme E; Krysa G; Kirilovsky J; Beinke S; McCleary S; Rioja I; Bamborough P; Chung C-W; Gordon L; Lewis T; Walker AL; Cutler L; Lugo D; Wilson DM; Witherington J; Lee K; Prinjha RK Identification of a Novel Series of BET Family Bromodomain Inhibitors: Binding Mode and Profile of I-BET151 (GSK1210151A). *Bioorg. Med. Chem. Lett* 2012, 22 (8), 2968–2972. [PubMed: 22437115]
- (50). Fraser JS; van den Bedem H; Samelson AJ; Lang PT; Holton JM; Echols N; Alber T. Accessing Protein Conformational Ensembles Using Room-Temperature X-Ray Crystallography. *Proc. Natl. Acad. Sci. U. S. A* 2011, 108 (39), 16247–16252.
- (51). Fischer M; Shoichet BK; Fraser JS One Crystal, Two Temperatures: Cryocooling Penalties Alter Ligand Binding to Transient Protein Sites. *Chembiochem* 2015, 16 (11), 1560–1564. [PubMed: 26032594]
- (52). Warren GL; Do TD; Kelley BP; Nicholls A; Warren SD Essential Considerations for Using Protein-Ligand Structures in Drug Discovery. *Drug Discov. Today* 2012, 17 (23–24), 1270–1281. [PubMed: 22728777]
- (53). 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; Young J. Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop. *Structure* 2016, 24 (4), 502–508. [PubMed: 27050687]
- (54). Berman HM; Westbrook J; Feng Z; Gilliland G; Bhat TN; Weissig H; Shindyalov IN; Bourne PE The Protein Data Bank. *Nucleic Acids Res* 2000, 28 (1), 235–242. [PubMed: 10592235]
- (55). Harder E; Damm W; Maple J; Wu C; Reboul M; Xiang JY; Wang L; Lupyan D; Dahlgren MK; Knight JL; Kaus JW; Cerutti DS; Krilov G; Jorgensen WL; Abel R; Friesner RA OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput* 2016, 12 (1), 281–296. [PubMed: 26584231]
- (56). Watts KS; Dalal P; Murphy RB; Sherman W; Friesner RA; Shelley JC ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J. Chem. Inf. Model* 2010, 50 (4), 534–546. [PubMed: 20373803]
- (57). Watts KS; Dalal P; Tebben AJ; Cheney DL; Shelley JC Macrocycle Conformational Sampling with MacroModel. *J. Chem. Inf. Model* 2014, 54 (10), 2680–2696. [PubMed: 25233464]
- (58). Hawkins PCD; Skillman AG; Warren GL; Ellingson BA; Stahl MT Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank

and Cambridge Structural Database. *J. Chem. Inf. Model* 2010, 50 (4), 572–584. [PubMed: 20235588]

- (59). Labute P. LowModeMD--Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops. *J. Chem. Inf. Model* 2010, 50 (5), 792–800. [PubMed: 20429574]
- (60). Ebejer J-P; Morris GM; Deane CM Freely Available Conformer Generation Methods: How Good Are They? *J. Chem. Inf. Model* 2012, 52 (5), 1146–1158. [PubMed: 22482737]
- (61). Pearce NM; Krojer T; Bradley AR; Collins P; Nowak RP; Talon R; Marsden BD; Kelm S; Shi J; Deane CM; von Delft F. A Multi-Crystal Method for Extracting Obscured Crystallographic States from Conventionally Uninterpretable Electron Density. *Nat. Commun* 2017, 8, 15123.
- (62). Merk A; Bartesaghi A; Banerjee S; Falconieri V; Rao P; Davis MI; Pragani R; Boxer MB; Earl LA; Milne JLS; Subramaniam S. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* 2016, 165 (7), 1698–1707. [PubMed: 27238019]
- (63). Cheng A; Tan YZ; Dandey VP; Potter CS; Carragher B. Strategies for Automated CryoEM Data Collection Using Direct Detectors In *Methods in Enzymology*; Crowther RA, Ed.; *Methods in Enzymology*; Academic Press: London, England, 2016; Vol. 579, pp 87–102. [PubMed: 27572724]
- (64). Stellato F; Oberthür D; Liang M; Bean R; Gati C; Yefanov O; Barty A; Burkhardt A; Fischer P; Galli L; Kirian RA; Meyer J; Panneerselvam S; Yoon CH; Chervinskii F; Speller E; White TA; Betzel C; Meents A; Chapman HN Room-Temperature Macromolecular Serial Crystallography Using Synchrotron Radiation. *IUCrJ* 2014, 1 (Pt 4), 204–212.
- (65). Tenboer J; Basu S; Zatsepin N; Pande K; Milathianaki D; Frank M; Hunter M; Boutet S; Williams GJ; Koglin JE; Oberthuer D; Heymann M; Kupitz C; Conrad C; Coe J; Roy-Chowdhury SU; James D; Wang D; Grant T; Barty A; Yefanov O; Scales J; Gati C; Seuring C; Srajer V; Henning R; Schwander P; Fromme R; Ourmazd A; Moffat K; Thor JJV; Spence JCH; Fromme P; Chapman HN; Schmidt M. Time-Resolved Serial Crystallography Captures High-Resolution Intermediates of Photoactive Yellow Protein. *Science* 2014, 346, 1242–1246. [PubMed: 25477465]
- (66). Kupitz C; Olmos JL Jr; Holl M; Tremblay L; Pande K; Pandey S; Oberthür D; Hunter M; Liang M; Aquila A; Tenboer J; Calvey G; Katz A; Chen Y; Wiedorn MO; Knoska J; Meents A; Majriani V; Norwood T; Poudyal I; Grant T; Miller MD; Xu W; Tolstikova A; Morgan A; Metz M; Martin-Garcia JM; Zook JD; Roy-Chowdhury S; Coe J; Nagarathnam N; Meza D; Fromme R; Basu S; Frank M; White T; Barty A; Bajt S; Yefanov O; Chapman HN; Zatsepin N; Nelson G; Weierstall U; Spence J; Schwander P; Pollack L; Fromme P; Ourmazd A; Phillips GN Jr; Schmidt M. Structural Enzymology Using X-Ray Free Electron Lasers. *Struct Dyn* 2017, 4 (4), 044003.
- (67). Olmos J; Pandey S; Martin-Garcia JM; Calvey G; Katz A; Knoska J; Kupitz C; Hunter MS; Liang M; Oberthuer D; Others. Authentic Enzyme Intermediates Captured“ on-the-Fly” by Mix-and-Inject Serial Crystallography. *bioRxiv* 2017, 202432.
- (68). Velazquez-Campoy A; Todd MJ; Freire E. HIV-1 Protease Inhibitors: Enthalpic versus Entropic Optimization of the Binding Affinity. *Biochemistry* 2000, 39 (9), 2201–2207. [PubMed: 10694385]
- (69). Chodera JD; Mobley DL Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. *Annu. Rev. Biophys* 2013, 42, 121–142. [PubMed: 23654303]
- (70). You W; Huang Y-MM; Kizhake S; Natarajan A; Chang C-EA Characterization of Promiscuous Binding of Phosphor Ligands to Breast-Cancer-Gene 1 (BRCA1) C-Terminal (BRCT): Molecular Dynamics, Free Energy, Entropy and Inhibitor Design. *PLoS Comput. Biol* 2016, 12 (8), e1005057.
- (71). Huggins DJ; Sherman W; Tidor B. Rational Approaches to Improving Selectivity in Drug Design. *J. Med. Chem* 2012, 55 (4), 1424–1444. [PubMed: 22239221]
- (72). Brock JS; Hamberg M; Balagunaseelan N; Goodman M; Morgenstern R; Strandback E; Samuelsson B; Rinaldo-Matthis A; Haeggström JZ A Dynamic Asp-Arg Interaction Is Essential for Catalysis in Microsomal Prostaglandin E2 Synthase. *Proc. Natl. Acad. Sci. U. S. A* 2016, 113 (4), 972–977. [PubMed: 26755582]

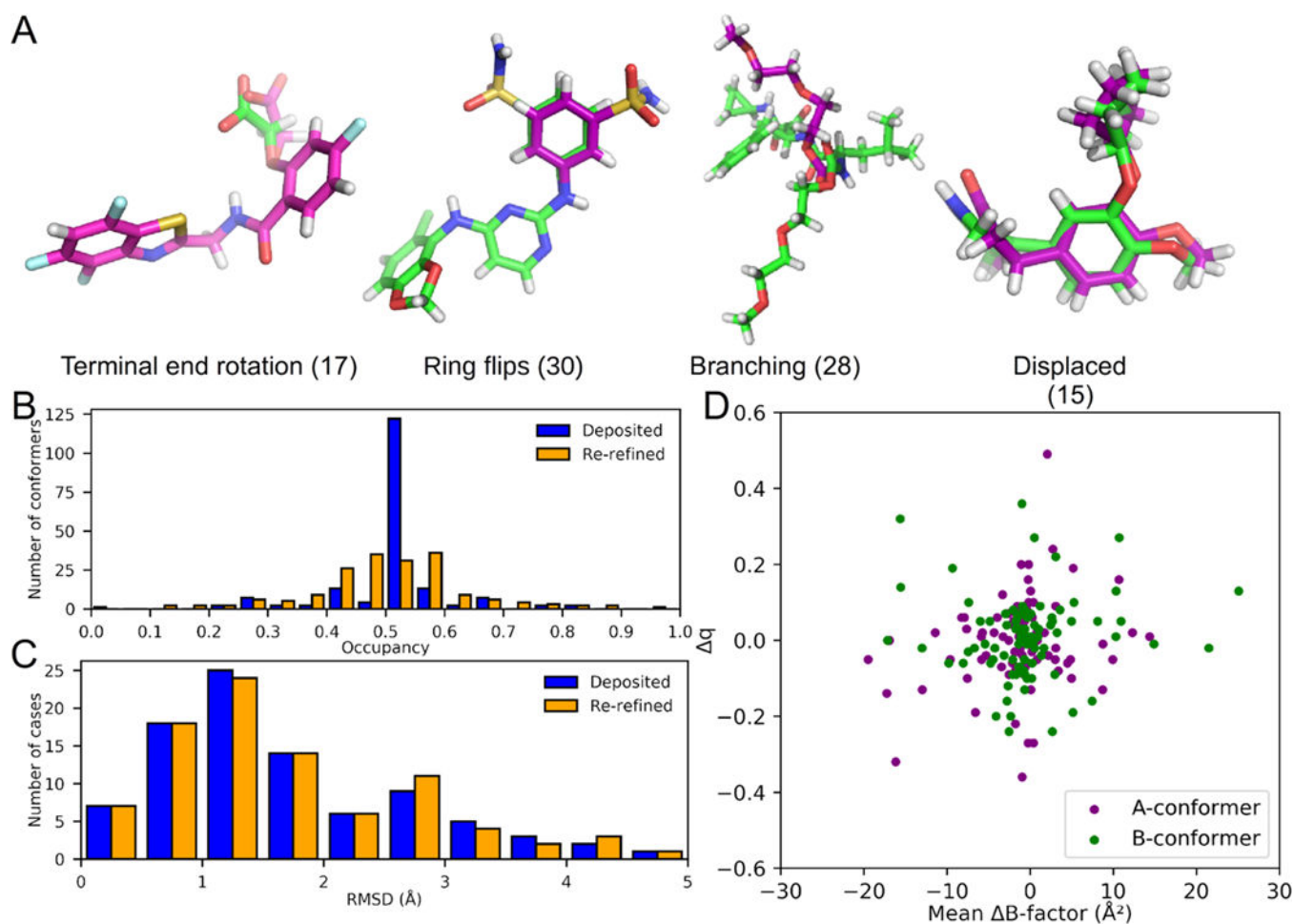
- (73). Terwilliger TC Automated Main-Chain Model Building by Template Matching and Iterative Fragment Extension. *Acta Crystallogr. D Biol. Crystallogr* 2003, 59 (Pt 1), 38–44. [PubMed: 12499537]
- (74). Bertsimas D; Shioda R. Algorithm for Cardinality-Constrained Quadratic Optimization. *Comput. Optim. Appl* 2009, 43 (1), 1–22.
- (75). Miyashiro R; Takano Y. Subset Selection by Mallows' Cp: A Mixed Integer Programming Approach. *Expert Syst. Appl* 2015, 42 (1), 325–331.
- (76). Lemmen C; Lengauer T; Klebe G. FLEXS: A Method for Fast Flexible Ligand Superposition. *J. Med. Chem* 1998, 41 (23), 4502–4520. [PubMed: 9804690]
- (77). Lang PT; Holton JM; Fraser JS; Alber T. Protein Structural Ensembles Are Revealed by Redefining X-Ray Electron Density Noise. *Proc. Natl. Acad. Sci. U. S. A* 2014, 111 (1), 237–242. [PubMed: 24363322]
- (78). Andersen MS; Dahl J; Vandenberghe L. CVXOPT: A Python Package for Convex Optimization, Version 1.1. 6. Available at [cvxopt.org](http://cvxopt.org) 2013, 54.
- (79). Painter J; Merritt EA mmLib Python Toolkit for Manipulating Annotated Structural Models of Biological Macromolecules. *J. Appl. Crystallogr* 2004, 37 (1), 174–178.
- (80). Shen H; Chou JJ MemBrain: Improving the Accuracy of Predicting Transmembrane Helices. *PLoS One* 2008, 3 (6), e2399.
- (81). Bray JK; Weiss DR; Levitt M. Optimized Torsion-Angle Normal Modes Reproduce Conformational Changes More Accurately than Cartesian Modes. *Biophys. J* 2011, 101 (12), 2966–2969. [PubMed: 22208195]
- (82). Fisher RA On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample. *Metron* 1921, 1, 3–32.
- (83). Volkman N. Confidence Intervals for Fitting of Atomic Models into Low-Resolution Densities. *Acta Crystallogr. D Biol. Crystallogr* 2009, 65 (Pt 7), 679–689. [PubMed: 19564688]
- (84). van Zundert GCP; Bonvin AMJJ Defining the Limits and Reliability of Rigid-Body Fitting in Cryo-EM Maps Using Multi-Scale Image Pyramids. *J. Struct. Biol* 2016, 195 (2), 252–258. [PubMed: 27318041]
- (85). Bochevarov AD; Harder E; Hughes TF; Greenwood JR; Braden DA; Philipp DM; Rinaldo D; Halls MD; Zhang J; Friesner RA Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *Int. J. Quantum Chem* 2013, 113 (18), 2110–2142.
- (86). Zhao Y; Truhlar DG The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06 Functionals and 12 Other Functionals. *Theor. Chem. Acc* 2008, 119 (5–6), 525–525.
- (87). Zhao Y; Truhlar DG Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res* 2008, 41 (2), 157–167. [PubMed: 18186612]



**Figure 1.**

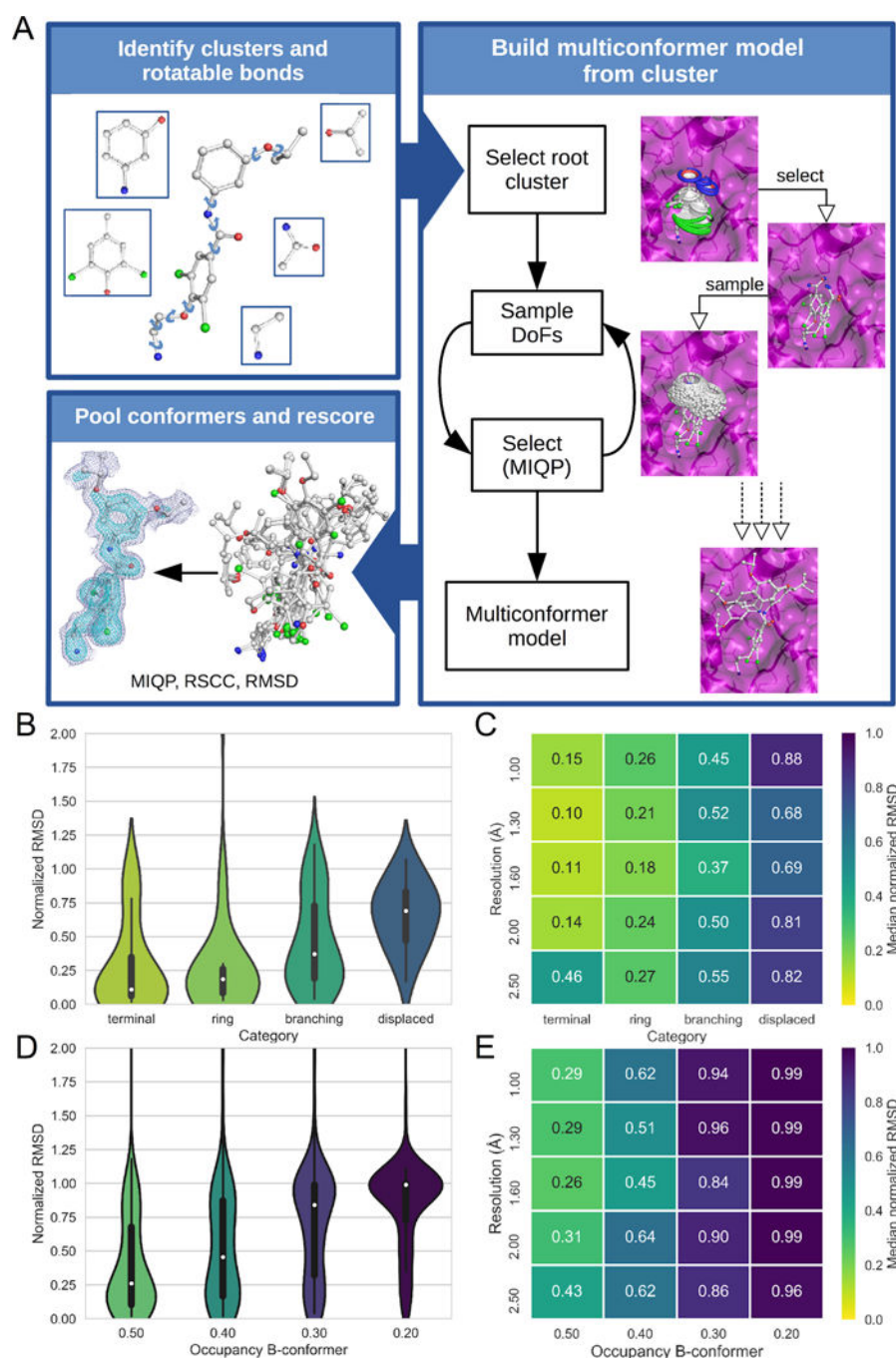
Ligand structural dynamics and minor changes during fragment optimization lead to new binding modes and drive drug design. A) Subtle changes in chemical structure of ligands can impose new binding modes. B) Near isoenergetic receptor-ligand conformations exchange in dynamic equilibrium in crystal structures. These conformations can inform the design of a ligand with higher affinity. C) Hsp90 inhibitors in gray (PDB ID 4CWO) and gold (PDB ID 4CWN). D) Subtle changes in chemical structures lead to changes in binding pose for Lp-PLA2, changing the course of design (PDB IDs 5JAL, 5JAO). E) New Lp-PLA2 inhibitor

designed as a result of observed alternate binding poses of fragments (PDB ID 5JAP). F) Evidence of difference density in x-ray crystal structure of Lp-PLA2 fragment shows alternate binding poses pre-exist at low occupancy (PDB ID 5JAL). G) Alternate conformations exploited in the design of EphB4-binding ligands (PDB ID 2VWX). Electron densities are shown at  $1.5\sigma$ . Positive (green) and negative (red) difference densities in panel F are shown at  $+3.0\sigma$  and  $-3.0\sigma$ , respectively.



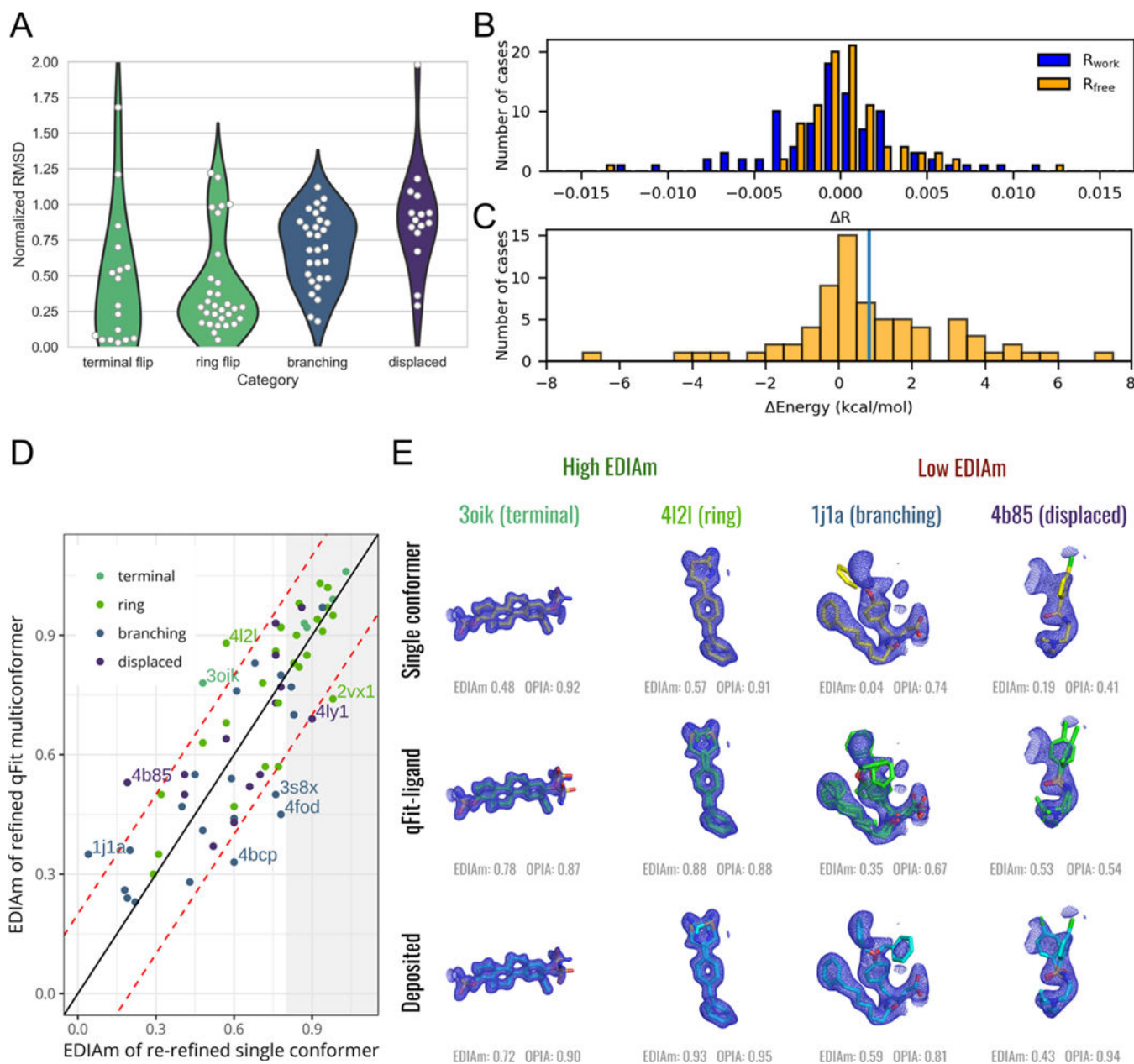
**Figure 2.** Benchmark statistics. A) Categories of alternate conformations present in the benchmark. B) Conformer occupancies pre- (blue) and post- (orange) re-refinement. C) Ligand A to B conformer RMSD, pre (blue) and post (orange) re-refinement. D) Occupancy shift versus mean B-factor difference after re-refinement.



**Figure 3.**

*qFit-ligand* workflow and statistics on synthetic data. A) Rigid clusters are defined as rings or terminal ends, and rotatable bonds as any bond that is not part of a ring. The local search finds possible positions and orientations of each cluster in the binding site, avoiding steric clashes with the protein. As clusters are joined, torsions and degrees of freedom are sampled. Up to 5 ligand conformations that best match the ligand density are selected and combined with the protein model to give a final ligand multiconformer model. B) Violin plot of categories at 1.60 Å resolution and occupancy 0.50 across 90 test cases. The white dot

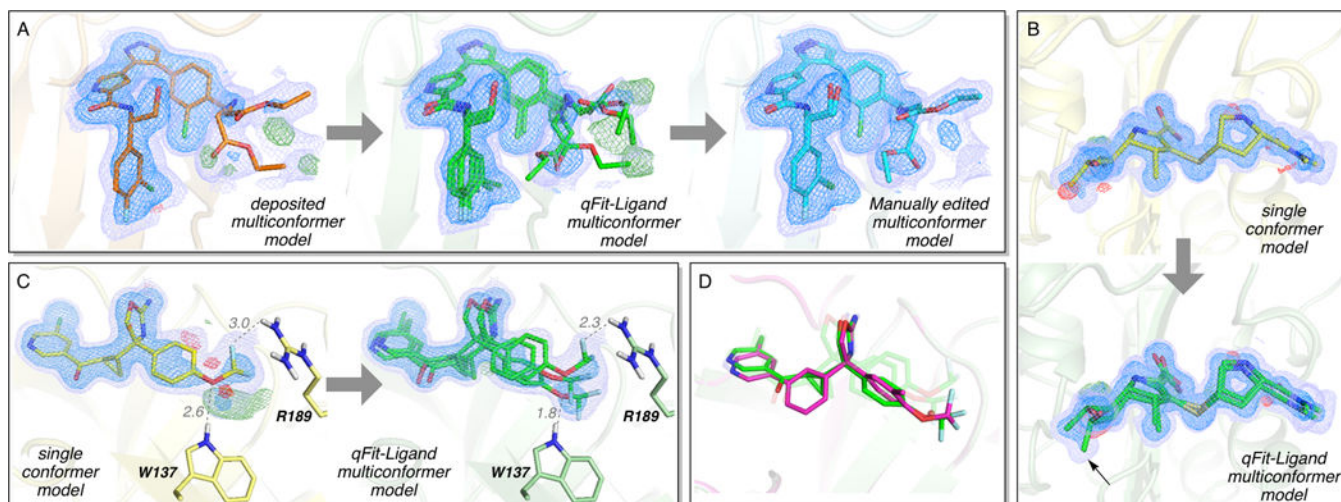
represents the median, the bold center line represent the interquartile range (IQR), and the thin center line represents the percentile range 25<sup>th</sup>-1.5 IQR to 75<sup>th</sup>+1.5 IQR. Lower nRMSD is better. C) Heatmap of category vs. resolution at 0.50 occupancy. Lower nRMSD is better. D) Violin plot for representative resolution 1.60Å and 0.50 / 0.50 occupancy at optimal parameters. E) Heatmap of normalized RMSD at optimal parameters (resolution vs. occupancy).



**Figure 4. Electron density support measures and energetics of *qFit-ligand* multi-conformers are similar to those of single-conformer models for the benchmark set, while reporting on conformational heterogeneity.**

A) nRMSD distributions by category of conformational heterogeneity. The nRMSD of terminal end flips was determined from only atoms affected by the dihedral changes. Lower nRMSD is better. B) Histogram of  $R_{work}$  and  $R_{free}$  differences between refined *qFit-ligand* models and single conformer structures. Negative values indicate a lower R for refined *qFit-ligand* models. C) The distribution of occupancy-weighted, ligand internal energies of *qFit-ligand* multiconformer ligands relative to single 'A' conformation. Negative values indicate that the multiconformer ensemble has lower internal energy than the single benchmark 'A' conformation. Positive values indicate the multiconformer ensemble has higher internal

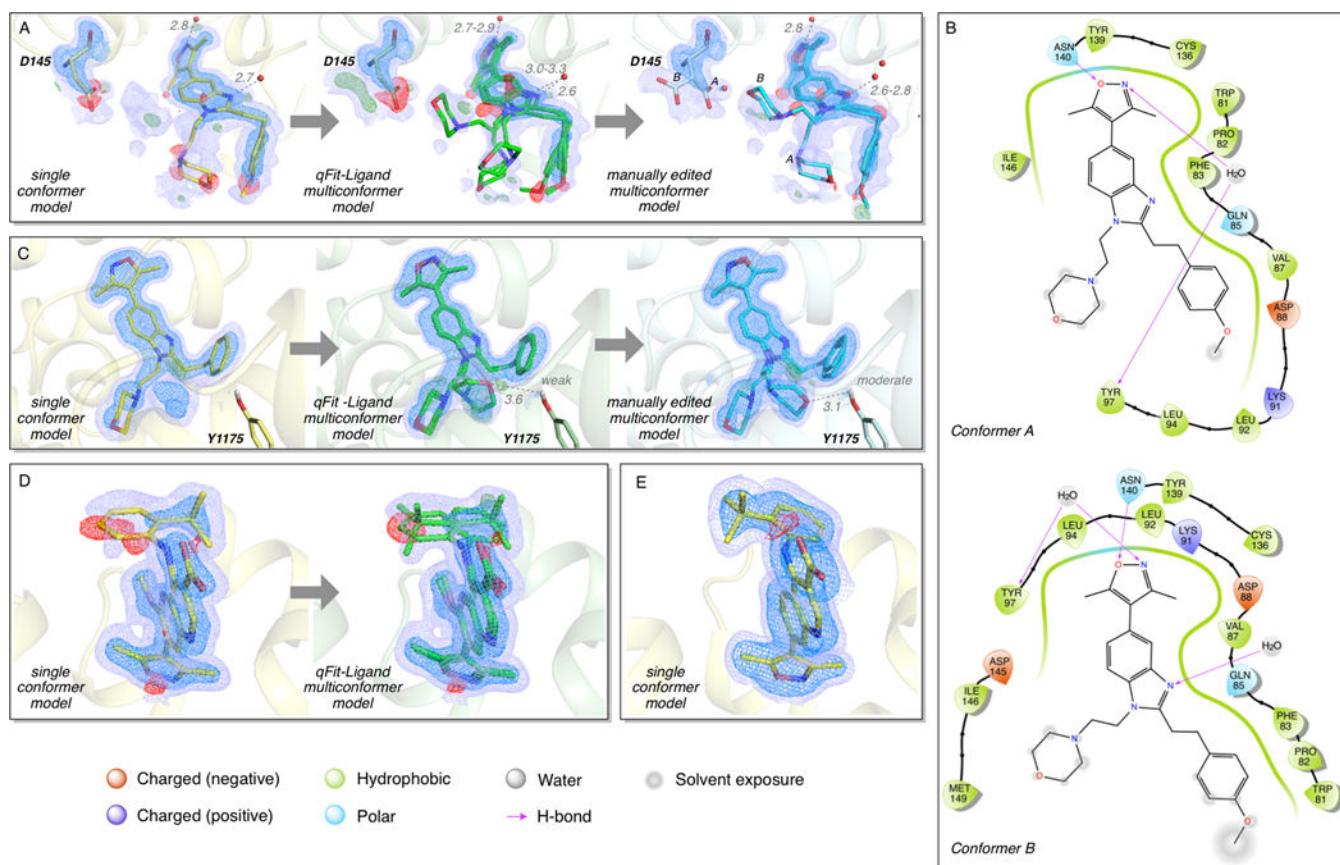
ligand energies. D) Ligand EDIAM scores for re-refined single conformer models against automatically refined *qFit-ligand* multiconformer models. Higher EDIAM scores are better, with scores greater than 0.8 indicating that the model is well-supported by the density (shaded area for single conformer models). E) Examples where the ligand EDIAM was improved by modeling alternate conformations with *qFit-ligand*. The 2mFo-DFc maps are shown in blue at a contour level of  $1\sigma$ .



**Figure 5.**

Prospective discovery of additional conformations and recovered conformations from the D3R and Twilight dataset. A) The deposited multiconformer model, *qFit-ligand* multiconformer model, and manually edited multiconformer model of ERK2 (PDB ID 4FV3; new PDB ID 6DMG). B) Single conformer and *qFit-ligand* multiconformer models of serratia fonticola carbapenemase E166A mutant with the acylenzyme intermediate of meropenem (PDB ID 4EV4; new PDB ID 6DMH). C) Prospective application of *qFit-ligand* to inhibitor 5T5 of BACE-1 (PDB ID 5EZX; new PDB ID 6DMI). Single conformer and *qFit-ligand* multiconformer models shown. D) Overlay of multiconformer model of inhibitor 5T5 (green) with inhibitor 5T6 (magenta) (PDB ID 5EZZ). Electron densities are shown at  $1.5\sigma$  (blue) and  $0.3\sigma$  (purple). Positive (green) and negative (red) difference densities are shown at  $+3.0\sigma$  and  $-3.0\sigma$ , respectively. All structures shown have been refined using Phenix.



**Figure 6.**

Prospective application of *qFit-ligand* to BRD compounds A) Single conformer crystal structure, *qFit-ligand* multiconformer, and manually edited multiconformer models including alternate conformations of Asp145 and compound BDOIA383 bound to BRD4 (PDB ID 5CFW; new PDB ID 6DMJ). B) Protein-ligand interactions of the major, crystal-contact stabilized 'A' and minor 'B' BDOIA383 conformations of the final *qFit-ligand* model (panel 6A). C) Single conformer crystal structure, *qFit-ligand* multiconformer, and manually edited multiconformer models of an isoxazolyl-benzimidazole ligand bound to CBP BRD (PDB ID 4NR5, new PDB ID 6DMK). D) Single conformer and *qFit-ligand* multiconformer models of ligand 9BM bound to BRD4 (PDB ID 4BW3; new PDB ID 6DML). Viewing orientation differs from panel A,C to clearly show ligand alternate conformations. E) Ligand S5B's single binding conformation in BRD2 (PDB ID 4AKN). Electron densities are shown at  $1.5\sigma$  (blue) and  $0.3\sigma$  (purple). Positive (green) and negative (red) difference densities are shown at  $\pm 3.0\sigma$ . Distances in Ångstroms.