Article

Shape-Restrained Modeling of Protein−Small-Molecule Complexes with High Ambiguity Driven DOCKing

Panagiotis I. Koukos,[#] Manon Réau,[#] and Alexandre M. J. J. Bonvin*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Small-molecule docking remains one of the most valuable computational techniques for the structure prediction of protein−small-molecule complexes. It allows us to study the interactions between compounds and the protein receptors they target at atomic detail in a timely and efficient manner. Here, we present a new protocol in HADDOCK (High Ambiguity Driven DOCKing), our integrative modeling platform, which incorporates homology information for both receptor and compounds. It makes use of HADDOCK's unique ability to integrate information in the simulation to drive it toward conformations, which agree with the provided data. The focal point is the use of shape restraints derived from homologous

compounds bound to the target receptors. We have developed two protocols: in the first, the shape is composed of dummy atom beads based on the position of the heavy atoms of the homologous template compound, whereas in the second, the shape is additionally annotated with pharmacophore data for some or all beads. For both protocols, ambiguous distance restraints are subsequently defined between those beads and the heavy atoms of the ligand to be docked. We have benchmarked the performance of these protocols with a fully unbound version of the widely used DUD-E (Database of Useful Decoys-Enhanced) dataset. In this unbound docking scenario, our template/shape-based docking protocol reaches an overall success rate of 81% when a reliable template can be identified (which was the case for 99 out of 102 complexes in the DUD-E dataset), which is close to the best results reported for bound docking on the DUD-E dataset.
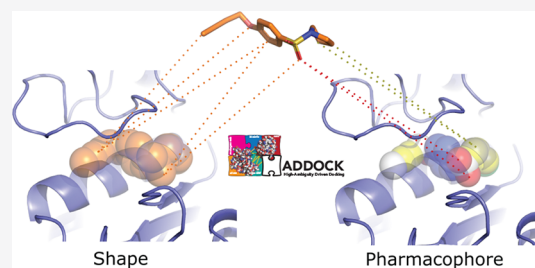
■ INTRODUCTION

The importance of reliable methods for the docking of small-molecule compounds to receptors of pharmaceutical interest cannot be understated. The nature of modern drug development practices dictates the gradual filtering of millions (perhaps even hundreds of millions) of compounds contained in virtual libraries to, ultimately, a few dozen lead compounds that can be further optimized before their clinical potential is investigated in animal and human trials.[1,2] This set of practices—collectively known as Computer-Aided Drug Design (CADD)—encompasses a variety of methods such as the virtual screening of compounds, molecular docking with recent developments making use of machine learning-based approaches,[3,4] and binding affinity prediction.

A relatively recent development in this space has been the advent of blind experiments (or exercises) focusing on pose prediction and binding affinity ranking/prediction. Examples of such initiatives are the aptly-named "Grand Challenges" (GC), held by the Drug Design Data Resource (D3R) consortium, on an annual basis between 2015 and 2018.[5−8] These experiments were built on the experience of the earlier Community Structure−Activity Resource (CSAR) experiment.[9] The stated drives behind these initiatives have been the desire to establish and push state-of-the-art performance in these challenging modeling problems, develop commonly agreed-upon standards of evaluating and assessing aforementioned performance,

codify best practices, and facilitate communication and sharing of data between pharmaceutical companies and academic investigators. In these targets, these initiatives have been, largely, successful.

It is through our participation[10−12] with HADDOCK (High Ambiguity Driven DOCKing), our integrative modeling platform,[13,14] in the D3R experiment that we initially developed docking protocols tailored to the peculiarities of protein−small-molecule complex modeling. Whereas, our protocol of choice for the 2016 iteration of the GC did not earn us a spot among the top performing groups of that round, it did allow us to better understand the problems specific to protein−small-molecule docking. The main takeaway point was that making use of the most closely related receptor for every target compound significantly improved the outcome of the modeling. In this case, the best receptor was identified after comparing crystallographic compounds with the target compound and selecting the receptor conformation with the
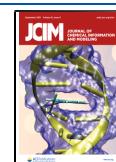
most similar ligand. With that knowledge, we optimized additional aspects of our approach for GC 2017, always prioritizing the use of high-quality experimental information for every step of the process. That revised protocol resulted in HADDOCK submitting one of the most accurate predictions for the pose prediction component of the challenge. We applied the same protocol in GC 2018 with equally good results.

Our successful D3R protocol can be summarized as follows: (1) After identifying highly homologous receptors with a co-crystallized compound, we compare the similarity of all crystallographic compounds to all target compounds and select the receptor conformation whose compound has the highest similarity to the compound to dock. (2) Prior to docking, we filter the generated conformers by comparing their 3D shape with that of the most similar crystallographic compound and select the 10 closest conformers in terms of shape similarity. (3) Finally, these 10 conformers are placed into the binding pocket by superimposing their shape onto the shape of the crystallographic compound and the model was refined in HADDOCK (i.e., no docking is performed).

While the use of shapes in protein–ligand modeling is not novel, even dating back to the earliest days of the field,[15,16] with common approaches emphasizing shape complementarity or overlap,[17−19] to the best of our knowledge, HADDOCK's ability to drive the desired compound into specific parts of the binding pocket via the use of shape information in combination with other kinds of restraints in a unified integrative modeling framework sets it apart from most modeling software.

Here, we present a new shape-based protocol, which incorporates all the lessons that we have learned over 3 years of participating in the D3R blind docking experiment into a protocol designed for HADDOCK, bypassing one of the main limitations of the previous protocols—their reliance on external software for significant parts of the ligand-based modeling process. This limitation did not allow us to use the integrative modeling and flexible refinement capabilities of HADDOCK as the rigid-body and semi-flexible refinement stages were bypassed and only the final flexible refinement stage was performed. Also, the fact that we are no longer dependent on commercial software for parts of the modeling workflow means we are free to develop this protocol into a pipeline accessible through the freely available HADDOCK webserver[14] (https://wenmr.science.uu.nl/haddock2.4/).

In the new shape-restrained protocols presented here, the principles underlying template identification and conformer generation procedures are similar to the ones described previously. The main difference is that after identifying a suitable receptor template for each target, its bound compound is transformed into a shape consisting of fake beads. In the pharmacophore-based version of the protocol, a pharmacophore shape is only introduced if features can be assigned for that part of the template compound; otherwise, a regular shape bead is used. Ambiguous distance restraints to guide the docking are then defined between those beads and the heavy atoms of the compound. There is no selection of conformers prior to docking; but instead, up to 50 conformers are docked into the receptor template. The most suited conformations are automatically selected during the docking by HADDOCK based on the shape restraints. These restraints are also drivers of possible conformational changes in the ligand during the flexible refinement stage of HADDOCK. We evaluate the

performance of these new shape-restrained protocols on the Database of Useful Decoys-Enhanced (DUD-E) dataset that contains crystallographic structures for 102 protein−small-molecule complexes.[20] Among those, 58 proteins can be classified as protein kinases, nuclear receptors, proteases, GPCRs, cleaving enzymes, cyclooxygenases, cytochromes P450, ion channels, or histone deacetylases and 54 are unique representatives of a protein family.[20,21] The DUD-E and its previous versions[22,23] have been mostly used to benchmark virtual screening tools.[22,23] When used to benchmark sampling tools, in most cases, the bound conformations of the target proteins were used.[21,24] The use of bound structures, however, biases the outcome of the docking since it suggests that we know the exact conformation of the target proteins when bound to their ligand and that we effectively ignore the intrinsic flexibility of their binding site and any binding-induced conformational changes.[25] In this work, we introduce a new unbound benchmark version of the DUD-E that allows us to assess the performance of the HADDOCK shape-restrained protocols in more realistic conditions.

■ RESULTS AND DISCUSSION

Dataset. We successfully identified templates for 99 of the 102 targets that are part of the DUD-E dataset (Tables S1 and S2). Those templates correspond to identical or homologous receptors that display the exact same sequence at the binding site level and that are co-crystallized with a different ligand. The amount of conformational changes in the binding sites ranges from 0.08 to 4.01 Å (see Tables S1 and S2). No usable templates could be identified for targets *cxcr4*, *drd3*, and *kpcb*. For *kpcb* and *drd3*, we failed to identify any homologous templates, whereas for cxcr4, we did identify some, but they were all in complex with the same compound as the reference receptor, so we discarded them. We identified at least one viable template for the remaining targets and selected one for docking based on the similarity of the reference and template compounds (see Materials and Methods). The similarity values range from almost 1 to below 0.2 indicating the presence of templates whose compounds are almost identical to their respective reference compounds on one end and templates that are almost entirely dissimilar on the other. The distribution of similarity values for the templates used for the two protocols can be seen in Figure S1. As we expected the similarity between the template and reference compounds to be a limiting factor for the outcome of the modeling, we decided to investigate that relationship for the shape protocol by identifying an additional low similarity template for all targets whose original template had a similarity value greater than 0.8 (see Figure S1). In this way, we can compare the performance in the subset of the dataset which includes both low- and high-similarity templates. In total, for 34 of the 35 targets for which the original similarity value was greater than 0.8, we could identify an additional low-similarity template (no alternative templates could be identified for target *kith*).

Conformer Generation. Prior to assessing the docking performance, we evaluate the performance of the conformer generation as we expect it to have a significant impact on the outcome of the docking since starting ligand conformations that are very different to the reference compound would need to undergo significant conformational rearrangements. In summary, RDKit and the settings we have chosen (see Materials and Methods) perform well as we are able to generate at least one acceptable pose (conformers whose
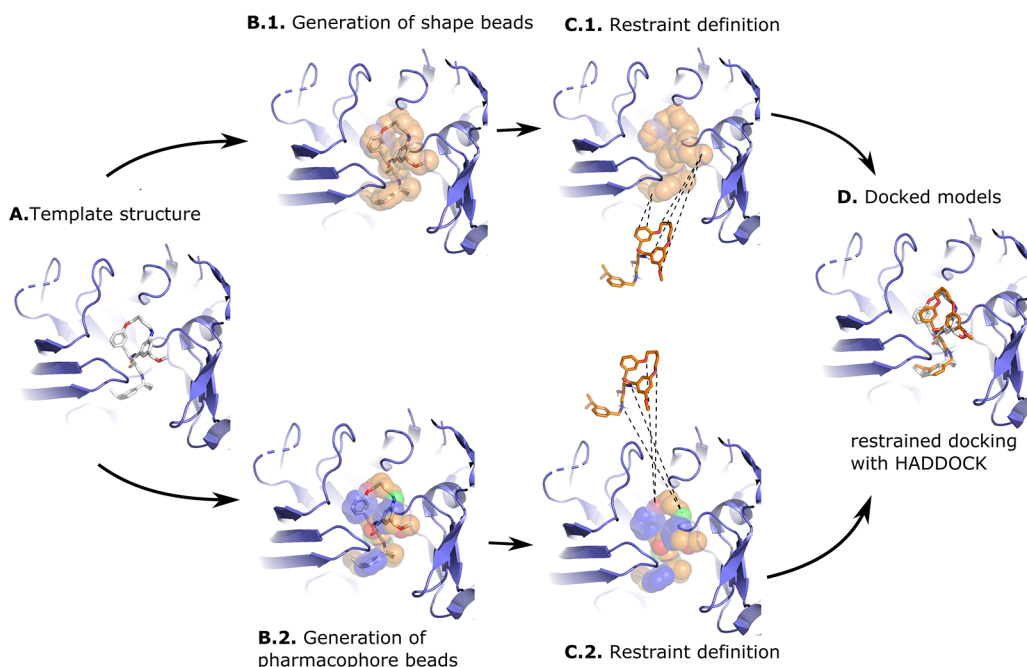
B.1. Generation of shape beads

C.1. Restraint definition

A. Template structure

D. Docked models

restrained docking with HADDOCK

B.2. Generation of pharmacophore beads

C.2. Restraint definition

Figure 1. Illustration of the shape-based and the pharmacophore-based docking protocol. Panel (A) shows a suitable receptor template identified based on the similarity of its bound ligand and the ligand to be docked (see Materials and Methods). Panel (B) shows the heavy atoms of the crystallographic compound transformed into shape beads (B.1) or pharmacophore beads with the colored beads representing different pharmacophore properties (B.2). The crystallographic compound is then removed from the pocket, and restraints are defined between the beads and the conformers: In the shape-based protocol (C.1) restraints are defined between all atoms of the compound and all beads of the shape, while in the pharmacophore-based protocol (C.2), restraints are defined between atoms of the compound and beads that share identical pharmacophore features. Panel (D) shows a docked model superimposed onto the template structure. The protein receptor is shown as slate cartoon, the crystallographic compound as white sticks, the generated and docked compounds as orange sticks, and the shape beads as transparent orange spheres. All molecular graphics were generated with PyMOL.[26]

heavy-atom RMSD to the reference bound ligand conformation is ≤2 Å for all but five target compounds. Specifically, the mean percentage of acceptable poses is 66 ± 36 (median: 78); however, there are 11 targets for which less than 10% of the generated conformers are acceptable, and five of those have no generated poses below 2 Å (see Figure S2).

Docking. Figure 1 shows the docking strategy behind the shape- and pharmacophore-driven protocols. In short, after selecting a suitable template, a shape based on the template compound is generated, restraints are defined between template shapes and target conformers, and docking is performed with HADDOCK (see Materials and Methods for more details).

Figure 2 shows the success rate for the two protocols with the "shape" and "pharm" groups highlighting the performance of the shape- and pharmacophore-driven protocols, respectively. The success rate is defined as the percentage of targets for which a model of acceptable (or better) quality has been generated within the top N ranked models based on the HADDOCK score. For the shape protocol, the performance is very poor when considering only the top ranked models during the rigid body stage (it0) but improves significantly in the refinement stage (it1) increasing to 58.6 and 71.7% when considering the top 1 and top 5 models, respectively, with an overall success rate of 81.8% when considering all refined models. The respective values for the pharmacophore protocol are 53.5, 63.6, and 75.8%, indicating a slightly lower performance compared to the shape protocol. Note again that those success rates are for an unbound docking scenario
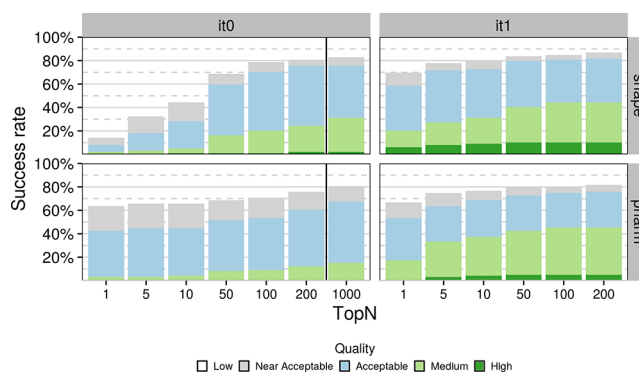

Figure 2. Success rate as a function of N-ranked models considered based on the HADDOCK score (top N) for tops 1, 5, 10, 50, 100, 200, and 1000 models for the initial rigid-body docking (it0) and after flexible refinement (it1). Each stacked bar plot corresponds to a specific cutoff with the color of the bar indicating the distribution of model quality for that cutoff. Dark green, light green, light blue, and light grey correspond to high-, medium-, acceptable-, and near-acceptable-quality models, respectively. The respective RMSD cutoffs are ≤0.5, ≤1, ≤2, and ≤2.5 Å when considering all heavy atoms of the compound after superimposing on the binding site backbone atoms of the receptors. Results have been grouped by protocol on the y axis and stage on the x axis. The black vertical line in the "it0" indicates the cutoff at which models proceed to flexible refinement.

(i.e., both ligand and receptor starting conformations deviate from their reference bound conformations in the complex).

This pattern holds even if we include the "near-acceptable" models—models whose RMSD to their respective receptor lies
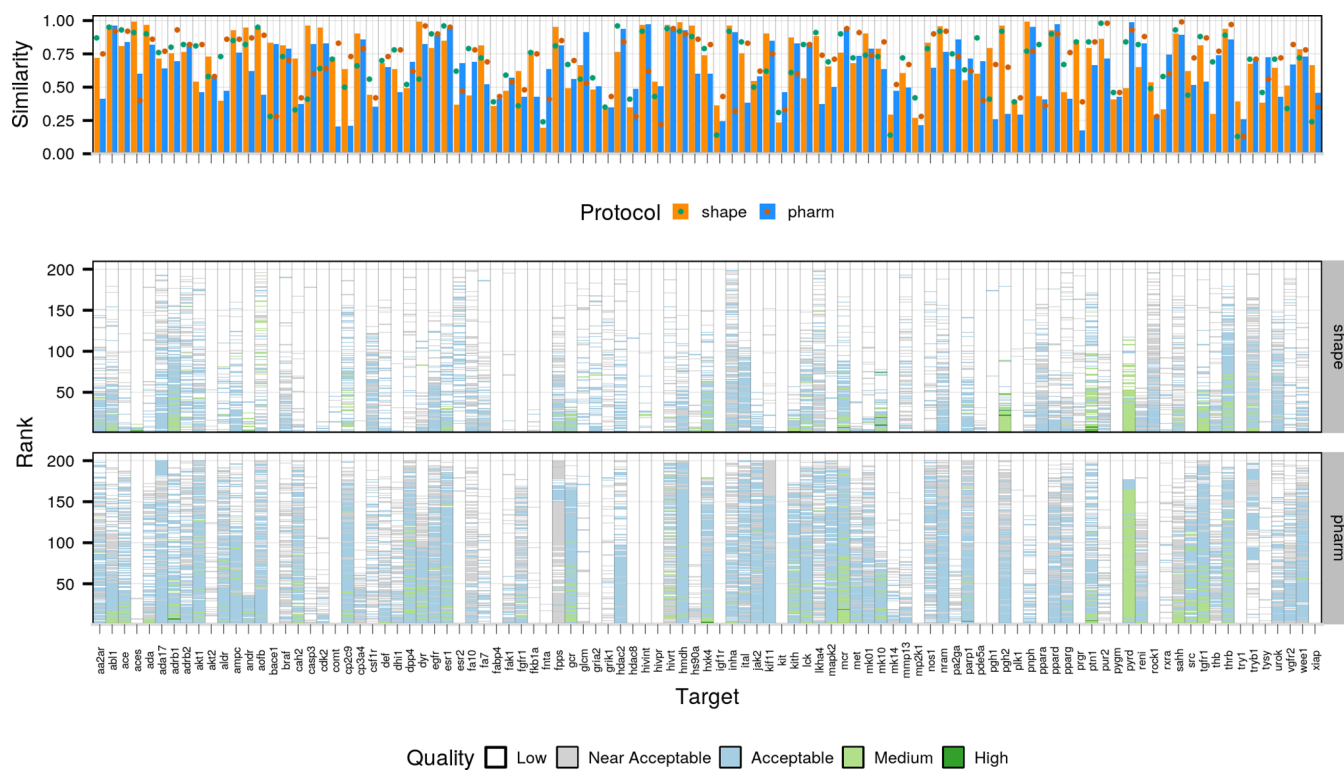
Figure 3. Comparison of template quality and performance for the two protocols. The top panel highlights the similarity (orange and blue bars for the shape- and pharmacophore-based protocols, respectively) and overlap (green and dark green dots for the shape- and pharmacophore-based protocols, respectively) between the reference and template compounds. The bottom panel compares the performance for the semi-flexible refinement stage of HADDOCK (it1) for the two protocols. Each column corresponds to one target with the *y* axis reflecting the ranking of models (ranks close to 0 refer to top-ranked models and those close to 200 to bottom-ranked models) and the color of each model reflecting its quality (see description of Figure 2).

between 2 and 2.5 Å—which cannot be considered near-native but are still likely to offer very relevant biological insights. When including these models, the respective percentages become 69.7, 77.8, and 86.7 for the shape top 1, top 5, and overall, and 66.7, 74.8, and 81.2 for the pharmacophore top 1, top 5, and overall, respectively. The performance of the pharmacophore protocol when considering only the top ranked model of it0 is worth mentioning as the success rate (when including the near-acceptable models) stands at 63.6%, which makes it particularly impressive considering only rigid-body motions are allowed in it0. Despite this, the distribution of model quality also seems to favor the shape protocol as higher-quality models are generated for more targets at all stages and cutoffs.

The success rate, however, only tells one side of the story, as can be seen in Figure 3, which shows the distribution of model quality as a function of the rank for all targets for the refinement stage (Figure 3, bottom panel). Comparing performance of the two protocols in this way reveals that the pharmacophore protocol produces models of acceptable (or better) quality much more consistently than the shape protocol but the shape protocol has higher coverage (see Figure S3 for rigid-body stage results). The top panel of Figure 3 shows the similarity between the template and reference compounds (using the Tversky and Tanimoto metrics, respectively; see Materials and Methods for details) in bars and the 3D shape overlap in points. Comparing the performance of the two protocols with these metrics in mind reveals the 3D shape overlap to be a limiting factor for the outcome of the modeling, acting almost like a binary classifier, with targets whose overlap

is below 0.5 almost never achieving good performance, whereas for those whose overlap is greater than 0.5, success is almost guaranteed, in particular for those whose overlap exceeds 0.75 (e.g., *adrb1* and *mcr*). Further comparing the performance of the two protocols reveals that overlap is the feature that determines the outcome even when considering the same target. For example, for target *aces*, the similarity and overlap values are 0.99 and 0.91 and 0.6 and 0.4, for the shape- and pharmacophore-based protocols, respectively, with the shape protocol yielding top-ranked, high-quality models and the pharmacophore-based protocol only producing near-acceptable models despite a lower binding site RMSD for the template chosen for the pharmacophore-based protocol (0.09 Å against 0.2 Å for the shape-based protocol).

To further investigate the impact of template selection on the outcome of the modeling, we compared the performance of the shape-based protocol in a subset of the dataset for which we identified both high- (similarity ≥0.8) and low-similarity templates (similarity around 0.4). Not only is the difference in terms of success rate stark, with the high-similarity subset observing success rates of 76.5, 88.2, and 91.2% when considering top 1, top 5, and overall, respectively, against 44.1, 47.1, and 73.5% for the low-similarity subset (still a very decent performance), but the simulations with the high-similarity templates also clearly results in distributions that are more populated with higher-quality models (see Figure S4).

Figure 4A,B illustrates a docking scenario in which both shape-restrained protocols are expected to perform well due to the high similarity between the template and the target compounds (Tv = 0.714 and Tc = 0.641). Both protocols

generate acceptable-quality models at top 1, with IL-RMSD (interface ligand-RMSD; see Materials and Methods for details) values of 1.60 and 1.37 Å for the shape- and the pharmacophore-based protocols, respectively. Figure 4C,D
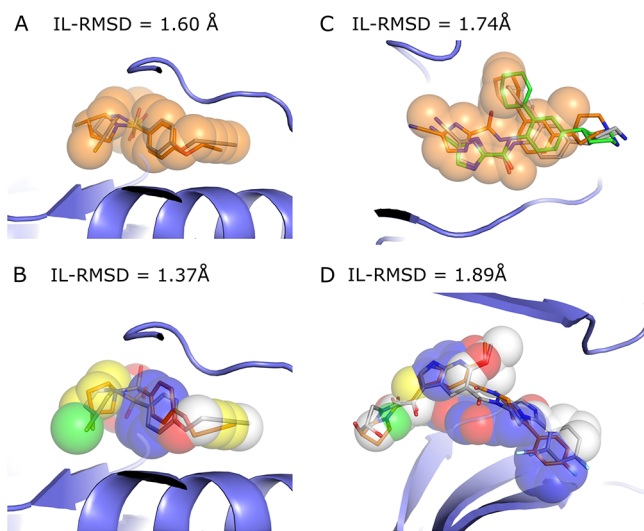

Figure 4. Illustration of shape-restrained docking outputs (orange sticks) given the input shape (spheres) as compared to the target binding mode (grey sticks). Panels (A) and (B) represent the top 1 models generated by the shape- and the pharmacophore- based docking respectively for ada17 (ideal scenario), based on template 3b92. Panel (C) represents the best docking model generated in top 5 (orange) and overall (green, with an IL-RMSD value of 1.74 Å) for csf1r with the shape-based protocol, based on template 2i0v. Panel (D) represents the best docking model generated in top 5 for mk14 with the pharmacophore-based protocol, based on template 3 ha8. The color code for the pharmacophore spheres is the following: H-bond donor (green), H-bond acceptor (red), hydrophobic (yellow), aromatic (blue), and regular shape bead (white)

illustrates a more difficult scenario where the selected template differs substantially from the target compound (csf1r, Tv = 0.44 and mk14, Tc = 0.47; see Figure 6). We still observe good performance, with acceptable models generated at top 2 for csf1r (1.74 Å) with the shape-based protocol and at top 5 for mk14 (1.89 Å) with the pharmacophore-based protocol. This is mainly explained by a good shape overlap between the template and the target compounds (0.56 and 0.52).

Impact of the Flexible Refinement. As mentioned previously, we observe a significant improvement in the success rate at it1 as compared to it0 for the shape-based protocol and a moderate improvement for the pharmacophore-based protocol. This underlines that the semi-flexible refinement plays a crucial role in improving the quality of the models generated at it0 as supported by Figure 5A. However, while we would expect moderate improvement related to local rearrangements in the binding site and relaxation of the small molecule, we also observe impressive and unexpectedly large improvement in the small-molecule binding mode between it0 and it1, especially in the shape-based protocol. This can be clearly seen in the distribution of ΔIL-RMSDs for the shape protocol (orange distribution in Figure 5A) where the shoulder extends to up to 5 Å (i.e., 5 Å improvement) for the acceptable models after refinement. The distributions of the RMSD between the aligned template and reference small molecules, referred to as the ligand−ligand RMSD, show that

both protocols also tend to generate models in which the conformation get closer to the reference for most targets (Figure S5). Three eloquent visual examples are provided in Figure 5B. The first one shows that the IL-RMSD of the top ranking it1 model for target *lck* (0.4 Å) is improved by 5.2 Å as compared to the corresponding model at it0. This target is considered an "easy case" since it is associated with an overlap of 0.82 and a binding site RMSD of 0.1 Å between the target and the template. The second example, an "intermediate case" (overlap of 0.57, binding site RMSD of 0.6 Å), shows that the IL-RMSD of the top 5 it1 model for target *gria2* (1.0 Å) is improved by 4.1 Å. Finally, we present a "difficult case" with the target *pa2ga* for which no acceptable conformer was generated. Based on the excellent template that is selected by the pharmacophore-based protocol, the generated model ranked top 10 at it1 significantly improves on the ligand−ligand RMSD metric (Figure S5) as well as the IL-RMSD to the reference (it1 IL-RMSD = 1.8 Å with a ΔIL-RMSD = 3.9 Å). The ligand−ligand RMSD values are 0.35, 0.76, and 1.72 Å for the three cases, indicating the excellent agreement in binding mode between our models and the reference structures.

The overall larger improvement observed for the shape-based protocol as compared to the pharmacophore-based protocol could be explained by the higher degree of freedom provided by the dummy beads as compared to the pharmacophore beads.

Peculiar Cases. Since we always learn from errors, we investigated cases where shape-restrained docking fails. As the first HADDOCK step (it0) keeps both the receptor and the ligand rigid, it is essential to provide as a starting point different conformers of the ligand to be docked. When the conformer generator fails in generating near-native conformations, the docking is likely to fail as well. Our results confirm this assumption as we observed that few to no acceptable models were generated in seven cases with less than 10% acceptable conformers generated (e.g., *cdk2*, *cp3a4*, *fkb1a*, *fnta*, *hivpr*, *pa2ga*, and *xiap*). A second source or error is the binding site location of the template compound that may differ from the binding site of the target compound (Figure 6B.5), as highlighted by the Pearson correlation of 0.72 and 0.59 observed between the distance in geometric center of the template and the target compounds and the best IL-RMSD obtained with the shape- and pharmacophore-based protocol, respectively (Figure S6). External expert-eye or experimental knowledge should alleviate this issue and help filtering out potential erroneous template compounds in early stages of the protocol. For this work, we however followed an automated pipeline. Third, if both protocols are capable of generating good-quality models where the template compounds show poor similarity with the target compounds (Figure 6A), the opposite also happens. In some cases, our initial assumption stating that similar compounds that bind the same target receptor should in principle adopt similar binding modes is not validated. This is the case for the *hivint* and *bace1* targets and template compounds as illustrated in Figure 6B. The different structures of *hivint* complexed with *N*-benzyl indolin-2-ones reported in the original paper of the target structure[27] display various binding modes of the compounds in the same binding site. In the case of *bace1*, two binding modes of iminohydantoin bace1 inhibitors are reported with a 180° flip of the iminohydantoin core.[28] Finally, a visual inspection could beneficially help in discarding template compounds that
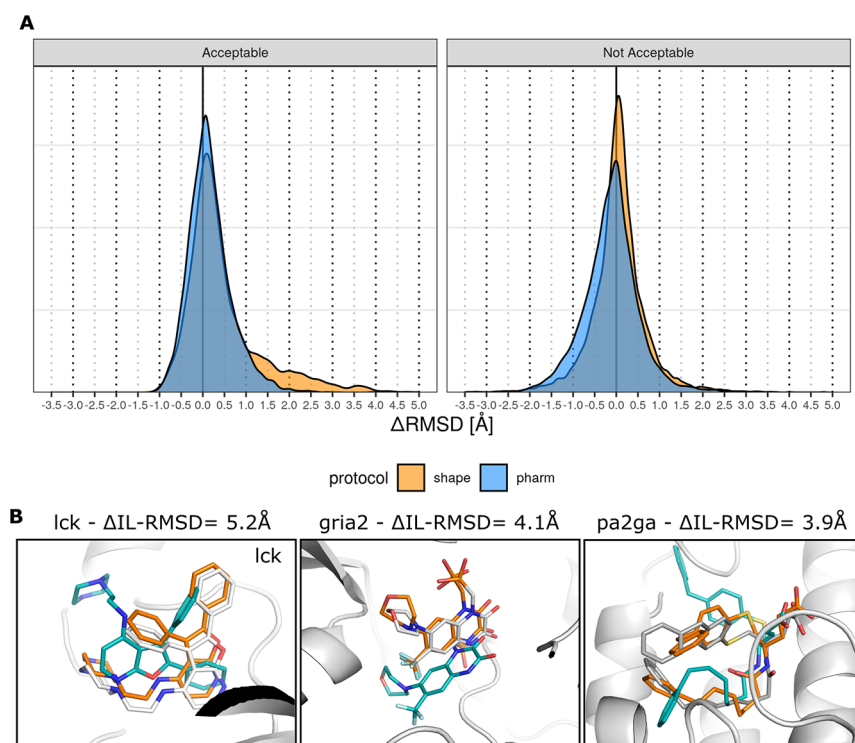
Figure 5. Illustration of the impact of the semi-flexible refinement on the docking model quality. Panel (A) shows the distribution of the ΔIL-RMSD between the models before and after semi-flexible refinement, computed as the IL-RMSD at it0, the IL-RMSD at it1 for a given model. The figure is divided into two parts: the left subpanel shows the distribution of ΔIL-RMSD for models that are acceptable at it1, the right subpanel shows the distribution of ΔIL-RMSD for models that are not acceptable at it1. Positive values in the distribution indicate better it1 performance. Panel (B) illustrates three cases with large IL-RMSD improvement between the model generated at it0 (blue) and the refined model at it1 (orange) as compared to the bound conformation in the reference structure (grey). Templates 2of4, 3bki, and 1kqu were used for cases lck, gria2, and pa2ga, respectively.

share poor similarity with the target compound in terms of shape, flexibility, and physicochemical properties despite a meaningful similarity score (e.g., Figure 6B.5).

Which Protocol Should Be Favored? The shape-based protocol yields better performances in terms of coverage, i.e., it generates near-native models, including near-acceptable models, for more DUD-E complexes (top 1: 69.7%, top 5: 77.8%, overall: 87%) than that of the pharmacophore-based protocol (top 1: 66.7%, top 5: 74.8%, overall: 81.2%). In an ideal scenario where we merge the results of the two protocols and pick up the best model (i.e., lowest IL-RMSD) per rank, the global coverage is remarkably improved (top 1: 81%, top 5: 86%, overall: 91%). As illustrated in Figure 2, it means that even though the protocols display comparable performance on most targets, one protocol could be favored over the other in some cases (e.g., *aces, fkb1a, gria2, hivint, mapk1, mk14, pde5a, prgr*). We identified the overlap between the template and the target compounds shapes as the determinant indicator of the success of the shape- and the pharmacophore-based docking (Figure S7). The overlap does not only inform on the proximity between the binding sites of the template and the target compounds, it also tells about the binding mode similarity. However, this information is accessible only in benchmarking conditions where the crystallographic binding mode of the target ligand is known. To mimic real-life conditions, we investigated if different combinations of descriptor filters extracted from the unbound form of the target and the template compounds (similarity, difference in molecular weight, number of rotatable bounds, and accessible

surface area) could lead to an enrichment in the success rate. We identified no evident conditions to favor one protocol over the other, but it is worth mentioning that both protocols achieve high success rates when the template and target share moderate to high similarity (≥∼0.4 or 0.3 for shape and pharmacophore, respectively) (Figure 7). As such, the shape-based protocol performance (including near-acceptable models) rises to 74.1, 83.5, and 92.9% at top 1, top 5, and overall (85 targets), respectively, when considering a Tversky similarity ≥0.4. With similar conditions, i.e., a Tanimoto coefficient over 0.3, the performance stands at 71.9, 80.9, and 87.6% at top 1, top 5, and overall (89 targets), respectively, with the pharmacophore-based protocol. Adding a molecular weight filter ($|MW_{ref} - MW_{temp}| < 50$ g/mol) slightly improves the performance, which rises to 77.5 and 89.8% for the shape top 1 and top 5, and reaches an impressive overall performance of 100% when considering all generated models. A lower overall performance is reached with the pharmacophore-based protocol (93.2%), but we observed a higher early enrichment of 81.4% at top 1. Of note, 49 and 59 DUD-E targets fulfill these two criteria for the shape- and pharmacophore-based protocols, respectively. All the aforementioned analyses suggest that the shape-based protocol should be favored for the sake of coverage. However, the conditions we identified could be used to favour one of these protocols depending on the available data and the intended goal of the study.

Comparison to State-of-the-Art Docking Tools. A similar benchmarking study was conducted with four commercial docking programs, namely Gold,[29] Glide,[30−32] Surflex,[33] and
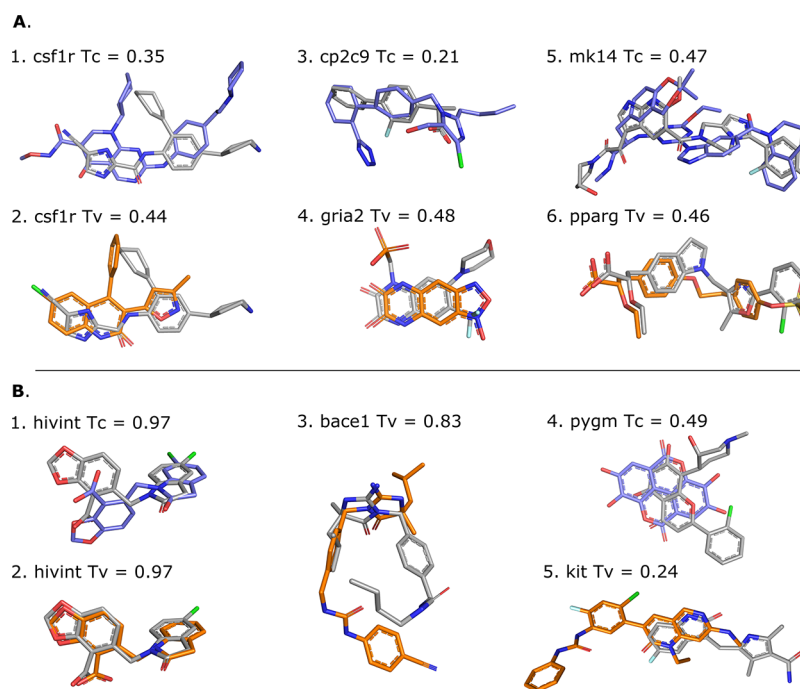
Figure 6. Illustration of template (shape-based protocol: orange, pharmacophore-based protocol: blue) and target (grey) compounds. Panel (A) shows different template/target compound couples of low similarity, i.e., Tanimoto (Tc) or Tversky (Tv) coefficient of <0.5, which lead to successful docking. Templates 3dpk, 2i0v, 5x23, 3bki, 3ha8, and 1i7i were used for cases 1−6, respectively. Panel (B) shows scenarios that lead to docking failure for different possible reasons. While displaying high similarity with the target compounds, the hivint pharmacophore-based template compound adopts a different binding mode (B.1), and the shape-based template compounds leads to an unexpected failure that cannot be explained neither by the coverage nor their respective binding site (B.2). The shape-based template compound for bace1, which is the same as the pharmacophore-based template, adopts a different binding mode compared to the target despite the high similarity (B.3). Low shape, flexibility, and physicochemical similarities are observed between the template and the target compound for pygm (B.4); case B.5 is an example of a poor overlap between template and target compounds that is responsible for poor performance. Templates 3nf9, 3nf6, 3l5c, 3g72, and 6mob were used for cases 1−5, respectively.
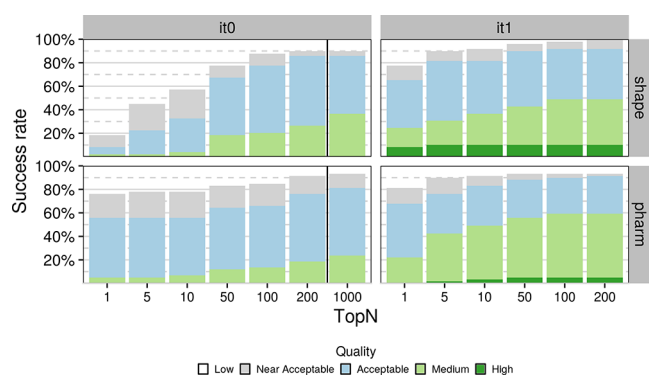

Figure 7. Success rate of the shape- and pharmacophore-based protocol on DUD-E complexes (49 and 59, respectively) that match two criteria: (1) the similarity coefficient between the template and the reference compounds is higher than 0.4 and 0.3, respectively, and (2) the difference in heavy atom molecular weight is below 50 g·mol^{-1}.

FlexX,[34] which are all widely used in the industry[21] and have better sampling power than most non-commercial docking programs.[35] In this study, two of the 102 DUD-E targets (*aofb* and *casp3*) were discarded as they show covalent binding to their respective receptor. Of note, those two targets are part of our dataset and the three we are missing are part of their dataset. To fairly compare the performance, we discarded *aofb* and *casp3* from our results. Commercial programs were evaluated in bound−unbound docking mode, i.e., unbound

conformations of the compounds have been docked onto the bound conformation of the receptor (i.e., no conformational changes required on the protein side), while we did perform fully unbound docking, which represents a significantly more difficult yet more realistic scenario.[11] It is worth mentioning that the four commercial docking software also make use of information in different ways to guide the docking, for example, by specifying the bounding box around the binding site or explicitly defining the binding site residues. Our two protocols show very competitive overall performance with those commercial docking programs, the shape-based protocol ranking second with a success rate of 81% just behind Surflex (84%) (Figure S8) and the pharmacophore-based protocol ranking fifth with 74%. We also observe a rather competitive early enrichment with 54 and 58% success at top 1 for the fully unbound pharmacophore- and shape-based protocols against 64 and 65% for Gold and Glide, respectively. It is interesting to note that all commercial software fail to generate near-acceptable models for six DUD-E targets on which at least one of our protocols also fails (*bace1*, *cdk2*, *fkb1a*, *fnta*, *hivint*, *pgh1*) for different reasons (see Peculiar Cases).

■ CONCLUSIONS AND PERSPECTIVES

We have presented two new protocols for the modeling of protein−small-molecule complexes with HADDOCK, which make excellent use of the integrative modeling capabilities of the platform, to yield very promising results. Selection of suitable templates for the modeling based on ligand similarity and use of shape data during the docking lie at the core of both

approaches. In the first, the shape is only defined in geometrical terms, whereas in the second, it is additionally annotated with the pharmacophore data, thus naming the two protocols shape- and pharmacophore-based. The dataset on which these protocols were benchmarked is based on the widely used DUD-E dataset, which was extended to make use of unbound receptors (or rather receptors bound to another ligand, which is commonly defined as cross-docking). This "unbound" DUD-E dataset, which is freely available in the ready-to-dock format on GitHub, is composed of diverse targets of various difficulty degrees. Our analysis has revealed that the main limiting factor for the outcome of the docking with our shape protocols is the 3D overlap between template and reference compounds, in other words, whether the template and reference compounds not only occupy the same binding pocket but have similar binding modes. We have also shown that we can predict the reliability of the modeling based on features that can be computed in real-world usage: the difference in molecular weight between the template and reference compounds and their molecular similarity. When we select the subset of our dataset for which the compounds are not too dissimilar in terms of molecular similarity (Tv and Tc greater or equal than 0.4 and 0.3 for the shape- and pharmacophore-based protocols, respectively) and molecular weight ($|MW_{ref} - MW_{temp}| < 50$ g/mol), we achieve success rates of 65.3, 81.6, and 91.8% and 67.8, 76.2, and 91.5% when considering top 1, top 5, and overall for the shape- and pharmacophore-based protocols, respectively. Using all targets, the performance remains rather high with 58.5, 71.7, and 81.8% and 53.5, 63.6, and 75.8% for the two protocols. The performance of our protocols is on par with that of the leading small-molecule docking software, despite the fact that our approach was benchmarked on a fully unbound dataset. As such, it is representative of the performance that one can expect when using this pipeline in a real-life scenario where one would not have access to the bound form of the receptor and/or that of the compound.

Despite this promising performance, our method still suffers from a few limiting factors. Chief among those is the dependency on a template whose compound occupies the same binding pose as that of the ligand to be modeled for a successful prediction. This was identified as the most severe limiting factor for both protocols. It is a problem that is not unique to these protocols but is shared by all data-driven methods, HADDOCK being no exception.[36] Overcoming this obstacle is no easy task and will require the development of a methodology for the automated identification, combination, and weighting of multiple template shapes and receptors in order to increase the coverage of the binding site. Another area where these protocols could benefit from further improvements is the ranking of solutions during the flexible stage as can be seen by the difference in the success rate at the top 1 or top 5 level and overall. Also worth noting is the fact that although regular users of the command-line version of HADDOCK will be able to make use of this pipeline, it might be well out of reach for most researchers who have no expertise in programming and the use of computational chemistry toolkits. Therefore, it would be of interest if these protocols were to be made available through the freely accessible HADDOCK webserver (https://wenmr.science.uu.nl/haddock2.4/). We hope to expand on these three points in the near future. A final point relevant for researchers who want to set up high-throughput screenings based on this protocol are

the CPU time requirements. The total time is going to heavily depend on the available computing resources. As an indication for a system of average size (target "aa2ar" with 15 binding site residues), the rigid-body and flexible stages (it0 and it1, respectively) take around 6 and 170 s, respectively, per generated model using a single thread on an AMD EPYC 7451 processor. As every generated model is independent of the others, HADDOCK can be run very effectively in HPC/HTC systems (the HADDOCK server actually makes use of the EGI-distributed high-throughput resources around the world[37]). Therefore, with enough computing resources allocated, the total simulation time can be in principle as low as the required CPU time per model for the two stages.

Limitations and future improvements mentioned in the previous paragraph notwithstanding, the protocols we are proposing achieve high performance on experimental conditions that mimic real-life ones as closely as possible using unbound structures for receptors as well as compounds. Furthermore, the performance we are reporting is on par with that of leading commercial docking platforms when benchmarked on partially bound datasets, and in many cases, exceeds it, thus making it an excellent option for the modeling of protein−small-molecule complexes.

■ MATERIALS AND METHODS

Dataset. To validate our new approach, we decided to benchmark the performance of these protocols against the Enhanced Database of Useful Decoys dataset (DUD-E, http://dude.docking.org/), one of the most widely used small-molecule docking benchmarks.[20] In total, the DUD-E consists of 102 targets with each target corresponding to a single protein receptor (usually of pharmaceutical interest) with a compound bound to it. The Protein Data Bank (PDB)[38] entries associated with every DUD-E target became our reference receptors. We discarded three targets (*cxcr4*, *drd3*, *kpcb*) because we could not identify any viable templates for them (see next section) bringing our total to 99 targets. The structures of the reference receptors were downloaded from the PDB instead of the DUD-E website to avoid any postprocessing that might have been applied to them.

Template Identification. The procedure we followed for identifying a template for each of the 99 targets in our dataset can be summarized in the following steps:

(1) Search the PDB for appropriate template structures. We used two approaches in conjunction with a few criteria to identify templates:

(a) Search for receptors whose sequence is more than a predetermined cutoff similar to the sequence of our reference receptor. The precise value we used for the sequence similarity cutoff differed from target to target because of the intrinsic features of some targets. As an example, some targets were membrane receptors that had been co-crystallized with a lysozyme domain. The presence of the lysozyme sequence, with lysozyme being one of the most commonly encountered domains in the PDB, affected the results, and we had to take that into account when adjusting the sequence similarity cutoff. The default value was 90% (calculated over the entire sequence) and, in cases such as the one mentioned above, was adjusted after manual inspection of the results.

(b) Search for receptors bound to compounds that are similar to the reference compound. For this, we made use of the "Fingerprint Similarity" feature of the "Chemical" tab of the advanced search functionality of the RCSB portal of the PDB (https://www.rcsb.org/search/advanced).

(2) After pooling the results from both searches, we iteratively removed unsuitable templates based on a few criteria:

(a) Templates that contained only irrelevant compounds such as crystallization buffer artifacts or some of the most commonly encountered compounds such as hemes.

(b) Incorrect templates based on sequence identity (regarding the templates identified through ligand similarity).

(c) Templates whose compounds were found to be stereoisomers of the reference compounds.

(3) After removal of all unsuitable templates, we calculated the pairwise similarity between all reference and template compounds.

(a) Shape-based protocol: The similarity metric we chose for this is the Tversky similarity[39] (with weights for the reference and template compounds set to 0.8 and 0.2, respectively)[40,41] computed over the maximum common substructure (MCS) as identified with the rdFMCS implementation of RDKit (version 2020.09.3).[42]

(b) Pharmacophore-based protocol: We computed the Tanimoto coefficient over 2D pharmacophore fingerprints generated with RDKIT[42] using the default pharmacophore fingerprint factory.

(4) After calculating all similarities, we rank the templates according to the similarity of their compound to the reference compound and select the one with the highest similarity with the condition that the template and the target share the same binding site (defined as <5 Å around the bound small molecule), i.e., 100% sequence identify in the binding site. We removed water and co-solvent molecules from the selected templates, while cofactors and ions were preserved.

Conformer Generation. To ensure that there is no bias during the docking, instead of starting from compounds bound to the reference or other receptors, we generate 3D conformations of the target compounds starting from their isomeric SMILES[43,44] with RDKit using the 2020 parameters (only the small aliphatic ring subset) with energy minimization and the ETKDG algorithm.[45,46] We cap the maximum number of conformers to 50 and provide the ensemble of conformers to HADDOCK for docking.

Docking. *System Preparation.* Prior to docking, we process the structures to ensure that the residue numbering between template and reference structures is consistent (a necessity for the analysis), rename residues, cofactors, and ions according to HADDOCK specifications, remove nonbiological symmetric units, and remove all irrelevant artifacts such as crystallization buffers. We generated topology files of the compounds to be docked using PRODRG (version 070118.0614)[47] compatible with the united-atom OPLS forcefield HADDOCK uses.

Shape-Based Protocol. After identifying one receptor template per target, we transform all heavy atoms of its

compound into dummy beads (those do not interact with the remaining of the system). We then define ambiguous distance restraints with an upper limit of 1 Å between the shape beads and the heavy atoms of the compound to be docked. The distance restraining function used in HADDOCK is based on the ambiguous NOE restraint implementation for NMR by Nilges.[48] The key aspect is the computation of an effective distance by summation over all possible distance pairs defined by the atom selections as $r_{\mathrm{eff}} = \left(\sum_1^N \frac{1}{r^6} \right)^{-1/6}$ where the sum runs over all N atom−atom distance pairs. This effective distance enters a harmonic "flat-bottom" potential (or square-well with harmonic walls) with a linear behavior for large deviations (see eq 8 in ref 48).

The nature of the restraints creates an additional consideration; specifically, what should be the "origin" and "target" of the restraints? In this protocol, the directionality of the restraints depends on the size of the reference and template compounds in terms of heavy atom count. Specifically, the restraints are always defined from smaller to larger. If the reference compound is larger than the template compound, the restraints are defined from each shape bead to any compound heavy atom. If the opposite is true (the template compound is larger than the reference compound), then restraints are defined between each heavy atom of the conformer to any shape bead. These ambiguous distance restraints effectively enforce that ligand atoms and beads must overlap. Depending on the directionality of the restraints, part of the ligand might remain outside the shape defined by the beads and vice-versa. All shape restraints are used during the simulation.

Pharmacophore-Based Protocol. In addition to the transformation of the template atoms into dummy beads, we assigned pharmacophore features to the beads to add physicochemical properties to the spatial information they hold. We assigned pharmacophore features (H-bond donor, H-bond acceptor, negative ionizable, positive ionizable, zinc binder, aromatic, hydrophobic, and lumped hydrophobic) to each atom with RDKIT using the default SMARTS-based feature definition. If no pharmacophore feature is assigned to an atom, a regular shape bead is defined. We defined distance restraints with an upper limit of 1 Å between the beads and the atoms of the compounds to be docked provided that they share the same pharmacophore feature or no pharmacophore feature. In this protocol, we cannot impose restraints from the smaller to the larger anymore because restraints are defined between features of the same class—some classes may be more populated in the reference compound, while others may be more populated in the template compound. Instead, we defined all restraints from the compound to the beads. Similar to the shape-based protocol, all restraints are considered during the docking process.

Figure 1 illustrates the shape-based and the pharmacophore-based docking protocol.

HADDOCK. For the docking, we used the January 2021 release of the command-line version of HADDOCK 2.4. The number of models generated in the initial rigid-body docking stage (it0) of HADDOCK is set to 20 times the number of starting ligand conformations. This sampling ratio was found to be the most efficient during benchmarking (data not shown). If the number of it0 models that are sampled is larger than 200, then only the top 200 it0 models proceeded to flexible refinement; otherwise, all models do. We only perform the semi-flexible refinement stage of HADDOCK (it1) and

skip full water refinement (itw) as this final stage is not improving the results as already remarked from our D3R participation. The positions of both the receptor and its associated shape are fixed in their original orientations, while the ligand is translated away from the protein and randomly rotated for each docking trial. The shape is kept rigid throughout the protocol, while the receptor interface and the ligand become flexible during the refinement stage. Systematic sampling of 180° rotations along the interface is disabled for it0. We also scale down the intermolecular interactions during the rigid-body stage to facilitate the insertion of the ligand into the binding pocket and accordingly exclude the vdW energy term during the scoring of the rigid-body models. The modified parameter settings for HADDOCK are summarized in Table S3.

Other than the above defined modifications, the scoring function used is the default scoring function of HADDOCK. Its functional form, specific for protein−ligand docking for the two stages is as follows:

$$HS - it0 = 0.0 \times E_{vdw} + 1.0 \times E_{elec} + 1.0 \times E_{desolv}$$
$$+ 0.01 \times E_{AIR} - 0.01 \times BSA$$

$$HS - it1 = 1.0 \times E_{vdw} + 1.0 \times E_{elec} + 1.0 \times E_{desolv}$$
$$+ 0.1 \times E_{AIR} - 0.01 \times BSA$$

where E_{vdw}, E_{elec}, and E_{desolv} stand for van der Waals, Coulomb electrostatics, and desolvation energies, respectively, and BSA stands for the buried surface area. The non-bonded components of the score (E_{vdw}, E_{elec}) are calculated with the OPLS forcefield.[49] The desolvation energy is a solvent-accessible surface area-dependent empirical term,[50] which estimates the energetic gain or penalty of burying specific sidechains upon complex formation.

Evaluation of Results. We evaluated the quality of the generated models according to their structural deviation from the reference structures. For this, we used the interface-ligand RMSD (IL-RMSD), which is the RMSD calculated over all heavy atoms of the ligand after superimposing on all backbone atoms of the interface of the receptor. Models with an IL-RMSD of less than 0.5 Å, between 0.5 and 1 Å, between 1 and 2 Å, between 2 and 2.5 Å, and over 2.5 Å are classified as high-, medium-, acceptable-, near-acceptable-, and low-quality models, respectively. The initial fitting was performed using the McLachlan algorithm[51] as implemented in the program ProFit (Martin, A.C.R., http://www.bioinf.org.uk/software/profit/, available through SBGrid[52]). Calculation of symmetry-corrected RMSD values for the compounds was performed with *obrms* from the Open Babel distribution (https://github.com/openbabel/openbabel, version 3.1.1).[53] As part of the analysis and interpretation of the results, we also present a classification of target difficulty based on the 3D shape overlap of the target and reference compounds after fitting on the binding site backbone atoms of their respective receptors. Fitting was performed with ProFit using the same settings as previously mentioned, and the 3D shape overlap was calculated with the *exact overlap* metric of the shape toolkit of OpenEye (release 2020.2.0).[54−56]

Data and Software Availability. All docking models generated during the benchmarking of the two protocols with HADDOCK 2.4 are made available through our laboratory data collection at https://data.sbgrid.org/labs/32/[57] (Note for

reviewers: the full dataset will be deposited upon acceptance). The unbound DUD-E benchmark data set, with all the code, docking input and parameter files, results, and analysis files are made available through GitHub (https://github.com/haddocking/shape-restrained-haddocking). Further, a detailed tutorial illustrating the setup of both shape-restrained and pharmacophore-restrained docking using the HADDOCK2.4 web portal is available at https://www.bonvinlab.org/education/HADDOCK24/shape-small-molecule/.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00796.

Table of shape-based template data and performance per target (Table S1); table of pharmacophore-based template data and performance per target (Table S2); differences between shape/pharmacophore-based protocols and default HADDOCK settings (Table S3); distribution of shape- and pharmacophore-based protocol template similarity values (Figure S1); distribution of RMSD values of generated conformers against reference compound structure (Figure S2); distributions of model quality against ranking based on HADDOCK score for the two protocols, for the rigid body and simulated annealing stages (Figure S3); distributions of model quality against ranking based on HADDOCK score for the shape-based protocol comparing the performance in the low- and high-similarity subsets and distribution of similarity and 3D shape overlap values between template and reference structures (Figure S4); distributions of ligand−ligand RMSD values after conformer generation and at the end of the simulated annealing stage for the two protocols (Figure S5); correlation of various metrics against performance for the two protocols (Figure S6); evaluation of the success rate as a function of the difficulty of each target based on two criteria: ligand similarity and 3D overlap (Figure S7); comparison of the success rate between HADDOCK and various commercial software (Figure S8) (PDF)

Table for all targets, listing the template used for both protocols, the SMILES of the reference and template compounds, the binding site RMSD between reference and template structures, and the performance when considering top 1, top 5, and all generated models (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Alexandre M. J. J. Bonvin − *Computational Structural Biology Group, Department of Chemistry, Faculty of Science, Utrecht University, Utrecht 3584CH, The Netherlands;* ⓞ orcid.org/0000-0001-7369-1322; Phone: +31 30 253 2652; Email: a.m.j.j.bonvin@uu.nl

Authors

Panagiotis I. Koukos − *Computational Structural Biology Group, Department of Chemistry, Faculty of Science, Utrecht University, Utrecht 3584CH, The Netherlands*

Manon Réau − *Computational Structural Biology Group, Department of Chemistry, Faculty of Science, Utrecht University, Utrecht 3584CH, The Netherlands*

■ ABBREVIATIONS USED
HADDOCK, High Ambiguity Driven DOCKing; DUD-E, Database of Useful Decoys: Enhanced; GC, Grand Challenge; D3R, Drug Design Data Resource; CSAR, Community Structure Activity Resource; IL-RMSD, interface ligand-RMSD; ΔIL-RMSD, delta IL-RMSD; MCS, maximum common substructure

■ REFERENCES
(1) Schneider, G.; Fechner, U. Computer-Based de Novo Design of Drug-like Molecules. *Nat. Rev. Drug Discovery* **2005**, *4*, 649−663.

(2) Mandal, S.; Moudgil, M.; Mandal, S. K. Rational Drug Design. *Eur. J. Pharmacol.* **2009**, *625*, 90−100.

(3) Schneider, G. Automating Drug Discovery. *Nat. Rev. Drug Discovery* **2018**, *17*, 97−113.

(4) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of Machine Learning in Drug Discovery and Development. *Nat. Rev. Drug Discovery* **2019**, *18*, 463−477.

(5) Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J. A.; Kang, Y. N.; Delproposto, J.; Kubish, G.; Dunbar, J. B., Jr.; Carlson, H. A.; Burley, S. K.; Walters, W. P.; Amaro, R. E.; Feher, V. A.; Gilson, M. K. D3R Grand Challenge 2015: Evaluation of Protein−Ligand Pose and Affinity Predictions. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 651−668.

(6) Gaieb, Z.; Liu, S.; Gathiaka, S.; Chiu, M.; Yang, H.; Shao, C.; Feher, V. A.; Walters, W. P.; Kuhn, B.; Rudolph, M. G.; Burley, S. K.; Gilson, M. K.; Amaro, R. E. D3R Grand Challenge 2: Blind Prediction of Protein−Ligand Poses, Affinity Rankings, and Relative Binding Free Energies. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 1−20.

(7) Gaieb, Z.; Parks, C. D.; Chiu, M.; Yang, H.; Shao, C.; Walters, W. P.; Lambert, M. H.; Nevins, N.; Bembenek, S. D.; Ameriks, M. K.; Mirzadegan, T.; Burley, S. K.; Amaro, R. E.; Gilson, M. K. D3R Grand Challenge 3: Blind Prediction of Protein−Ligand Poses and Affinity Rankings. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 1−18.

(8) Parks, C. D.; Gaieb, Z.; Chiu, M.; Yang, H.; Shao, C.; Walters, W. P.; Jansen, J. M.; McGaughey, G.; Lewis, R. A.; Bembenek, S. D.; Ameriks, M. K.; Mirzadegan, T.; Burley, S. K.; Amaro, R. E.; Gilson, M. K. D3R Grand Challenge 4: Blind Prediction of Protein−Ligand Poses, Affinity Rankings, and Relative Binding Free Energies. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 99−119.

(9) Carlson, H. A. Lessons Learned over Four Benchmark Exercises from the Community Structure-Activity Resource. *J. Chem. Inf. Model.* **2016**, *56*, 951−954.

(10) Kurkcuoglu, Z.; Koukos, P. I.; Citro, N.; Trellet, M. E.; Rodrigues, J. P. G. L. M.; Moreira, I. S.; Roel-Touris, J.; Melquiond, A. S. J.; Geng, C.; Schaarschmidt, J.; Xue, L. C.; Vangone, A.; Bonvin, A. M. J. J. Performance of HADDOCK and a Simple Contact-Based

(11) Koukos, P. I.; Xue, L. C.; Bonvin, A. M. J. J. Protein−Ligand Binding Affinity Predictor in the D3R Grand Challenge 2. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 175−185.

(11) Koukos, P. I.; Xue, L. C.; Bonvin, A. M. J. J. Protein−Ligand Pose and Affinity Prediction: Lessons from D3R Grand Challenge 3. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 83−91.

(12) Basciu, A.; Koukos, P. I.; Malloci, G.; Bonvin, A. M. J. J.; Vargiu, A. V. Coupling Enhanced Sampling of the Apo-Receptor with Template-Based Ligand Conformers Selection: Performance in Pose Prediction in the D3R Grand Challenge 4. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 149.

(13) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.

(14) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720−725.

(15) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(16) DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Docking Flexible Ligands to Macromolecular Receptors by Molecular Shape. *J. Med. Chem.* **1986**, *29*, 2149−2153.

(17) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, 935−949.

(18) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A Novel Method for the Shape-Directed Rapid Docking of Ligands to Protein Active Sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289−307.

(19) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74−82.

(20) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.

(21) Chaput, L.; Mouawad, L. Efficient Conformational Sampling and Weak Scoring in Docking Programs? Strategy of the Wisdom of Crowds. *Aust. J. Chem.* **2017**, *9*, 37.

(22) Chaput, L.; Martinez-Sanz, J.; Saettel, N.; Mouawad, L. Benchmark of Four Popular Virtual Screening Programs: Construction of the Active/Decoy Dataset Remains a Major Determinant of Measured Performance. *Aust. J. Chem.* **2016**, *8*, 56.

(23) Wójcikowski, M.; Ballester, P. J.; Siedlecki, P. Performance of Machine-Learning Scoring Functions in Structure-Based Virtual Screening. *Sci. Rep.* **2017**, *7*, 1−10.

(24) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455−1474.

(25) Zhang, W.; Bell, E. W.; Yin, M.; Zhang, Y. EDock: Blind Protein-Ligand Docking by Replica-Exchange Monte Carlo Simulation. *Aust. J. Chem.* **2020**, *12*, 37.

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

(27) Rhodes, D. I.; Peat, T. S.; Vandegraaff, N.; Jeevarajah, D.; Le, G.; Jones, E. D.; Smith, J. A.; Coates, J. A. V.; Winfield, L. J.; Thienthong, N.; Newman, J.; Lucent, D.; Ryan, J. H.; Savage, G. P.; Francis, C. L.; Deadman, J. J. Structural Basis for a New Mechanism of Inhibition of HIV-1 Integrase Identified by Fragment Screening and Structure-Based Design. *Antiviral Chem. Chemother.* **2011**, *21*, 155−168.

(28) Zhu, Z.; Sun, Z.-Y.; Ye, Y.; Voigt, J.; Strickland, C.; Smith, E. M.; Cumming, J.; Wang, L.; Wong, J.; Wang, Y.-S.; Wyss, D. F.; Chen, X.; Kuvelkar, R.; Kennedy, M. E.; Favreau, L.; Parker, E.; McKittrick, B. A.; Stamford, A.; Czarniecki, M.; Greenlee, W.; Hunter, J. C. Discovery of Cyclic Acylguanidines as Highly Potent and Selective β-

Site Amyloid Cleaving Enzyme (BACE) Inhibitors: Part I - Inhibitor Design and Validation. *J. Med. Chem.* **2010**, *53*, 951−965.

(29) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(30) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(31) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(32) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177−6196.

(33) Jain, A. N. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J. Med. Chem.* **2003**, *46*, 499−511.

(34) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(35) Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive Evaluation of Ten Docking Programs on a Diverse Set of Protein-Ligand Complexes: The Prediction Accuracy of Sampling Power and Scoring Power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964−12975.

(36) van Dijk, A. D. J.; de Vries, S. J.; Dominguez, C.; Chen, H.; Zhou, H.-X.; Bonvin, A. M. J. J. Data-Driven Docking: HADDOCK's Adventures in CAPRI. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 232−238.

(37) Honorato, R. V.; Koukos, P. I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlato, M.; Giachetti, A.; Rosato, A.; Bonvin, A. M. J. J. Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **2021**, *8*, 708.

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

(39) Tversky, A. Features of Similarity. *Psychol. Rev.* **1977**, *84*, 327−352.

(40) Horvath, D.; Marcou, G.; Varnek, A. Do Not Hesitate to Use Tversky - And Other Hints for Successful Active Analogue Searches with Feature Count Descriptors. *J. Chem. Inf. Model.* **2013**, *53*, 1543−1562.

(41) O'Hagan, S.; Kell, D. B. MetMaxStruct: A Tversky-Similarity-Based Strategy for Analysing the (Sub)Structural Similarities of Drugs and Endogenous Metabolites. *Front. Pharmacol.* **2016**, *7*, 266.

(42) Landrum, G.; Tosco, P.; Kelley, B.; sriniker; gedeck; NadineSchneider; Vianello, R.; Ric; Dalke, A.; Cole, B.; AlexanderSavelyev; Swain, M.; Turk, S.; N, D.; Vaucher, A.; Kawashima, E.; Wójcikowski, M.; Probst, D.; guillaume godin; Cosgrove, D.; Pahl, A.; JP; s; strets123; JLVarjo; O'Boyle, N.; Fuller, P.; Jensen, J. H.; Sforna, G.; DoliathGavid. Rdkit/Rdkit: 2020_03_1 (Q1 2020) Release. *Zenodo* March 2020, DOI: 10.5281/zenodo.3732262.

(43) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.

(44) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for Generation of Unique SMILES Notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97−101.

(45) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562−2574.

(46) Wang, S.; Witek, J.; Landrum, G. A.; Riniker, S. Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *J. Chem. Inf. Model.* **2020**, *60*, 2044−2058.

(47) Schüttelkopf, A. W.; van Aalten, D. M. F. PRODRG : A Tool for High-Throughput Crystallography of Protein−Ligand Complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1355−1363.

(48) Nilges, M. A calculation strategy for the structure determination of symmetric dimers by 1H NMR. *Proteins* **1993**, *17*, 297−309.

(49) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(50) Fernández-Recio, J.; Totrov, M.; Abagyan, R. Identification of Protein-Protein Interaction Sites from Docking Energy Landscapes. *J. Mol. Biol.* **2004**, *335*, 843−865.

(51) McLachlan, A. D. Rapid Comparison of Protein Structures. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1982**, *38*, 871−873.

(52) Morin, A.; Eisenbraun, B.; Key, J.; Sanschagrin, P. C.; Timony, M. A.; Ottaviano, M.; Sliz, P. Collaboration Gets the Most out of Software. *Elife* **2013**, *2*, e01456.

(53) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *Aust. J. Chem.* **2011**, *3*, 1−14.

(54) Haigh, J. A.; Pickup, B. T.; Grant, J. A.; Nicholls, A. Small Molecule Shape-Fingerprints. *J. Chem. Inf. Model.* **2005**, *45*, 673−684.

(55) Boström, J.; Berggren, K.; Elebring, T.; Greasley, P. J.; Wilstermann, M. Scaffold Hopping, Synthesis and Structure-Activity Relationships of 5,6-Diaryl-Pyrazine-2-Amide Derivatives: A Novel Series of CB1 Receptor Antagonists. *Bioorg. Med. Chem.* **2007**, *15*, 4077−4084.

(56) Shape Toolkit 2020.2.2 OpenEye Scientific Software, Santa Fe, NM. Https://Www.Eyesopen.Com.

(57) Meyer, P. A.; Socias, S.; Key, J.; Ransey, E.; Tjon, E. C.; Buschiazzo, A.; Lei, M.; Botka, C.; Withrow, J.; Neau, D.; Rajashankar, K.; Anderson, K. S.; Baxter, R. H.; Blacklow, S. C.; Boggon, T. J.; Bonvin, A. M. J. J.; Borek, D.; Brett, T. J.; Caflisch, A.; Chang, C.-I.; Chazin, W. J.; Corbett, K. D.; Cosgrove, M. S.; Crosson, S.; Dhe-Paganon, S.; Di Cera, E.; Drennan, C. L.; Eck, M. J.; Eichman, B. F.; Fan, Q. R.; Ferré-D'Amaré, A. R.; Christopher Fromme, J.; Garcia, K. C.; Gaudet, R.; Gong, P.; Harrison, S. C.; Heldwein, E. E.; Jia, Z.; Keenan, R. J.; Kruse, A. C.; Kvansakul, M.; McLellan, J. S.; Modis, Y.; Nam, Y.; Otwinowski, Z.; Pai, E. F.; Pereira, P. J. B.; Petosa, C.; Raman, C. S.; Rapoport, T. A.; Roll-Mecak, A.; Rosen, M. K.; Rudenko, G.; Schlessinger, J.; Schwartz, T. U.; Shamoo, Y.; Sondermann, H.; Tao, Y. J.; Tolia, N. H.; Tsodikov, O. V.; Westover, K. D.; Wu, H.; Foster, I.; Fraser, J. S.; Maia, F. R. N. C.; Gonen, T.; Kirchhausen, T.; Diederichs, K.; Crosas, M.; Sliz, P. Data Publication with the Structural Biology Data Grid Supports Live Analysis. *Nat. Commun.* **2016**, *7*, 10882.