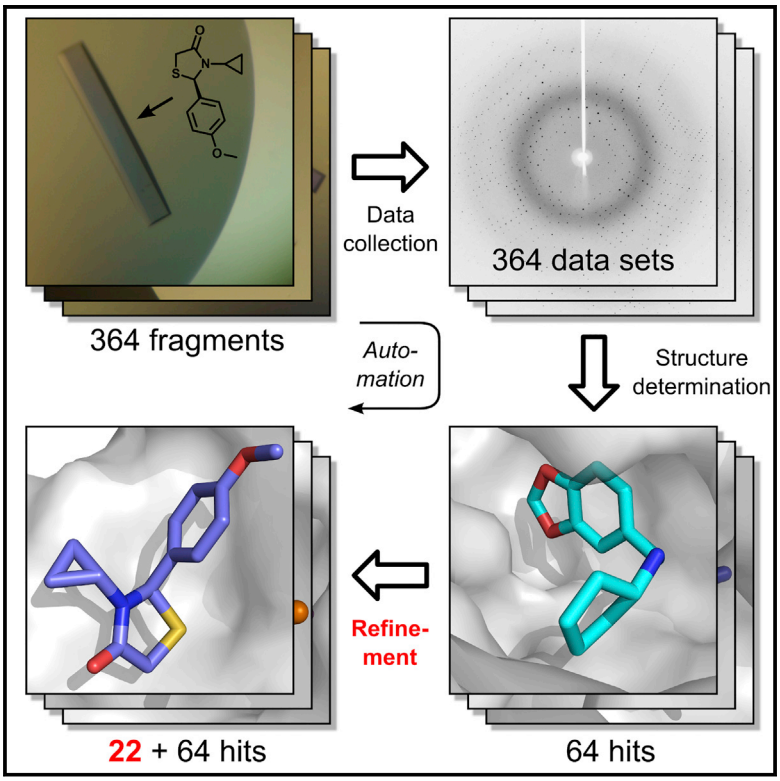
Resource

Structure

High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits

Graphical Abstract



Authors

Johannes Schiebel, Stefan G. Krimmer, Karine Röwer, ..., Uwe Mueller, Gerhard Klebe, Andreas Heine

Correspondence

heinea@mailer.uni-marburg.de

In Brief

Schiebel et al. describe a large collection of diffraction data obtained from protein crystals treated with 364 different fragment molecules. An automated structural refinement of these datasets suggests that more complete refinement enables detection of additional fragments binding in the target's active site.

Highlights

- Reliable fragment hit identification requires phase improvement through refinement
- The modeling of water molecules particularly enhances fragment electron densities
- 364 diffraction datasets are provided for the design and validation of new methods
- An automated procedure was developed to streamline structural refinement

Schiebel et al., 2016, Structure *24*, 1398–1409
August 2, 2016
http://dx.doi.org/10.1016/j.str.2016.06.010

CellPress

CellPress

Structure
Resource

High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits

Johannes Schiebel,[1] Stefan G. Krimmer,[1] Karine Röwer,[2] Anna Knörlein,[1] Xiaojie Wang,[1] Ah Young Park,[1] Martin Stieler,[1] Frederik R. Ehrmann,[1] Kan Fu,[1] Nedyalka Radeva,[1] Michael Krug,[2] Franziska U. Huschmann,[1,2] Steffen Glöckner,[1] Manfred S. Weiss,[2] Uwe Mueller,[2,3] Gerhard Klebe,[1] and Andreas Heine[1,*]

[1]Institute for Pharmaceutical Chemistry, Philipps-University Marburg, Marbacher Weg 6, 35032 Marburg, Germany

[2]Helmholtz-Zentrum Berlin für Materialien und Energie, HZB, BESSY II, Abteilung Makromolekulare Kristallographie, Albert-Einstein-Straße 15, 12489 Berlin, Germany

[3]Present address: MAX IV Laboratory, Lund University, 221 00 Lund, Sweden

*Correspondence: heinea@mailer.uni-marburg.de

http://dx.doi.org/10.1016/j.str.2016.06.010

SUMMARY

Today the identification of lead structures for drug development often starts from small fragment-like molecules raising the chances to find compounds that successfully pass clinical trials. At the heart of the screening for fragments binding to a specific target, crystallography delivers structural information essential for subsequent drug design. While it is common to search for bound ligands in electron densities calculated directly after an initial refinement cycle, we raise the important question whether this strategy is viable for fragments characterized by low affinities. Here, we describe and provide a collection of high-quality diffraction data obtained from 364 protein crystals treated with diverse fragments. Subsequent data analysis showed that ~25% of all hits would have been missed without further refining the resulting structures. To enable fast and reliable hit identification, we have designed an automated refinement pipeline that will inspire the development of optimized tools facilitating the successful application of fragment-based methods.

INTRODUCTION

A major reason for the attrition of new drug candidates during clinical trials are suboptimal physicochemical properties that can result in poor pharmacokinetics and an impaired safety profile (Murray and Rees, 2009). Such characteristics often develop because high-throughput screening (HTS) libraries contain already relatively large and lipophilic molecules. Since hit-to-lead-to-drug optimization typically further increases molecular weight (MW) and lipophilicity, the design of drug-like clinical candidates is hampered (Hann, 2011; Rees et al., 2004). Fragment-based lead discovery (FBLD), in contrast, builds on much smaller and predominantly more hydrophilic compounds. It therefore emerges as an orthogonal approach to HTS and is now established in most large pharmaceutical companies (Chessari and Woodhead, 2009; Joseph-McCarthy et al., 2014). A recent study published by the company Astex shows that leads developed via FBLD, compared with HTS, are indeed significantly smaller and less lipophilic thus holding promise to increase success rates during clinical trials (Murray et al., 2012).

Another advantage of FBLD is that typical fragment libraries, consisting of only 10^2 to 10^4 molecules, sample the chemical space much more efficiently compared with HTS libraries, containing up to 10^6 entries, because each low-MW compound represents a large number of higher-MW compounds comprising the same fragment core (Roughley and Hubbard, 2011; Scott et al., 2012). However, hit detection is considerably more challenging in the case of FBLD and requires sensitive biophysical methods since the small size of fragments directly results in low binding affinities between 0.1 and 10 mM (Murray and Rees, 2009). Techniques such as nuclear magnetic resonance, surface plasmon resonance, high-concentration biochemical screens, mass spectrometry (MS), and thermal shift assays are often used to filter for promising fragments prior to X-ray crystallographic experiments (Hubbard and Murray, 2011; Joseph-McCarthy et al., 2014). In particular, the latter method plays a key role in FBLD because fragment-bound structures are ultimately needed to guide the subsequent chemical optimization process (Murray and Blundell, 2010; Schiebel et al., 2015). Analyzing currently applied screening strategies in a case study, we have recently found that biophysical screening methods are not able to reliably predict the majority of X-ray binders (Schiebel et al., 2016). This suggests that X-ray crystallography should be used as a primary screening tool, since it is more sensitive and information-rich than any other biophysical approach.

While crystallography had been a resource-intensive and time-consuming method in the past, current and future developments such as the miniaturization and automation of crystallization trials, brighter synchrotron sources, faster detectors as well as automatic crystal mounting robots aid in converting crystallography into a high-throughput technique that can be used for primary screening purposes (Davies and Tickle, 2012; Yin et al., 2014). In addition, it is essential to analyze the resulting data in a streamlined and automated manner avoiding laborious manual data processing, structure solution, and refinement (Echols et al., 2014; Mooij et al., 2006). In this regard, all fragment hits must be detected reliably, especially since a lot of resources are already spent on crystallographic screening and because the outcome of such an enterprise is of utmost importance for the

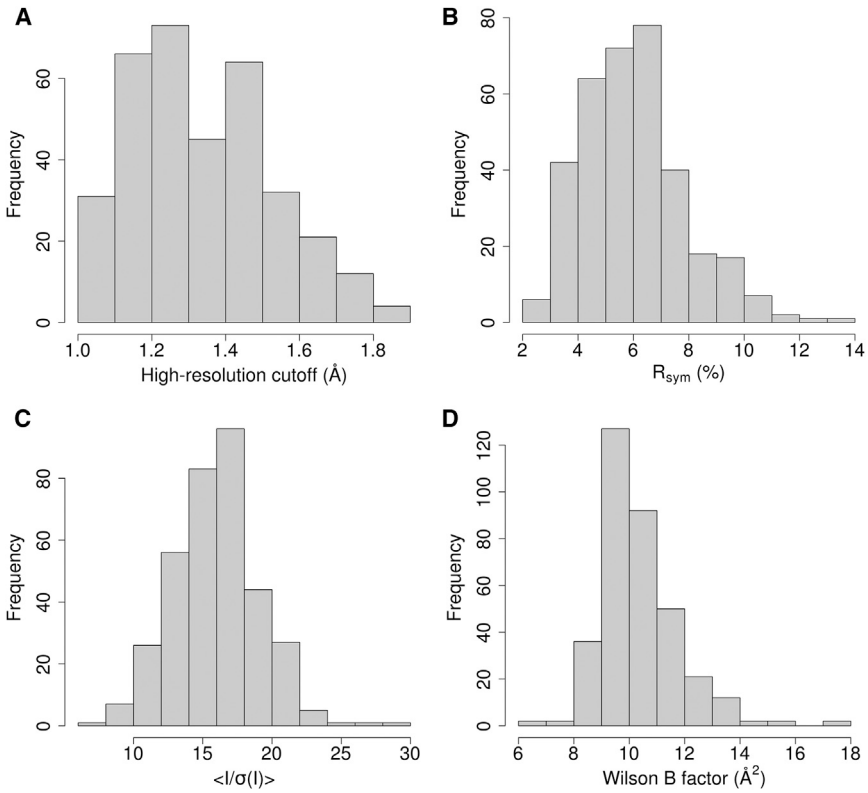


Figure 1. Data Collection Statistics

(A–D) The distributions of (A) the resolution, (B) R_{sym}, (C) <I/σ(I)>, and (D) Wilson B factor values are shown for all 348 $P2_1$ structures. Cell parameters are specified in Figure S1. Analogous results for the 16 $P2_12_12_1$ structures are given in Figure S2.

following hit-to-lead-to-drug evolution. This is particularly challenging as the low affinity of fragments frequently leads to only partial occupation of binding sites and thereby aggravates hit identification by ambiguous electron density features (Deller and Rupp, 2015; Echols et al., 2014). Hence, new and improved methods, which ensure that no X-ray fragment hits are missed, should be made available to FBLD practitioners and the broad crystallographic community.

For the development and validation of robust data analysis and hit identification tools, publicly available crystallographic data from a typical fragment screening campaign are urgently needed. Here, we provide and characterize a collection of 364 high-quality X-ray diffraction datasets (mean resolution of 1.35 ± 0.19 Å) obtained from crystals of the aspartic protease endothiapepsin (EP) that were individually soaked with different fragments from our in-house library (Köster et al., 2011). Analyzing this unique dataset using an automated refinement pipeline that was developed for this purpose, we found that completion of the solvent structure and refinement of each model prior to the search for fragment hits is critical and enables exceptionally high hit rates.

RESULTS

Collection and Provision of 364 High-Resolution Datasets

For the present study, our in-house compound collection that contains 364 molecules of the size of typical fragments has been used (for details please see the Experimental Procedures) (Köster et al., 2011). Applying six different biochemical and biophysical techniques, we screened the full fragment library for inhibitors of EP, a pepsin-like aspartic protease that serves as a surrogate for β-secretase, renin, and plasmepsins to develop new drugs for the treatment of Alzheimer's disease, hypertension, and malaria (Geschwindner et al., 2007; Schiebel et al., 2015). As frequently observed in fragment screening campaigns, the hit lists suggested by the different methods did not overlap to a convincing extent. Because crystal structures are ultimately needed for successful FBLD, we decided to screen all library fragments by X-ray crystallography without the application of any pre-filter. This project generated a wealth of structural information and allowed a performance evaluation of different screening strategies and techniques (Schiebel et al., 2016). In general, almost half of all crystallographic binders had not been detected by any of the six applied screening techniques. For instance, the best-performing technique, a reporter-displacement assay (Schiebel et al., 2015), resulted in 50 putative binders, of which 27 could be confirmed by crystallography. However, 44 fragment-bound structures would have been missed in a crystallographic screen of only those 50 compounds. Filtering fragment libraries by screening prior to crystallographic experiments can thus lead to an enrichment but not to an exhaustive identification of all X-ray hits.

In the crystallographic screening project, we were able to collect diffraction data for EP crystals that were individually soaked with each of the 364 fragments. Interestingly, 16 of the 364 crystals chosen for final data collection belonged to the orthorhombic $P2_12_12_1$ space group instead of the common monoclinic $P2_1$ form. While the mean cell parameters for $P2_1$ are a = 45.3 ± 0.1 Å, b = 73.0 ± 0.2 Å, c = 52.7 ± 0.1 Å, and β = 109.4° ± 0.2° (Figure S1), the $P2_12_12_1$ unit cell is closely related with cell parameters of a = 45.2 ± 0.1 Å, b = 72.6 ± 0.2 Å, c = 104.1 ± 0.3 Å (Figure S2). The rare incidence of the orthorhombic unit cell might be explained by the impact of certain chemicals on the crystal lattice during the soaking process as also observed for other systems (Skarzynski and Thorpe, 2006).

Crystals used in an FBLD project should usually diffract to at least 2.5 Å resolution to enable the unambiguous identification of fragments in the electron density (Scott et al., 2012). Here, exclusively datasets with resolutions better than 2 Å have been obtained resulting in a mean resolution of 1.34 ± 0.19 Å for the 348 $P2_1$ structures (for the respective histogram, see Figure 1A).

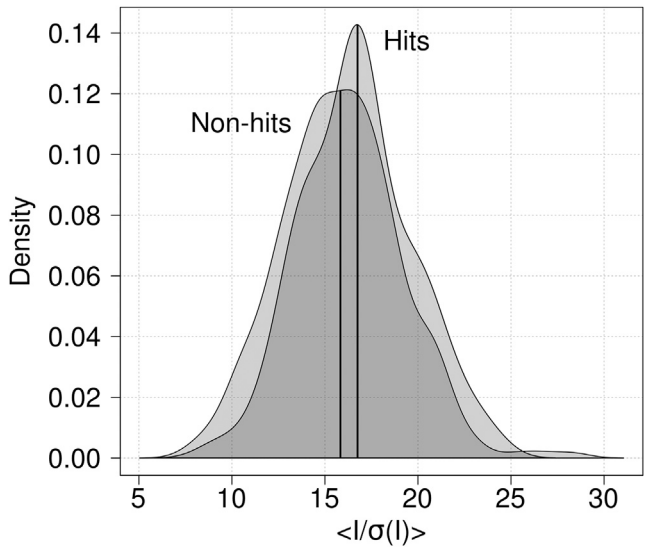


Figure 2. Mean I/σ(I) Distribution of Hits versus Non-hits

This graph compares the kernel density estimations for the 72 hits (right distribution) and the 292 non-hits (left distribution; for an explanation of these numbers, see the Experimental Procedures). Corresponding average <I/σ(I)> values are indicated by vertical black lines (16.8 and 15.8, respectively). The p values of two different tests are below 0.05 ($p_{t\ test}$ = 0.021, $p_{Mann\text{-}Whitney\ U\ test}$ = 0.026), suggesting that the observed difference between hits and non-hits is significant.

Importantly, also the other data collection quality indicators for these datasets are very convincing with a mean R_{sym} of 6.1% ± 1.9%, an <I/σ(I)> of 15.9 ± 3.1 and Wilson B factors of 10.4 ± 1.5 Å^2 (Figure 1). Moreover, sufficiently redundant and complete data were collected with a mean multiplicity of 3.9 ± 0.4 and an average completeness of 98.0% ± 2.1%, despite the low monoclinic symmetry. The additional $P2_12_12_1$ data are of similar high quality as shown in Figure S2. Overall, we provide a collection of exceptionally high-quality datasets to the crystallographic community that can be used for the development and validation of new methods for streamlined crystallographic fragment screening.

In an attempt to understand which parameters might be important for fragment binding in soaking experiments, we compared the diffraction properties of all crystals leading to fragment-bound structures with those yielding only empty apo structures of EP. In our study, the impressively high number of 71 fragments were found to bind to EP, corresponding to a hit rate of 20% (Schiebel et al., 2016). Interestingly, the hit subset displays a significantly higher mean <I/σ(I)> (Figure 2), which can either suggest that fragment binding improves the crystal diffraction power, for instance due to the ordering of flexible parts of the protein, or that crystals with stronger diffraction facilitate the detection of fragments in the electron density maps.

Automated Application of an Optimized Refinement Protocol Yields High-Quality Structures

Early on in the project, we recognized that such an enormous enterprise can only be tackled with a proper amount of automation. This is not only true for data collection but also holds for the processing of diffraction data. For this reason, we developed and used a script that automatically processes diffraction data of multiple crystals synchronized with data collection using *XDSAPP* (K.R., M.S.W., and U.M., unpublished data; Sparta et al., 2016) and subsequently performs an initial refinement using *PHENIX* (Adams et al., 2010). At this stage, a team of eight crystallographers was assembled to analyze data quality, to re-process and re-collect low-quality data, to identify fragment hits, and to refine ligand-bound crystal structures. Many of these tasks had to be repeated manually for each of the individual datasets, thus slowing down our progress considerably. In particular, the manual refinement of fragment-bound structures turned out to be very inefficient. In order to accelerate the present and future projects, we developed a standardized refinement strategy that was subsequently integrated into an automated refinement pipeline applicable to different systems. In a first step, molecular replacement (MR) is used to place the model and to obtain the initial phases. This is then followed by simulated annealing, standard coordinate, and isotropic atomic displacement parameter (ADP, B factor) refinement and then TLS refinement within *PHENIX*. The MR search was included to account for unit cell variations and to ensure that the pipeline is independent of the target space group. After the four initial steps, *COOT* is used to add water molecules to the model while taking care that electron density features representing larger ligands are rarely populated by water (Emsley et al., 2010). The four final refinement steps include the anisotropic treatment of ADPs from the protein and subsequently also from water molecules, the placement of protein-bound hydrogen atoms, as well as a final round of water update.

The development of this refinement pipeline enabled us to determine and refine each of the 364 structures in an automated and identical manner. Based on the wealth of the thereby generated information, it is particularly interesting to investigate which refinement protocol is most appropriate at which resolution (Figures 3 and S3). For instance, the anisotropic treatment of protein ADPs does not significantly change R_{free} values at resolutions between 1.5 and 1.6 Å while the gap between R and R_{free} on average increases by 1.2% ± 0.2%, clearly indicating over-fitting of the diffraction data. In contrast, R_{free} values decrease by 0.2% ± 0.2% and 0.4% ± 0.2% at resolutions in the range of 1.4–1.5 and 1.3–1.4 Å, respectively, highlighting that at least a resolution of 1.5 Å is necessary before ADPs should be refined anisotropically (cf. Figures 3A–3C). For the additional anisotropic treatment of water ADPs, a slightly higher resolution of 1.4 Å is usually required, underscoring that, particularly at resolutions between 1.4 and 1.5 Å, anisotropic refinement of only protein but not water ADPs might be the most appropriate strategy (Figure S3). Independent of the resolution, the inclusion of riding hydrogen atoms into the model resulted in a clear drop of R_{free} values (1.2% and 0.9% for resolutions of 1.0–1.2 and 1.6–1.8 Å, respectively) without a significant increase in the gap between R and R_{free} (Figure 3D). Although each H atom accounts for only one electron, which usually renders them invisible in the electron density, their large overall number in the protein co-determines the diffraction properties of EP crystals. Interestingly, this observation is not limited to EP but extends to other systems at various resolutions between 1 and 3 Å (Afonine et al., 2010).

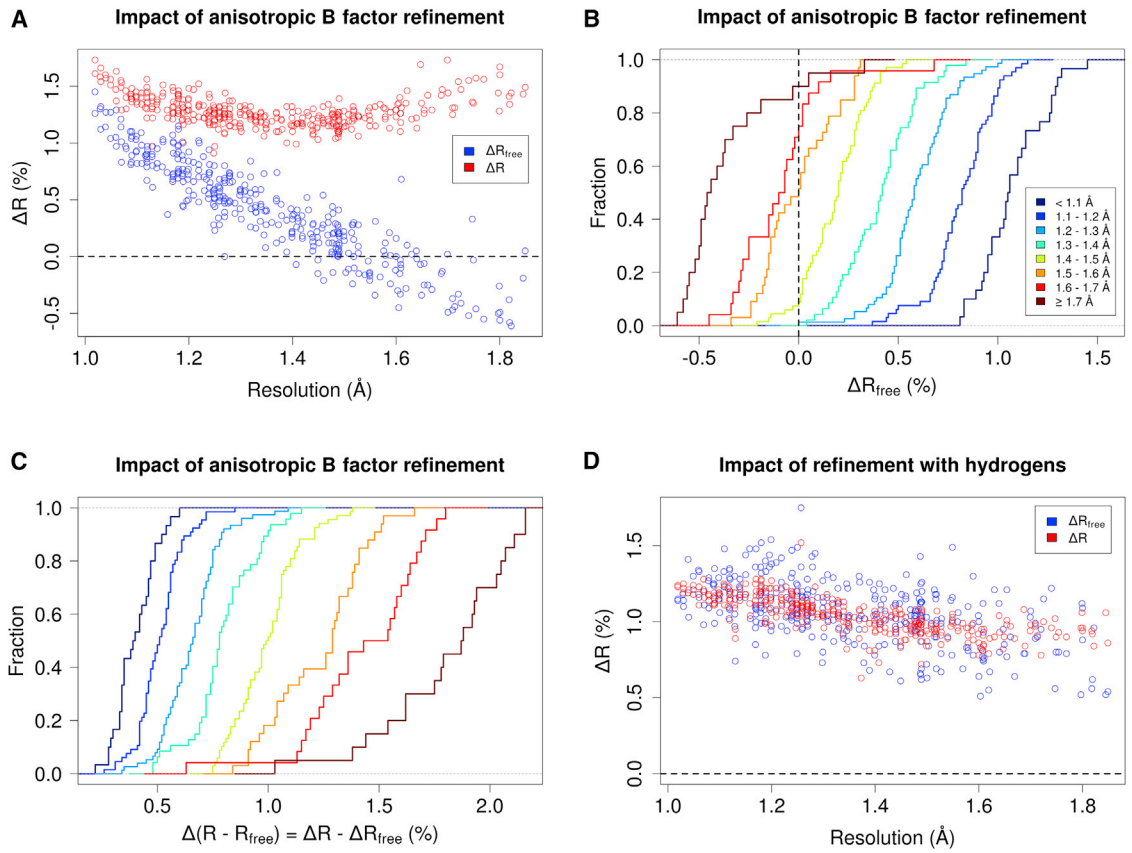


Figure 3. Choice of the Appropriate Refinement Strategy Based on Resolution Criteria

(A) Differences in R (red) and R_{free} (blue) before and after anisotropic ADP refinement of protein atoms ($\Delta R_{(free)} = R_{(free),before} - R_{(free),after}$) for all 364 datasets. The horizontal dashed line subdivides the graph into regions where R factors are improved (upper part) or deteriorated (lower part), respectively. The influence of the additional anisotropic treatment of water ADPs is visualized in Figure S3.

(B) Cumulative distributions of ΔR_{free} values (cf. (A)) in 0.1 Å resolution bins.

(C) Cumulative distributions of $\Delta R - \Delta R_{free}$ values (cf. (A); equal to $\Delta(R - R_{free})$) in 0.1 Å resolution bins (color code as defined in the insert of (B)).

(D) Differences in R (red) and R_{free} (blue) before and after refinement including hydrogens.

To analyze whether the automated refinement procedure delivers reasonable models, a statistical analysis of the most important refinement parameters was performed for all 364 datasets (Figure 4). The resulting structures, which on average contain 218 water molecules, are of very high quality as evidenced by a mean R value of 13.5% ± 1.0% and a mean R_{free} value of 16.3% ± 2.0%. Moreover, the mean figure of merit (FOM) obtained is 0.92 ± 0.02, the phase error for the free reflections is 14.5° ± 2.6°, and the mean B factor is 11.8 ± 1.6 $Å^2$.

Addition of Water Molecules Affects Model Quality Most Significantly

After validation of our refinement pipeline in general terms, we investigated the influence of each of the refinement steps on the quality of the resulting models. Clearly, every additional refinement step that was included in the pipeline leads, on average, to an improvement of the most relevant quality indicators that are mutually correlated (Figure 5). Importantly, the two steps involving the placement of water molecules account for 68% of the total phase error reduction with respect to the free reflections. A further 19% originates from the anisotropic treatment of ADPs (including the TLS refinement step) and 13% from the addition of hydrogen atoms. Since the anisotropic refinement of ADPs is only appropriate at resolutions better than 1.5 Å (Figures 3 and S3), this partitioning is resolution dependent. For structures with resolutions worse than 1.5 Å, anisotropic refinement deteriorates the phases by 0.3°. Consequently, the addition of water molecules to the model is even more important at these resolutions and accounts for almost 90% of the total phase improvement with some beneficial effect of hydrogen atom addition. In contrast to structures with resolutions worse than 1.5 Å, the overall phase error reduction is ~40% larger for structures with resolutions better than 1.5 Å. Indeed, this can be almost entirely explained by the positive effect of the anisotropic ADP refinement accounting for a phase improvement of 2.1° ± 1.1°. At the same time, the reduction in phase error due to the addition of water molecules remains unchanged at 5.7° compared with all structures with resolutions worse than 1.5 Å and only increases marginally from 1.0° to 1.1° when H atoms are added.

Furthermore, we estimated to what extent the refinement of the structures had converged using the automated procedure. For this purpose, we calculated the refinement parameter means

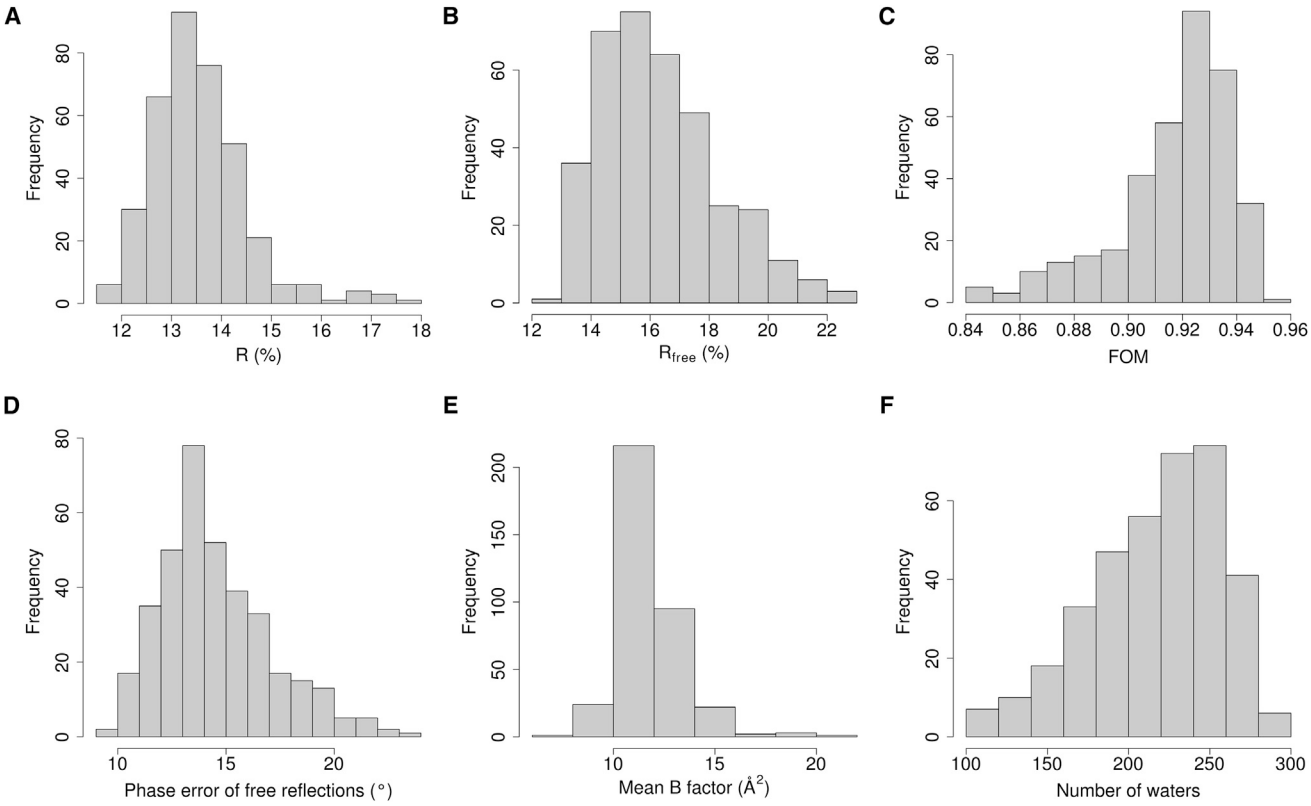


Figure 4. Statistics for the Results from the Automated Refinement

(A–F) Histograms are shown for (A) the R value, (B) the R_{free} value, (C) the figure of merit (FOM), (D) the phase error with respect to the free reflections, (E) the mean B factor, and (F) the number of water molecules in the model. For this analysis, all 364 structures were taken into account after the final step of the refinement pipeline. The PDB accession numbers of these structures can be found in Table S1.

for all manually finalized fragment-bound crystal structures (our unpublished data), defining the endpoint of the refinement, and compared them with the same parameters from the automated routine (Figure 5). While our refinement pipeline resulted in an average phase error of 14.9° for the hit subset, the crystallographers involved in this project deposited models with a mean phase error of 13.6°. Taking the phase error of 23.1° from the first refinement step of the pipeline as the starting point, the automated refinement has thus converged to 87%. The finalization of a structure including model building remains a predominantly manual task and, hence, it is not very surprising that the quality of the automatically generated structures could be further improved by manual intervention (Echols et al., 2014). In this regard, it is important to note that we have not attempted to automatically fit any ligands into electron density blobs because we wanted to allow the crystallographer to evaluate whether a fragment is present or absent in a structure based on unbiased $mF_o - DF_c$ maps. Moreover, we decided to place water molecules in a rather conservative fashion to avoid the population of ligand electron density and to provide a preferably bias-free model that can be conveniently used for manual improvement. Consequently, the automatically generated structures contain on average only 215 water molecules with respect to the hit subset, while 293 waters are present in the final ligand-bound models. Although a modified version of our pipeline that uses the more generous *PHENIX* water placement protocol resulted in a similar number of 287 water molecules, refinement convergence improved only slightly to 90%.

The Extent of Refinement Influences the Probability of Hit Detection

Crystallographers often determine a structure and perform only a crude initial refinement before they inspect the resulting electron density maps for the presence of a bound ligand. Structures without any sign of additional bound molecules are usually discarded, a strategy that is also frequently applied in FBLD (Drinkwater et al., 2010; Koh et al., 2015; Newman et al., 2012). However, fragments are characterized by low affinities and thus often only partially occupy the binding site of the target (Tiefenbrunn et al., 2014), which makes it much more difficult to detect them in difference electron density maps compared with inhibitors with improved affinity. This intrinsic difference between fragments and more potent inhibitors suggests that special procedures will be required to ensure reliable detection of all hits in a crystallographic fragment screening campaign. Indeed, during the initial manual treatment of our data, we recognized that more complete refinement of the structures positively influences fragment detection due to improved electron density features. Based on this experience, we developed a first version of an automated refinement pipeline to more exhaustively refine

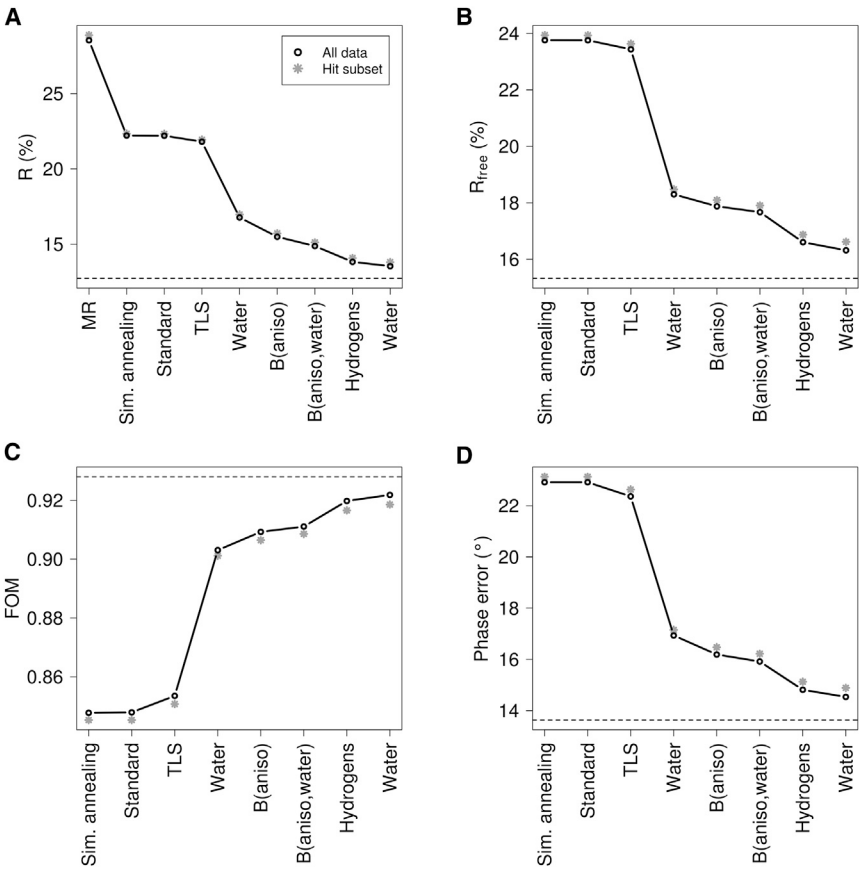


Figure 5. Development of Quality Indicator Values during the Refinement Procedure

(A–D) For each refinement step (A) mean R values, (B) mean R_{free} values, (C) the mean FOM, and (D) the mean phase error with respect to the free reflections are plotted. Values for all 364 datasets are shown as black circles, while those for the crystallographic hits are depicted as gray asterisks. As a reference, the black dashed lines indicate the mean values for the manually refined, deposited fragment-bound structures. Please note that our automated refinement procedure places initial water molecules after the TLS refinement. Additional waters are then added during the final refinement step. For further details regarding the refinement protocol, see the Experimental Procedures.

all structures before inspecting electron density maps for putatively bound fragments. This directly resulted in the identification of additional hits and, along with the above-described rationalization of our workflow, motivated us to improve and generalize the pipeline for use in current and future projects.

Applying this automated procedure, it was analyzed whether and how strongly the refinement process influences the likelihood to detect certain fragments in their electron density. The correlation coefficient (CC) between the $mF_o - DF_c$ map after a certain refinement step (no fragment modeled) and the F_c map calculated from the fragment molecule in its final binding position was used as a measure of the quality of fit between a certain observed electron density feature and the respective modeled fragment. Relative to the results from the initial standard refinement that are usually used for the detection of hits, fragment electron densities are clearly improved after the final refinement step as indicated by an 8.8% increase in the mean CC (Figure 6A). This and the following CC analyses refer to the subset of all crystallographic hits and also include additional fragment molecules in cases where the occupation of multiple binding sites could be detected, overall leading to 86 analyzable fragments (for details, cf. Experimental Procedures). Each of the individual refinement steps contributes to this positive effect likely because the phases gradually improve as the refinement converges (Figures 5 and 6B). Importantly, the most significant CC enhancement occurs along with the largest phase error drop during the placement of water molecules. All modeling steps involving the addition of waters result in 81.6% of the total CC increase. Moreover, the anisotropic treatment of ADPs and the addition of H atoms to the EP model account for the residual 13.7% and 4.7%, respectively. As already observed for the phase error, the overall CC improvement is higher for better resolved structures with a more beneficial effect of anisotropic ADP refinement. In addition, we estimated how much further the CC could have been reduced when manually refining all 364 structures. Based on the above approximation that the automated refinement has converged to 87%, the putative CC maximum for all fragments is 0.80 while our pipeline improves the mean CC from 0.72 to 0.79 (for an explanation of these seemingly low values, see the Experimental Procedures).

Visually screening all difference electron density maps for the presence of fragments after the initial standard refinement, and thereby following the frequently applied crystallographic workflow, we would have only identified 74% of our fragment hits (considering all copies of multiple binders). An additional 22 binders could, from our point of view, only be recognized after the application of the complete automated refinement procedure and, thus, would have been missed when relying only on initial $mF_o - DF_c$ maps. During the automated refinement of these structures, a gradual improvement of the fragment electron densities can be observed with the most significant impact of the addition of water molecules (Figures 7A–7I). Several examples illustrating the effect of the refinement on the quality of the map sections surrounding fragment molecules are given in Figures 7J–7M and S4. Interestingly, all 22 fragments that are hardly recognizable in the initial maps without prior knowledge of their binding position and conformation display a CC of less than 0.7 (Figures 6C, 7, and S4A–S4Q). A further 17 fragment copies also have an initial CC below this cutoff and are likewise characterized by a clear refinement-induced electron density improvement. In contrast, however, they may already be recognized in the initial maps based on certain electron density features, such as ring-shaped structures (cf. Figure S4R). This is particularly true when inspecting difference electron density maps at

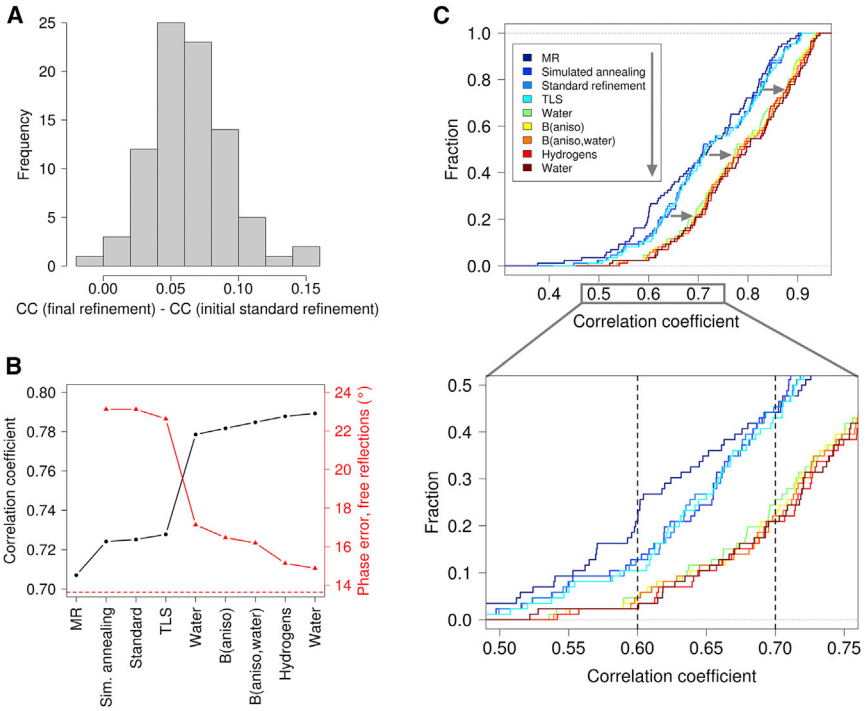


Figure 6. Improvement of Electron Density for Fragment Identification

Correlation coefficients (CC) were calculated between $mF_o - DF_c$ and F_c maps of each crystallographically detected fragment.

(A) Distribution of the CC improvements observed after the final refinement in comparison with the initial standard refinement.

(B) Development of the mean CC (black spheres) and phase error (red triangles) during the automatic refinement procedure for the hit subset. The mean phase error for all deposited fragment-bound structures is highlighted by the red dashed line.

(C) Cumulative distributions of CC values for the different refinement steps (color code as given in the legend). As indicated by the gray arrows, the distributions are shifted toward higher CC values as the refinement progresses. By far the most significant shift is caused by the initial addition of water molecules. The black dashed lines in the close-up view illustrate CC thresholds of 0.6 and 0.7, respectively.

a σ level of 2.5, which in our experience greatly facilitates the identification of fragment hits. Even for fragments with an initial CC above 0.7, further refinement often leads to an improvement of the map quality enabling a more unambiguous hit identification (Figures S4S and S4T).

Furthermore, our analysis shows that, based on the experience of the eight crystallographers involved in this project, fragment hits should at least display a CC of 0.6 with respect to the fully refined $mF_o - DF_c$ map (Figure 6C). This CC threshold resulted automatically since, for each fragment, we decided individually whether it had bound to EP based on the respective maps. Only two of our 86 fragment hits had a final CC below 0.6, which can be easily explained: the refinement pipeline is currently not able to recognize drastic conformational changes as observed for the complex between EP and fragment **218** where the flap region comprising residues 78–84 is displaced by the ligand and obscures the $mF_o - DF_c$ fragment density in its original position. On the other hand, fragment **31** can be characterized as a hit despite a CC below 0.6 because the planar phenyl ring is clearly visible in the density even though the ester functionality remains undetected.

Finally, the results from displacement isothermal titration calorimetry experiments were used (Schiebel et al., 2016) to investigate whether those hits only detected after proper structural refinement are of any value for hit-to-lead-to-drug chemistry or whether they only represent extremely weak binders (Figure S5). While fragments that were detected without the application of the refinement pipeline are characterized by a mean K_d of 2.7 ± 3.0 mM and a ligand efficiency (LE) of 1.22 ± 0.26 kJ mol^{-1} $atom^{-1}$, the additional hits found display an average K_d of 5.1 ± 3.8 mM and an LE of 1.21 ± 0.27 kJ mol^{-1} $atom^{-1}$. It is well known that an LE above 1.2 kJ mol^{-1} $atom^{-1}$, corresponding to ~0.3 kcal mol^{-1} $atom^{-1}$, is particularly important for successful FBLD and, hence, the additional hits identified may prove essential during the chemical optimization process (Scott et al., 2012). In particular, four of these fragments have LEs clearly above this threshold, and two fragments are sub-millimolar binders (Figure S5). In this regard, it is also important to note that the strongest binders are not always the best choice for chemical optimization and that less potent fragments can provide essential structural knowledge inspiring the medicinal chemist (Jhoti et al., 2013; Schiebel et al., 2016).

DISCUSSION

Today, FBLD plays an increasingly important role in the development of new drugs. Applications of this method can result in clinical candidates even for targets that had long been considered undruggable based on more classical approaches such as HTS campaigns (Coyne et al., 2010). The generation of structural data on fragment-target complexes is one of the key factors important for the success of an FBLD study (Murray and Blundell, 2010). In particular, X-ray crystallography is applied to obtain this essential information and was recently suggested to be most powerful when used as a primary screening technique (Schiebel et al., 2016). For this reason, new and improved methods need to be developed that allow the reliable and to a large degree automated detection of all fragment hits from a full crystallographic fragment screen.

Here, we provide a very high-quality X-ray diffraction dataset collected from 364 different fragment-soaked EP crystals that is ideally suited for the development and validation of such methods and can be downloaded for this purpose from the PDB (see Accession Numbers section and Table S1). Importantly, our dataset can also support the development, optimization, and validation of refinement programs and pipelines. Moreover, the data can be used for various other purposes such as the training or validation of ligand placement routines

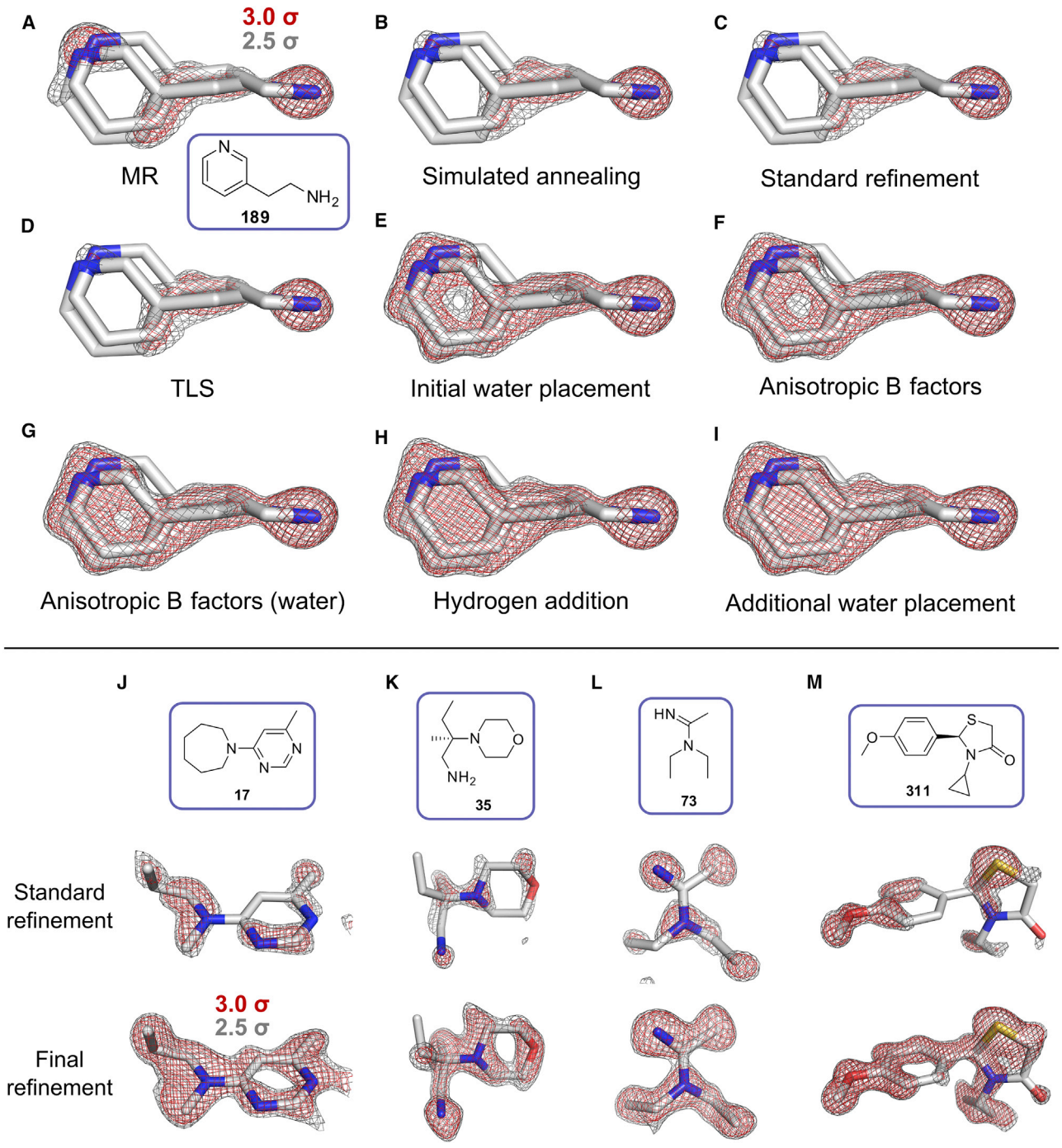


Figure 7. Improvement of Fragment Electron Densities through Automatic Refinement

(A–I) Gradual density enhancement highlighted for fragment 189. $mF_o - DF_c$ difference electron density maps are shown around fragment 189 (gray sticks) after each of the individual refinement steps as gray and red meshes at the 2.5 and 3.0 σ level, respectively. The individual refinement steps are as follows: (A) MR. The chemical structure of fragment **189**, which was modeled in two alternative conformations, is depicted in the blue box. (B) Simulated annealing; (C) standard refinement; (D) TLS; (E) initial water placement; (F) refinement of anisotropic B factors; (G) refinement of anisotropic B factors including water molecules; (H) hydrogen addition; (I) additional water placement.

(J–M) Further examples for electron density enhancements. For fragments **17** (J), **35** (K), **73** (L), and **311** (M), $mF_o - DF_c$ maps are depicted after the initial standard refinement (top) and after the final refinement (bottom) as gray and red meshes at the 2.5 and 3.0 σ level, respectively. The fragments are shown as gray sticks and respective chemical structures are presented in the blue boxes. The azepane ring of fragment **17** is never completely visible in the electron density and thus was only partially modeled. While the EP-**17** and EP-**35** complex structures contain the fragment twice, only one of these molecules is shown. Similar cases for which the electron density only allows ligand identification after the final refinement step are outlined in Figure S4. For information on the fragments' affinities and ligand efficiencies, see Figure S5.

as well as investigation of the conservation of side-chain conformations and water sites.

Within this project, all experiments that did not fulfill our strict quality criteria were repeated so that all crystals that were finally chosen diffracted to resolutions better than 2.0 Å and almost exclusively (96%) yielded R_{sym} values below 10% (Figures 1 and S2). This comprehensive crystallographic screening campaign was only feasible due to sufficient synchrotron beamtime and a high level of automation during data collection, processing, and analysis. In this context, we developed an automated in-house refinement pipeline that produced excellent models with 95% of all R_{free} values below 20%.

Analyzing the difference electron density maps resulting from this exercise, it was observed that, in contrast to common practice, elaborate structural refinement and especially water placement is crucial for the reliable detection of all fragment hits. This can be traced back to the improved phases of a fully refined model in comparison with a crude initial structure that is typically used for the search of bound fragments (Figure 6B). In a different fragment-based project, our refinement pipeline was applied to nine diffraction datasets and aided in the identification of three, instead of one, human carbonic anhydrase binders, underlining that our observations should also hold for other targets (S.G., J.S., A.H., and G.K., unpublished data). Although the positive influence of a more complete refinement may be particularly prominent in the case of fragments due to their low affinity and hence partial binding-site occupancy, similar trends have also been observed for larger ligands.

Based on these findings, we think that there is an urgent need for robust and efficient methods that are dedicated to FBLD. Clearly, most pharmaceutical companies have their in-house solutions, but these are not available to the broad crystallographic community and are, with probably one exception (Mooij et al., 2006), not well documented (Echols et al., 2014; Oster et al., 2015). Before we decided to compile our own customized refinement pipeline that satisfied our specific needs such as careful water placement, we also considered using the *PHENIX.ligand_pipeline* that was described previously (Echols et al., 2014). However, the authors explicitly note that the pipeline was not optimized for use in fragment-based projects. In contrast, the *AutoDrug* software has been developed specifically for this purpose but rather focuses on data collection and processing with only a crude initial refinement step (Tsai et al., 2013).

Because the accuracy of phases, however, turned out to be pivotal for the reliable detection of all crystallographic fragment hits, strategies need to be developed to automatically refine a large set of structures to almost complete convergence. Our pipeline fulfills this requirement already to 87%, and therefore is responsible for the detection of 22 additional hits, a number that admittedly can vary slightly depending on the expertise of the crystallographer. As learned from our study, the reliable identification of fragment hits primarily requires the modeling of water molecules with some beneficial effects originating from the addition of hydrogen atoms and from the anisotropic refinement of ADPs at resolutions better than 1.5 Å. Ultimately, the refinement pipeline should be directly coupled to data collection and processing software to further accelerate the screening process.

Such methods will permit a broad range of practitioners across industry and academia to use X-ray crystallography as primary fragment screening platform leading to a maximum of ligand-bound crystal structures of the targeted macromolecule. The resulting in-depth structural insights will provide the framework required to design new and urgently needed drugs for various indications.

EXPERIMENTAL PROCEDURES

Fragment Library

In a previous study, we described the assembly of a 364-entry fragment library, which accidentally contained three duplicates (Köster et al., 2011; Schiebel et al., 2015). The fragments can be divided into 292 molecules not binding to EP and 72 compounds identified as crystallographic binders (Schiebel et al., 2016). While the former group contains two of the three duplicates, the latter includes fragments **39** and **321**, which are chemically identical and both result in fragment-bound crystal structures. Thus, the complete dataset contains 71 non-redundant crystallographic hits.

Crystallization, Data Collection, and Processing

EP was isolated in 0.1 M sodium acetate buffer at a pH of 4.6 according to our previously described protocol (Köster et al., 2011). Subsequently, crystals were grown and soaked with individual fragments prior to data collection as previously described (Schiebel et al., 2016). In sitting-drop vapor diffusion experiments, 2 μl of a 0.15 mM EP solution was mixed with the same volume of reservoir solution containing 0.1 M NaAc (pH 4.6), 0.1 M NH_4Ac, and 24–30% (w/v) PEG 4000. After equilibration at 17°C for at least 1 week, individual crystals were transferred into 11 μl of 70 mM NaAc (pH 4.6), 70 mM NH_4Ac, 16%–20% (w/v) PEG 4000, 19%–23% (v/v) glycerol, 9% (v/v) DMSO, and 90 mM of one of the 364 fragments. Crystals were harvested from these soaking drops after 48 hr and immediately flash-frozen in liquid nitrogen.

Diffraction data were collected at the BESSY MX beamlines 14.1, 14.2, and 14.3 (Mueller et al., 2012, 2015) and processed on site using the beamline version of *XDSAPP* in command-line mode (K.R., M.S.W., and U.M., unpublished data; Sparta et al., 2016) for each of the collected datasets, including a subsequent *PHENIX*-based refinement step utilizing a user-supplied model (Adams et al., 2010; Huschmann et al., 2016). The resulting data collection statistics were reviewed for each of the 364 datasets and, if necessary, manual re-processing was performed using *XDS*, *XDSAPP*, *HKL2000*, or *IMOSFLM*/*SCALA* (Evans, 2006; Kabsch, 2010; Leslie, 1992; Otwinowski and Minor, 1997; Sparta et al., 2016). In cases where data collection statistics still did not fulfill our strict quality requirements, the complete crystallographic experiment was repeated. All datasets can be downloaded from the PDB (see Accession Numbers section and Table S1) for further use and evaluation.

All final data were statistically analyzed and plotted using *R* (R Development Core Team, 2010). Data collection parameters are given in the text as means ± SD.

Automated Refinement

Our customized refinement pipeline basically consists of nine individual steps and solely builds on available crystallographic software. The script, written in bash command language, successively processes all reflection files that are specified as input. In the initial step, each structure is determined via MR taking the user-supplied coordinate file as search model after removal of all alternative side-chain conformations, hydrogen atoms, and ANISOU records by *PDBCUR* of the *CCP4* package (Winn et al., 2011). During the MR procedure, the *PHENIX*-implemented version of *PHASER* searches for as many molecules in the asymmetric unit as suggested by a Matthews coefficient analysis while testing all alternative space groups (McCoy et al., 2007). If a space group different from the input has been identified by *PHASER*, it is updated in the reflection file using *MTZUTILS* of the *CCP4* suite. All subsequent refinement steps are performed with the command-line version of *PHENIX.refine* (*PHENIX* version dev2006) at the constant number of five macrocycles and include bulk solvent correction as well as anisotropic scaling of the data (Adams et al., 2010). The initial Cartesian simulated annealing step allowing for rearrangements of flexible protein parts is followed by a standard XYZ coordinate refinement with isotropic ADPs. Subsequently, a TLS refinement is performed based on those TLS groups identified for the search model via

the *PHENIX.find_tls_groups* routine. In the next step, *COOT*'s *find-waters* routine is used to generate a preliminary water model based on the $mF_o - DF_c$ map at the 4 σ level (Emsley and Cowtan, 2004; Emsley et al., 2010). Alternatively, we also tested water placement via *PHENIX* but finally decided to stick to the more conservative *COOT*-based version as the default. After an interim refinement applying the same strategy as in the previous step, waters that are characterized by an $mF_o - DF_c$ map variance above 2 σ are detected by the check-waters-by-difference-map tool in *COOT*. Such waters likely were placed into ligand electron density and thus are subsequently removed from the model prior to the actual refinement step (again with the same TLS refinement strategy). In the sixth step, the TLS refinement is replaced by the anisotropic treatment of all ADPs except those of water molecules. In contrast, the refinement of anisotropic ADPs is switched on for all atoms including water oxygens starting from the following step. The resulting model is then hydrogenated using *PHENIX.ready_set* and refined with riding hydrogens prior to a final round of water addition. This is again performed in the above-described two-step manner with the exception that waters are now placed based on an $mF_o - DF_c$ σ level of 5.5 in order to avoid over-interpretation of dubious electron density features. After the water model has been updated, the final refinement is carried out with anisotropic ADPs and riding hydrogen atoms. Finally, we would like to note that our pipeline was not only used for EP but also tested on a number of other proteins where it produced reliable results (data not shown). The refinement pipeline tool described here was written in bash command language and is available from the authors upon request. Based on the original ideas and results reported here, this tool is presently being developed into a highly sophisticated and robust Python-based structure refinement workflow with a particular focus on fragment screening campaigns (K.R., M.S.W., and U.M., unpublished data). Applying a *PHENIX*- and *COOT*-based refinement strategy and additional concepts for robust ligand identification, and making use of multi-core processors, the new program will perform a completely automatic refinement of a PDB model against multiple reflection datasets in parallel. It is also planned to link this tool to *XDSAPP* (Sparta et al., 2016) so that it can be invoked seamlessly after the automated processing of multiple diffraction datasets.

For the present study, all 364 reflection files arising from the above-described data collection were processed by the current refinement pipeline. A 0.99 Å resolution structure of the EP apo form lacking all water and ligand molecules was used as an MR search model (PDB: 4Y5L). To warrant comparability, we consistently used the results from the initial standard refinement and opposed them to those from the final refinement step because the majority of structures are characterized by resolutions for which this refinement strategy is most appropriate. All figures displaying structural elements and electron densities were produced using *PYMOL* (DeLano, 2002). Refinement parameter data were statistically analyzed and plotted within the *R* environment.

Calculation of Correlation Coefficients

The refinement pipeline adds water molecules in a conservative fashion to the model and avoids population of densities that might represent larger ligands. This strategy was successful in 72% of cases, while in the residual cases, on average 1.6 waters had been placed into individual fragment electron densities. Nevertheless, those fragment hits could still be reliably detected because the reduction in the $mF_o - DF_c$ signal is compensated by an emerging $2mF_o - DF_c$ peak. For the purpose of our study, however, we wanted to compare unbiased $mF_o - DF_c$ maps with F_c maps calculated for the respective fragment in its final binding pose, and therefore ran the refinement pipeline in a slightly modified version. In particular, water molecules closer than 2.2 Å to the respective fragment were additionally removed from the model prior to the two secondary water refinement steps. Seventy of the 72 diffraction datasets resulting in crystallographic hits were handled in this way. To avoid redundancy, the duplicate fragment **39** was omitted from the CC analysis. Fragment **177** was detected in the binding pocket of EP but lost its fragment character due to a reaction (J.S., F.R.E., G.K., and A.H., unpublished data) and thus was likewise left out. The 70 datasets finally used for this analysis encompass 86 bound fragment copies for which individual CCs could be calculated (Figure 6). Similarly, fragments **39** and **177** were omitted from the hit subset used for the production of the graphs depicted in Figure 5.

For the calculation of CC values, the respective deposited fragment-bound structure (our unpublished data) was superimposed onto the output models of every refinement step using the *superpose* routine as implemented in the *CCP4* package (Krissinel and Henrick, 2004). Subsequently, F_c maps were calculated for each individual fragment copy using *SFALL* and *MAPMASK* of the *CCP4* suite (Agarwal, 1978). Associated $mF_o - DF_c$ maps were generated with *FFT* based on the reflection files produced by the individual refinement steps (Ten, 1973). Using the *CCP4* program *OVERLAPMAP* (Branden and Jones, 1990), CCs were then calculated between corresponding F_c and $mF_o - DF_c$ maps for every individual fragment copy and refinement step. The resulting data were finally analyzed using the statistical framework *R*.

Importantly, we would like to note that our CC values are inevitably lower compared with ordinary CCs because $mF_o - DF_c$ maps prior to the placement of any ligands were used for the correlations to avoid model bias and to better reflect the fact that crystallographers typically search this map for the presence of bound ligands. All ordinary real-space CCs (RSCC; F_o versus F_c map) calculated based on our deposited fragment-bound crystal structures using *PHENIX.real_space_correlation* display values above 0.9 with a mean of 0.97 ± 0.02. In a recent publication, Deller and Rupp (2015) argue that such ligands with an RSCC >0.9 can be regarded as trustworthy, underlining that the interpretation of our data is reasonable.

ACCESSION NUMBERS

The structure factor amplitudes of all 364 datasets and their corresponding structural coordinates have been deposited in the PDB via the newly established group deposition tool. For now, the data can be downloaded using the 364 individual PDB accession numbers given in Table S1, which also includes instructions for easy download of all structures. However, in the future it will be possible to retrieve all data via the PDB group ID G_1002001.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.str.2016.06.010.

AUTHOR CONTRIBUTIONS

Data were collected and processed by J.S., S.G.K., X.W., A.Y.P., M.S., F.R.E., K.F., N.R., F.U.H., M.S.W., U.M., and A.H while K.S. and M.K. developed the computational tools for data processing. The resulting electron density maps were screened for the presence of fragments by J.S., S.G.K., A.K., X.W., A.Y.P., M.S., F.R.E., K.F., N.R., and A.H. Data analysis and interpretation were performed by J.S. and A.H. The automated refinement pipeline was written and applied by J.S. Moreover, S.G. contributed the results concerning the human carbonic anhydrase. The research was designed by M.S.W., U.M., G.K., and A.H with contributions from J.S. and S.G.K. The manuscript was written by J.S., G.K., and A.H.

ACKNOWLEDGMENTS

This study was supported by the German Ministry of Science and Education in the BioChancePlus Program (Project FragScreen No. 0315161C) and the BMBF-project Frag2Xtal (No. 05K13RM1). Additional financial support was received from the European Research Council (ERC) of the European Union (grant 268145-DrugProfilBind). Moreover, we gratefully acknowledge the generous allocation of synchrotron radiation beamtime and travel support through the Helmholtz-Zentrum Berlin. We would like to thank the PDB team, in particular Stephen K. Burley and Jasmine Young, for support and enabling group deposition of all datasets.

Received: April 5, 2016
Revised: June 7, 2016
Accepted: June 8, 2016
Published: July 21, 2016

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010).

PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Afonine, P.V., Mustyakimov, M., Grosse-Kunstleve, R.W., Moriarty, N.W., Langan, P., and Adams, P.D. (2010). Joint X-ray and neutron refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *66*, 1153–1163.

Agarwal, R. (1978). A new least-squares refinement technique based on the fast Fourier transform algorithm. Acta Crystallogr. *A34*, 791–809.

Branden, C.-I., and Jones, T.A. (1990). Between objectivity and subjectivity. Nature *343*, 687–689.

Chessari, G., and Woodhead, A.J. (2009). From fragment to clinical candidate - a historical perspective. Drug Discov. Today *14*, 668–675.

Coyne, A.G., Scott, D.E., and Abell, C. (2010). Drugging challenging targets using fragment-based approaches. Curr. Opin. Chem. Biol. *14*, 299–307.

Davies, T.G., and Tickle, I.J. (2012). Fragment screening using X-ray crystallography. Top. Curr. Chem. *317*, 33–59.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System on World Wide Web (Delano Scientific) http://www.pymol.org.

Deller, M.C., and Rupp, B. (2015). Models of protein-ligand crystal structures: trust, but verify. J. Comput. Aided Mol. Des. *29*, 817–836.

Drinkwater, N., Vu, H., Lovell, K.M., Criscione, K.R., Collins, B.M., Prisinzano, T.E., Poulsen, S.A., McLeish, M.J., Grunewald, G.L., and Martin, J.L. (2010). Fragment-based screening by X-ray crystallography, MS and isothermal titration calorimetry to identify PNMT (phenylethanolamine N-methyltransferase) inhibitors. Biochem. J. *431*, 51–61.

Echols, N., Moriarty, N.W., Klei, H.E., Afonine, P.V., Bunkoczi, G., Headd, J.J., McCoy, A.J., Oeffner, R.D., Read, R.J., Terwilliger, T.C., and Adams, P.D. (2014). Automating crystallographic structure solution and refinement of protein-ligand complexes. Acta Crystallogr. D Biol. Crystallogr. *70*, 144–154.

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

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

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. *62*, 72–82.

Geschwindner, S., Olsson, L.L., Albert, J.S., Deinum, J., Edwards, P.D., de Beer, T., and Folmer, R.H. (2007). Discovery of a novel warhead against beta-secretase through fragment-based lead generation. J. Med. Chem. *50*, 5903–5911.

Hann, M.M. (2011). Molecular obesity, potency and other addictions in drug discovery. MedChemComm *2*, 349–355.

Hubbard, R.E., and Murray, J.B. (2011). Experiences in fragment-based lead discovery. Methods Enzymol. *493*, 509–531.

Huschmann, F.U., Linnik, J., Sparta, K., Ühlein, M., Wang, X., Metz, A., Schiebel, J., Heine, A., Klebe, G., Weiss, M.S., and Mueller, U. (2016). Structures of endothiapepsin-fragment complexes from crystallographic fragment screening using a novel, diverse and affordable 96-compound fragment library. Acta Crystallogr. F Struct. Biol. Commun. *72*, 346–355.

Jhoti, H., Williams, G., Rees, D.C., and Murray, C.W. (2013). The 'rule of three' for fragment-based drug discovery: where are we now? Nat. Rev. Drug Discov. *12*, 644–645.

Joseph-McCarthy, D., Campbell, A.J., Kern, G., and Moustakas, D. (2014). Fragment-based lead discovery and design. J. Chem. Inf. Model. *54*, 693–704.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Koh, C.Y., Siddaramaiah, L.K., Ranade, R.M., Nguyen, J., Jian, T., Zhang, Z., Gillespie, J.R., Buckner, F.S., Verlinde, C.L., Fan, E., and Hol, W.G. (2015). A binding hotspot in *Trypanosoma cruzi* histidyl-tRNA synthetase revealed by fragment-based crystallographic cocktail screens. Acta Crystallogr. D Biol. Crystallogr. *71*, 1684–1698.

Köster, H., Craan, T., Brass, S., Herhaus, C., Zentgraf, M., Neumann, L., Heine, A., and Klebe, G. (2011). A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. J. Med. Chem. *54*, 7784–7796.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. *60*, 2256–2268.

Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography *26*.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Mooij, W.T., Hartshorn, M.J., Tickle, I.J., Sharff, A.J., Verdonk, M.L., and Jhoti, H. (2006). Automated protein-ligand crystallography for structure-based drug design. ChemMedChem *1*, 827–838.

Mueller, U., Darowski, N., Fuchs, M.R., Forster, R., Hellmig, M., Paithankar, K.S., Puhringer, S., Steffien, M., Zocher, G., and Weiss, M.S. (2012). Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. J. Synchrotron Radiat. *19*, 442–449.

Mueller, U., Forster, R., Hellmig, M., Huschmann, F.U., Kastner, A., Malecki, P., Puhringer, S., Rower, M., Sparta, K., Steffien, M., et al. (2015). The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: current status and perspectives. Eur. Phys. J. Plus *130*, 141–150.

Murray, C.W., and Blundell, T.L. (2010). Structural biology in fragment-based drug design. Curr. Opin. Struct. Biol. *20*, 497–507.

Murray, C.W., and Rees, D.C. (2009). The rise of fragment-based drug discovery. Nat. Chem. *1*, 187–192.

Murray, C.W., Verdonk, M.L., and Rees, D.C. (2012). Experiences in fragment-based drug discovery. Trends Pharmacol. Sci. *33*, 224–232.

Newman, J., Dolezal, O., Fazio, V., Caradoc-Davies, T., and Peat, T.S. (2012). The DINGO dataset: a comprehensive set of data for the SAMPL challenge. J. Comput. Aided Mol. Des. *26*, 497–503.

Oster, L., Tapani, S., Xue, Y., and Kack, H. (2015). Successful generation of structural information for fragment-based drug discovery. Drug Discov. Today *20*, 1104–1111.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

R Development Core Team. (2010). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Rees, D.C., Congreve, M., Murray, C.W., and Carr, R. (2004). Fragment-based lead discovery. Nat. Rev. Drug Discov. *3*, 660–672.

Roughley, S.D., and Hubbard, R.E. (2011). How well can fragments explore accessed chemical space? A case study from heat shock protein 90. J. Med. Chem. *54*, 3989–4005.

Schiebel, J., Radeva, N., Köster, H., Metz, A., Krotzky, T., Kuhnert, M., Diederich, W.E., Heine, A., Neumann, L., Atmanene, C., et al. (2015). One question, multiple answers: biochemical and biophysical screening methods retrieve deviating fragment hit lists. ChemMedChem *10*, 1511–1521.

Schiebel, J., Radeva, N., Krimmer, S.G., Wang, X., Stieler, M., Ehrmann, F.R., Fu, K., Metz, A., Huschmann, F.U., Weiss, M.S., et al. (2016). Six biophysical screening methods miss a large proportion of crystallographically discovered fragment hits: a case study. ACS Chem. Biol. *11*, 1693–1701.

Scott, D.E., Coyne, A.G., Hudson, S.A., and Abell, C. (2012). Fragment-based approaches in drug discovery and chemical biology. Biochemistry *51*, 4990–5003.

Skarzynski, T., and Thorpe, J. (2006). Industrial perspective on X-ray data collection and analysis. Acta Crystallogr. D Biol. Crystallogr. *62*, 102–107.

Sparta, K.M., Krug, M., Heinemann, U., Mueller, U., and Weiss, M.S. (2016). XDSAPP2.0. J. Appl. Crystallogr. *49*, 1085–1092.

Ten, L. (1973). Crystallographic fast Fourier transforms. Acta Crystallogr. *A29*, 183–191.

Tiefenbrunn, T., Forli, S., Happer, M., Gonzalez, A., Tsai, Y.S., Soltis, M., Elder, J.H., Olson, A.J., and Stout, C.D. (2014). Crystallographic fragment-based drug discovery: use of a brominated fragment library targeting HIV protease. Chem. Biol. Drug Des. *83*, 141–148.

Tsai, Y., McPhillips, S.E., Gonzalez, A., McPhillips, T.M., Zinn, D., Cohen, A.E., Feese, M.D., Bushnell, D., Tiefenbrunn, T., Stout, C.D., et al. (2013). AutoDrug: fully automated macromolecular crystallography workflows for fragment-based drug discovery. Acta Crystallogr. D Biol. Crystallogr. *69*, 796–803.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

Yin, X., Scalia, A., Leroy, L., Cuttitta, C.M., Polizzo, G.M., Ericson, D.L., Roessler, C.G., Campos, O., Ma, M.Y., Agarwal, R., et al. (2014). Hitting the target: fragment screening with acoustic in situ co-crystallization of proteins plus fragment libraries on pin-mounted data-collection micromeshes. Acta Crystallogr. D Biol. Crystallogr. *70*, 1177–1189.