Blinded predictions of binding modes and energies of HSP90-ligands for the 2015 D3R grand challenge

Blinded predictions of binding modes and energies of HSP90-α ligands for the 2015 D3R Grand Challenge

Antonia S J S Mey[†$], Jordi Juárez-Jiménez[†$], Alexis Hennessy[†], and Julien Michel[†*]

[†]EaStCHEM School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, United Kingdom

[$] These authors have contributed equally to this work

* Corresponding author. E-mail: mail@julienmichel.net

Abstract

In the framework of the 2015 D3R inaugural grand challenge, blind binding pose and affinity predictions were performed for a set of 180 ligands of the Heat Shock Protein HSP 90-α protein, a relevant cancer target. Spectral clustering was used to rapidly identify alternative binding site conformations in publicly available crystallographic HSP90-alpha structures. Subsequently, multiple docking and scoring protocols employing the software *Autodock Vina* and *rDock* were applied to predict binding modes and rank order ligands. Alchemical free energy calculations were performed with the software *FESetup* and *Sire*/*OpenMM* to predict binding affinities for three congeneric series subsets. Some of the protocols used here were ranked among the top submissions according to most of the evaluation metrics. Docking performance was excellent, but the scoring results were disappointing. A critical assessment of the results is reported, as well as suggestions for future similar competitions.

1. Introduction

Since the advent of computer-aided drug design (CADD) in the early 1980s a major goal of the field has been the quantitative prediction of protein-ligand interactions. Though CADD is now a major technique deeply entrenched in drug discovery [1–9], molecular modelling protocols that reliably and quantitatively predict precise biophysical measurements remain sought after. Obstacles are well understood and include limits in the accuracy of potential energy functions [10–13], and accounting of possible conformational changes in protein and ligands upon complexation [14–19]. While numerous methodologies have been proposed to address these challenges, initial evaluation typically relies on retrospective analysis of their performance. Though such practice is necessary to initiate methodological work, it is insufficient to validate techniques for application to drug discovery. The chief objection is that it is difficult to guarantee removal of biases in the selection of optimal and transferable protocol parameters. A more realistic assessment of the performance of molecular modelling protocols should include true predictions that are subsequently corroborated by follow-up experimental measurements. Yet a major difficulty that stands in the way of this scenario is that most academic labs that contribute innovative CADD methodologies are ill-equipped to follow-up predictions with adequate experimental measurements. To address this obstacle, the 2015 D3R's (Drug Design Data Resource) inaugural grand challenge was setup by a consortium of academics and pharmaceutical companies to assess the state of the art of molecular modelling protocols with blinded predictions of binding poses and binding affinities of drug-like small molecules on previously undisclosed relevant datasets contributed by the pharmaceutical industry.

Structured as a two-stage contest, the first D3R grand challenge aimed to put different computational approaches to the test to predict binding modes and binding affinities. This report is concerned with the competition that focussed on the Heat shock protein HSP 90-alpha (HSP90-α) dataset. Participants were provided with a library of structures of putative ligands for the ATP binding site of this cancer target [20–22]. The objective of the first stage of the competition was to perform a blinded pose prediction on a small subset of six compounds of this library, and an affinity-based rank ordering for the full library. Participants were also given the opportunity to submit binding affinity predictions for subsets of this dataset. This report contrasts the performance of several protocols used by our group for docking and scoring this dataset and concludes with

suggestions for future similar competitions. All datasets discussed are available for download from the webpage of the Drug Design Data Resource consortium [23]. The supplementary information gives details on how to compare the presented work here to other competitors results.

2. Theory and Methods

2.1 Datasets

The ligand library provided by the organizers consisted of 180 ligands based on different scaffolds, all provided as SMILES strings. Subsets of the 180 ligands were defined by the organizers as set1, 2 and 3 containing five benzophenone derivatives, four amino pyrimidine derivatives and ten molecules built from benzimidazolone scaffolds respectively. A 2D structure of each of the subsets can be found in figure SI1. For HSP90-α four crystal structures were provided (PDBID: 2JCC, 2XDX, 4YKR and 4YKY), capturing two conformations of the ATP binding site of HSP90-α, although participants were encouraged to further evaluate the receptors conformational variability.

Once the challenge was concluded the experimental affinity data was released by the organizers, consisting of IC_{50} data obtained using a FRET based assay as described by Huth et al. [24]. To compare computational prediction with experimental results the binding free energy of a ligand L2 relative to a ligand L1 was calculated with equation 1:

$$\Delta\Delta G_{\boldsymbol{L1\rightarrow L2}} = k_B T \ln \frac{\mathrm{IC50}_{L2}}{\mathrm{IC50}_{L1}} \quad , \qquad (1)$$

where k_B is the Boltzmann constant, and T the temperature.

2.2 Selection of protein structures for docking and scoring

Initial analysis of available X-ray diffracted crystal structures of HSP90-α indicated that significant ligand-induced conformational changes in the ATP binding site were possible [25,26]. Given the relatively short period available for contestants to submit predictions it was deemed unpractical to initiate binding site conformational dynamics studies via enhanced sampling methods such as accelerated Molecular Dynamics [27,28] (aMD) or Metadynamics [29]. Yet the literature evidence was sufficient to suggest that simplistic treatments of protein flexibility for this binding site (i.e: considering pre-computed rotamer libraries for flexible sidechains [30], using reduced

van der Waals radii on the receptor atoms [31,32], or the refinement of docking poses by means of Molecular Dynamics (MD) simulations [33,34]) would fail to capture potential significant conformational rearrangements. The alternative pursued here was to analyse collections of previously solved X-ray structures to construct a representative ensemble of binding site conformations suitable for docking calculations [26].

Given the large number of available HSP90-α structures in public databases an algorithm was needed to classify structures by representative patterns. This was carried out with a spectral clustering method [35,36]. Specifically, 195 HSP90-α crystal structures from the PDB were identified with a resolution of at least 2 Å. A root mean square deviation (RMSD) matrix **M** between these 195 structures based on C-α atoms was then computed, see figure 1. The RMSD matrix then served as a distance matrix for spectral clustering. The idea behind spectral clustering is to achieve a dimensionality reduction of a large dataset and grouping (clustering) of similar structures in the process. The goal here was to find representative structures from the resulting clusters that would then be used for docking calculations. The RMSD matrix was used to formulate a Gaussian diffusion kernel which gives an artificial diffusion distance between the different crystal structures. The diffusion probability between two crystal structures is given by:

$$\boldsymbol{K_{ij}} = \mathbf{exp}\left(-\frac{\left|\boldsymbol{M_{ij}^2}\right|}{\boldsymbol{2\varepsilon}}\right), \qquad (2)$$

where M_{ij} are the entries of the RMSD matrix, ε is a cutoff, and **K** is a matrix holding information about the diffusion probabilities. The cutoff ε was chosen to be 0.25 $Å^2$. **K** can be normalized such that it is a stochastic matrix **T**, whose eigenvalues are the objects of interest for the spectral clustering. The dominant eigenvalues of **T** indicate how many clusters the dataset contains. Subsequently the so called Perron Cluster Cluster Analysis (PCCA+) algorithm [37], which uses information in the eigenvector structure to assign invariant clusters, was used to assign each of the crystal structure to one of the clusters. The eigenvalues of the stochastic matrix **T** and the PCCA+ clusters, were computed using the software package *pyEMMA* [38].

2.3 Docking

Docking protocols

PDB structures retained for further analyses were prepared for docking calculations. Addition of hydrogen atoms and removal of co-crystallized ligands was conducted using the software *fconv* [39]. Protonation states for titratable residues in the receptor were assigned using *PROPKA* [40,41] at a pH value of 7, while the pKA calculator plugin from Marvin 15.3.30.0 was used to determine the most abundant species of ligands at pH 7.

Two different docking methodologies were employed. The first protocol, denoted 'unbiased', was carried out using the software *Autodock Vina.* A cube of 20 Å of side was centred on Asp 93 to define the docking search space, the exhaustiveness level was set up at 50 and default options were used for the remaining parameters. *AutodockTools* was used to prepare the ligands and to assign Gasteiger charges [42]. The second protocol, denoted 'guided', was carried out with the software *rDock* [32]. In this case, the 'guided' protocol was similar to that reported by Barril and co-workers [43]. In brief, a hydrogen bond donor and hydrogen bond acceptor groups from the ligand were set to a distance of 4.0±0.75 Å from C_γ of Asp 93, which was also the centre of a sphere of radius 24 Å that enclosed the docking cavity. Each compound was subjected to 100 iterations of the genetic algorithm using default scoring parameters.

Additionally, water molecules are known to play a key role in the stabilization of ligands in the HSP90-α binding site [43], therefore two interstitial water molecules that interact with the side chain of Asp 93 were retained in both the 'unbiased' and the 'guided' docking protocols.

Ranking protocols

Docking protocols employed for the pose prediction were also used for ranking of the whole dataset. For clarity protocols are named according to the docking program used (*v* for *Autodock Vina*, *rd* for *rDock*), and the PDB ID of the HSP90-α structure used for the dockings (4w7t, 4l94, 4cwf, 2fwz, 2ccu). This yielded 10 different ranking for the ligands dataset. In addition two further protocols were defined as rd/v-average(avg) and rd/v-best. rd/v-avg is obtained by taking the average score of the top-ranked pose from all five receptor structures for each of the 180 compounds. rd/v-best is obtained by taking the results from the best scoring structure for each of the 180 compounds. Some compounds were unable to satisfy the imposed restraints, and these were arbitrarily assigned a score of 0.0 kcal/mol.

2.4 Alchemical free energy simulations

Theoretical background

Relative alchemical free energy have shown to be a useful tool for the computational prediction of binding affinities in the past [11,44–50]. In this methodology, the relative free energy of binding for two ligands to the same receptor can be described by a thermodynamic cycle. Rather than using molecular dynamics (MD) simulations to compute binding trajectories directly for the two different ligands, an alchemical transformation, 'morphing' L1 to L2 is carried out. The relative free energy of binding between two ligands is then given by:

$$\Delta\Delta G_{bind}(L1, L2) = \Delta G_{solvated}(L1, L2) - \Delta G_{bound}(PL1, PL2). \quad (3)$$

This means that in order to compute $\Delta\Delta G_{bind}(L1, L2)$ two independent sets of simulations have to be carried out. One simulation for L1 solvated in water and one simulation for L1 bound to the protein. In both sets of simulations L1 is transformed to L2 over a set of $\lambda \in [0,1]$ windows, where λ =0 corresponds to L1 and λ =1 corresponds to L2, either in just solvent or solvated and complex. Any intermediate λ value corresponds to a simulation with a linear interpolation between L1 and L2 in the force field. In this work a single topology alchemical perturbation protocol was employed, using 17 evenly spaced λ-windows. The MBAR estimator as implemented in pymbar was used to recover $\Delta G_{solvated}$ and ΔG_{bound} values from the 17 separate λ simulations [51]. All simulations were carried out with the *Sire*/*OpenMM* (*SOMD*) software revision 2015.0.0. [52].

Alchemical free energy molecular dynamics simulation setup

The relative free energy setup and simulation protocol from the predicted docking poses is laid out in the following. Aside from the set 1, 2, and 3 compounds a small test dataset was initially simulated to validate the proposed simulation protocol for the challenge. The compounds of the test dataset were taken from Bruncko et al. [53]. The simulation setup pipeline was identical in all cases although a different representative structure of the receptor was used for each set. PDB structures 2YK2 [54] and 4L90 were used for sets 1 and 3 respectively based on trial docking calculations, whereas PDB structures 2XDX [55], and 3OWD [53] were used for the test set and set 2

respectively due to the perceived similarity between the co-crystallized and query compounds.

The *FESetup* [56] software was used to set up relative free energy calculations Ligands were automatically parameterized using the GAFF force field [57] and assigned AM1-BCC [58,59] charges. More details for the setup parameters are provided in the SI.

Alchemical free energy molecular dynamics simulation protocol

Single-topology alchemical free energy calculations were carried out using the *SOMD* simulation engine with the following protocol. The CUDA OpenMM [60] platform was used to execute the simulations on a cluster of GTX980 GPUs. Both the complexes consisting of protein and ligand as well as the ligand in water were run in order to be able to extract a relative binding free energy from these simulations using eq. 3. Each simulation was repeated twice to assess reproducibility of the computed relative binding free energies. During all simulations a hydrogen mass repartition [61] scheme was employed to allow for an integration time step of 4 fs and each λ-simulation was run for 8 ns using a Leap-Frog Verlet algorithm. The initial velocities were drawn randomly according to a Maxwell-Boltzmann distribution at 298 K. A mean temperature of 298 K was achieved using an Andersen thermostat with a collision frequency of 10 ps^{-1}, whereas pressure was maintained at 1 atm using a Monte Carlo barostat as implemented in OpenMM [62,63] with an update frequency of 25 MD steps. The simulation box was treated with periodic boundary conditions and non-bonded interactions were evaluated by using a 10 Å atom-based Barker-Watts reaction field cut off scheme, with the medium dielectric constant set to $\varepsilon_{solvent} = 78.3$. Simulation input files are available for download, with more information given in the SI.

Alchemical free energy molecular dynamics simulation analysis

As a first step in conducting an alchemical free energy calculation, a series of perturbations between ligands needs to be defined. An example of such a perturbation network is shown in figure 5A and can be interpreted as a directed graph, where the directed edges indicate the direction of simulations carried out. Each edge represents two sets of calculations; the changing of the ligand in solvent and the changing of the ligand bound to the protein. The value along the edge (in kcal/mol) is the computed $\Delta\Delta G_{bind}(L1, L2)$, based on the intermediate simulations, from equation 3. In order to obtain relative free energies with respect to a given reference compound, e.g**. 9e** from

figure 5A, different approaches can be taken [64] In this study a straight forward path averaging was used. Each of the two 'edge simulations' was repeated twice and an average of both relative binding free energies ΔΔG between the two compounds was computed. This resulted in an 'averaged' perturbation network as shown in Fig. 5A. Relative binding free energies of each individual run were compared before averaging in order to make sure that both runs resulted in a similar ΔΔG. The averaged perturbation network of relative binding free energies was then analysed using the python network package networkX [65]. Two simple checks were carried to test the robustness of the network before proceeding. According to Kirchhoff's law, the sum of flow into a node and out of a node must be zero and simple cycles consisting of three nodes should therefore be zero. In practice cycles with a cycle closure error less than 0.3 kcal/mol were considered a sign of robust convergence. Since the overall goal was to compute a relative binding free energy of all compounds with respect to a particular target compound, networkX was used to compute all possible simple paths, i.e. non-self crossing paths, between a target compound (e.g. node **9e** of figure 5A) and an initial compound (e.g. **9b** of figure 5A). The resulting free energy paths were averaged and their standard deviations used for the error analysis. Doing an analysis in this way neglects information on the fact that different paths will have different lengths and should give different contributions to the overall relative free energy between **9b** and **9e**. However, it is unclear based on the assumption that all edges along the path carry a certain uncertainty, which may vary depending on each of the edge, what the most robust way of estimating relative free energies is. All relative free energies shown in the results section have been computed based on the path averages and their standard deviations based on the perturbation networks of each dataset.

Once relative binding free energies have been computed both for the experimental data and from the computational perturbation network analysis, their correlation has to be assesed. Typical measures for correlation are the Pearson correlation coefficient R and Kendall τ as well as the mean unsigned error (MUE) which gives information about the quality of the free energy estimate. In order to test the robustness of the estimated R, τ, and MUE values an error analysis was undertaken in order to get a lower and upper bound for their estimates [46,66] . Each computational relative binding free energy estimate has a measured value and an error associated with it. For each computational

data point an artificial normal distribution was constructed, with a mean of the free energy estimated and the standard deviation given by the error of the estimate. Then for each computational data point 1000 random samples were drawn from the constructed normal distribution and correlated with the experimental result. This then gives 1000 data points for R, τ and MUE, which are in turn a distribution from which the mean and the standard deviation can be determined and represent the values that are reported in the results section.

3. Results and Discussion

3.1 Receptor structure selection via spectral clustering

During the first stage of the D3R challenge the objective was to obtain binding modes for six compounds, defined by the organizers, belonging to the three different chemical families: two benzimidazol-2-ones (**40** and **44**) two hydroxibenzophenones (**160** and **175**) and two aminopyrimidines (**73** and **179**). Given the known plasticity of the ATP binding site of HSP90-α, it was deemed necessary to dock each compound against multiple structures. Searches in the PDB returned 195 crystal structures with resolution < 2 Å. Although performing docking calculations on all such structures was feasible, this was not judged suitable given short-comings in scoring, and the increased human time needed for analysis of the results. Instead spectral clustering using a C-α RMSD metric was used to identify groups of highly related structures (see methods). The results indicated five dominant eigenvalues; hence the choice of five different binding site RMSD families seemed appropriate. After visual inspection, however, the first three clusters were unified, since their structural difference were not very large, and both cluster 1 and cluster 2 initially only held a small number of crystal structures, all stemming from the same crystallographic study [53]. As shown in figure 1, the final clustering groups the 195 structures into three large families differing in the arrangement of the region encompassed between residues 104 and 111: The first cluster (c1) groups structures where these residues form a loop that occludes the binding site. The second cluster (c2) represents a conformation where the same region forms an α-helix. The remaining cluster (c3) gathers structures where residues 104 to 111 are arranged in a configuration that retained features from both α-helix and disordered loop and where the ATP binding site is only partially occluded. It is worth noting that,

although the clustering procedure was solely based on the analysis of the C-α RMSD matrix, and no constraints were imposed regarding the final number of clusters, the resulting three clusters agree well with the three major conformational states of HSP90-α (open, closed and helical) previously reported in the literature [25,26].

Representative structures for each cluster where then selected after visual inspection. Care was taken to represent sidechain rearrangements not captured by the C-α RMSD metric. Noticeably, only residues LYS58 and ASP54 were identified in two significantly different conformations in clusters c1 and c2, but not among structures in cluster c3. Consequently, structures ID:2ccu [67] and ID:4l94 [68] were used to represent c1, 2fwz [69] and 4cwf [70] to represent c2 and 4w7t [71] was the representative structure of c3. Therefore, solely by means of binding site RMSD spectral clustering and a slight refinement of the results, it was possible to capture major differences in the receptor's binding site and to represent the whole available crystallographic data in a reduced data set.

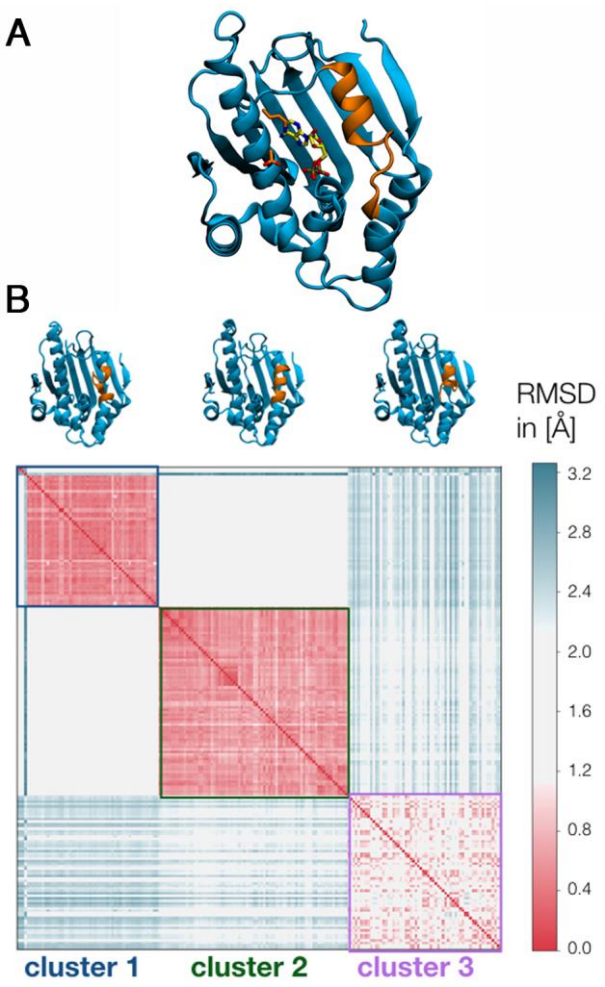

Figure 1. A) Structure of the complex between HSP90- α and ADP (PDB ID: 1byq[72]). Regions of highest variability on the active site among HSP90-α crystals are highlighted in orange, including residues ASP54 and LYS58. **B)** RMSD matrix of 195 crystal structures of HSP90-α, with entries ordered according to the spectral clusters. The conformational difference between the three chosen clusters is highlighted in the top structures in orange.

3.2 Pose prediction results

A total of thirteen different predictions were generated using the protocols described in the methods section. Each prediction included up to five ranked poses. In addition to the protocols used to rank compounds, solely for the pose prediction part of the challenge, two protocols involved manual selection of the top 5 poses by visual inspection of the top docked poses across all tested structures (denoted *–visual* hereafter). In this protocol, the criteria followed to rank docking poses was based on the shape complementarity between each docking pose and the binding site, the chemical complementarity between ligand substituents and binding site lining sidechains (specially the number of hydrogen bonds established and the distances predicted by the docking program) and the overall resemblance of the docking pose with the crystallographic binding mode of related compounds. The quality of each prediction was evaluated by the D3R organizers using two metrics: the mean RMSD of the top ranked pose to the crystal structure, and the mean value of the lowest RMSD among the five poses belonging to each compound. A comparison of the performance of individual predictions is shown in figure 2.

In general, among protocols with no *a posteriori* human driven refinement, protocols based on r*Dock* outperform those based on *Vina* when results are ranked according to the RMSD of the top ranked pose: four out of six *rDock* protocols achieved RMSD values below 2.5 Å in this metric, while in the case of *Vina* protocols all scored above 3.0 Å. The same trend is observed when results are ranked according to the lowest RMSD obtained within the top 5 ranked poses, although differences become less evident in this case: all *rDock* protocols scored below 2.5 Å while four out of six *Vina* protocols were below that threshold. Due to the small number of predictions, which does not allow for a significant statistical analysis, these results do not report on the quality of any of the codes or scoring functions. Instead, it is assumed that the better performance of protocols employing *rDock* highlights that the use of pharmacophoric

restraints enables a more accurate sampling of region of interest of the conformational space. Indeed, this explains why reduced differences between *rDock* and *Vina* protocols are observed when predictions are ranked according to the lowest RMSD of any top 5 ranked poses. The correct pose is also sampled by the protocols based on *Vina* but it is wrongly ranked. By contrast, the use of pharmacophoric restraints avoids exploring solutions away from the region of interest, and allows a more exhaustive exploration of the local minima, produced more refined poses that score better.

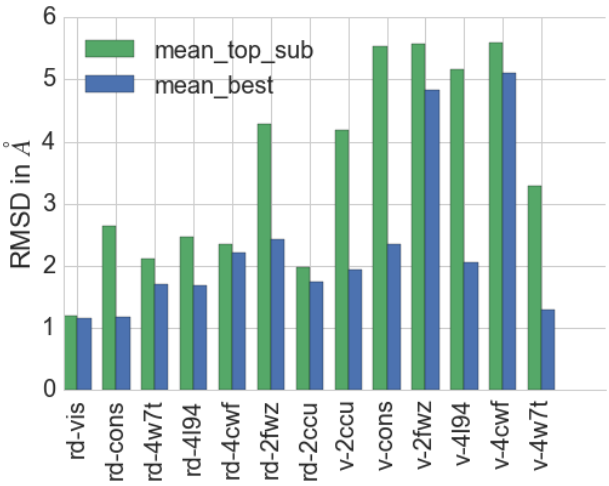

Figure 2. Bar plot of all different Docking protocols and the mean RMSD of the best submitted pose (green) and beast mean RMSD from all submissions (blue).

The best overall performance was obtained by the *rDock-visual* protocol, were a prediction was produced after visual inspection of top scored poses across multiple structures (unfortunately the equivalent *Vina* protocol was not ranked by the organizers due to a formatting error for our submission. Similar results were expected due to the high resemblance of the predicted poses with the *rDock-visual* and *Vina-visual* protocols). Precisely, the mean RMSD of the first pose for the *rDock-visual* protocol was 1.2 Å (ranking 4th out of 47 submissions in the challenge). As shown in figure 4 this protocol yielded predictions with an RMSD below 1.0 Å for four compounds (**73**, **164**, **40** and **179**), slightly above for **175** and only for **44** the RMSD peaked above 3.0 Å. Even in the case of **44**, the main features of the binding mode were retained and the high RMSD value is due to a different orientation of the pyridine sulphonamide moiety. It is worth emphasing that this human driven protocol outperformed automated docking predictions on single structures, as well as automated consensus protocols.

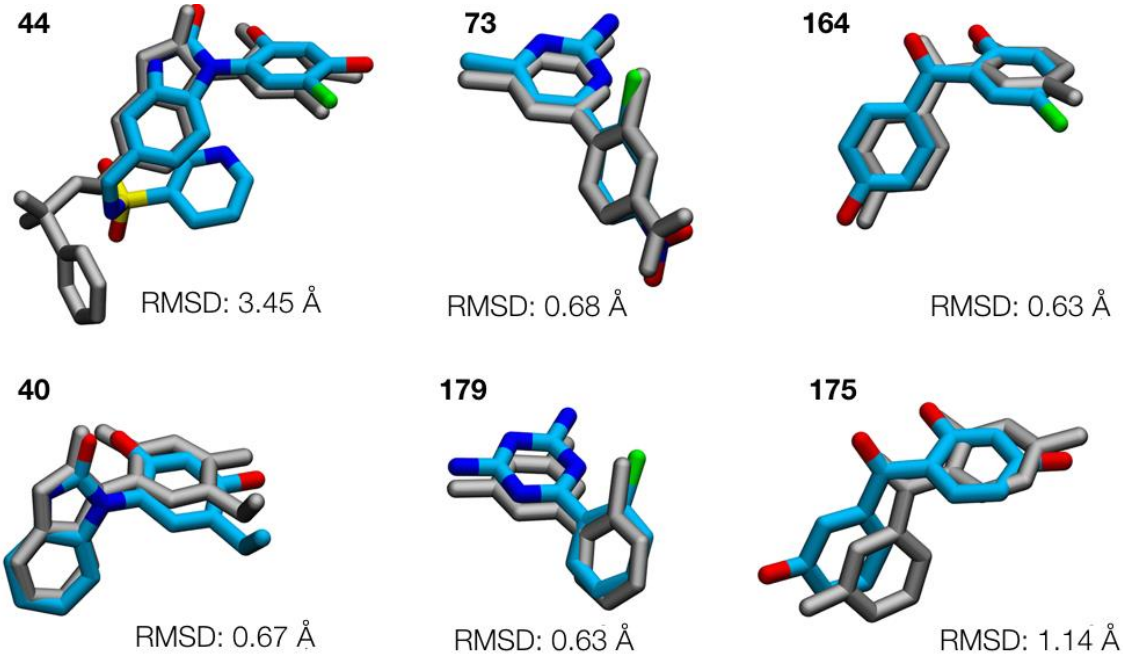

Figure 3. Superposition and RMSD performance of the *rd-visual* protocol prediction of query compounds with respect to their revealed crystal structure.

Scoring of compounds

Similarly to the evaluation of the pose prediction stage, performance of the different scoring protocols was assessed using three different metrics: the area under curve (AUC) of a receiver operator curve (ROC), the Pearson correlation coefficient (R value) and Kendall τ. The first metric reports on the ability of a protocol to distinguish active

(defined as those compounds with an $IC_{50} < 1\ \mu$ M) from inactive compounds, while the other two metrics assess the correlation of the rankings between experimental and docking results for the 180 compounds. To assess statistical treatment, the data was bootstrapped and for each docking dataset of the 180 ranked compounds, a 'new' set of 180 compounds was drawn at random with replacement from the existing dataset and ordered according to their rank obtained from the docking program. Figure 4 summarises the overall results of the scoring of the 180 ligands using the different protocols. From figure 4A it can be deduced, that three of the *rDock* protocols (*rd-best*, *rd-avg*, and *rd-4w7t*) performed significantly better than the other *rDock* and *Autodock Vina* protocols for the AUC measure. All *Autodock Vina* protocols perform similarly and the best average AUC value obtained is 0.72. The overall correlation between the rankings of compounds is rather poor for both the R value and the Kendall τ measure. Figure 4B and 4C show that only the *rDock* protocol *rd-4l94* performs significantly worse. This poorer performance may be due to the different arrangement of the side chains of LYS58 and ASP54 in this crystal structure. In comparison to other participants of the challenge, both correlation scores lead to a top 3 position rank. However, there is no statistically significant difference with other top scoring protocols, largely due to an overall poor correlation.

Interestingly, nine compounds were excluded from the 'guided' docking protocol since they could not satisfy the imposed pharmacophoric restraints. Among these, compounds **51**, **122**, **180**, and **182** were later revealed to be inactive ($IC_{50} >> 1$ μM); two other (**145** and **165**) were found to be ~100 fold less active than the most potent inhibitor in the dataset (**44**, $IC_{50} = 0.005$ μM). The remaining three compounds **128**, **161** and **174** belong to the top 10% of most potent inhibitors. Therefore, although AUC measure was significantly better for the 'guided' docking protocols, this family of protocols failed to identify three interesting compounds that arguably deviated significantly from other scaffolds in the dataset (Figure SI4).

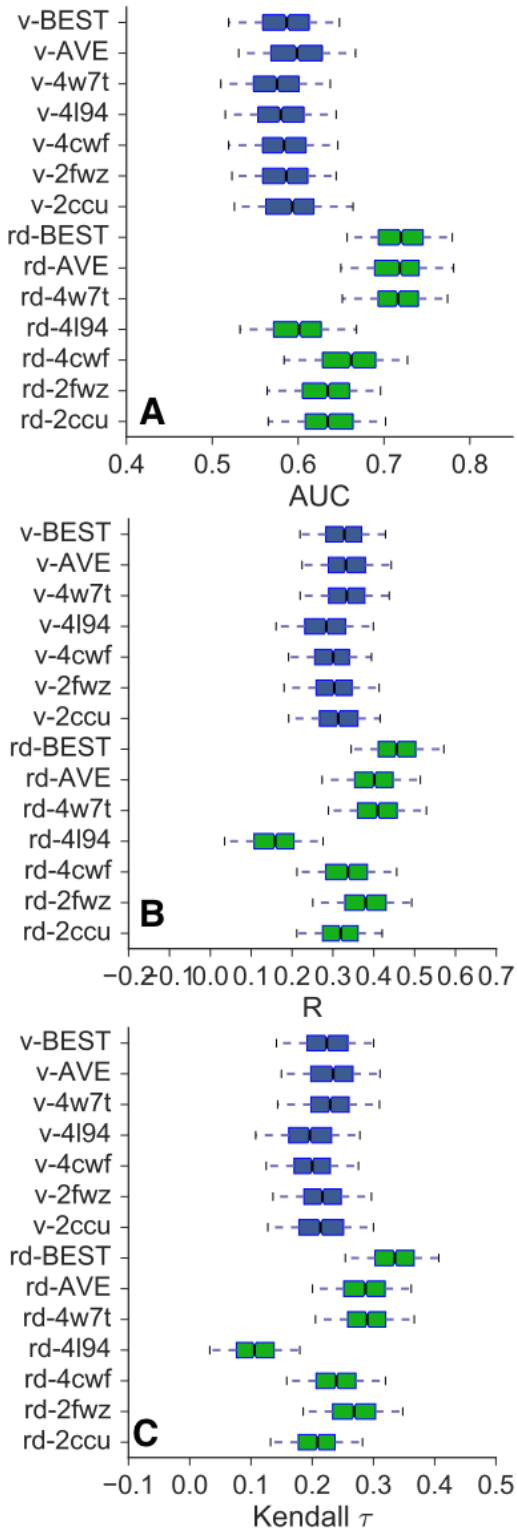

Figure 4. Scoring Results. Autodock protocols are shown in blue and rDock protocols in green. Mean and 95% confidence intervals are shown. **A)** Area under Curve of 1000 bootstrapped ROC curves. **B)** R metric. **C)** Kendall τ metric.

3.2 Alchemical relative free energy calculations

Test data set – amino pyrimidine scaffolds

The robustness of the proposed simulation and analysis protocol was first tested on a small dataset of amino pyrimidine derivatives for which both crystal structure data, as well as IC_{50} data from a FRET assay were available [53]. The results of the test data set are summarized in figure 5. Figure 5A shows the relative alchemical calculations carried out, needed for the computation of relative binding free energies with respect to a reference compound, here compound **9e**. Errors on the computed $\Delta\Delta G$ values are omitted for clarity in figure 5A, but can be found in table SI1. Figure 5B shows the results of the experimentally obtained $\Delta\Delta G_{\mathrm{exp}}$ with respect to the computational results. Trends are correctly captured apart from one compound (**9b**). The correlation between the computational and experimental estimate are analysed computing the Pearson rank correlation coefficient R, as well as Kendall τ. A relatively good correlation of R=0.74±0.03 and τ = 0.79±0.04, is observed. A third measure that was used is the mean unsigned error (MUE) which after bootstrapping had a clear maximum in its distribution given by MUE = 1.2±0.1 kcal/mol. Since the test dataset gave reasonable confidence in the implemented protocol the same protocol and analysis method were employed for blind predictions of the D3R datasets.

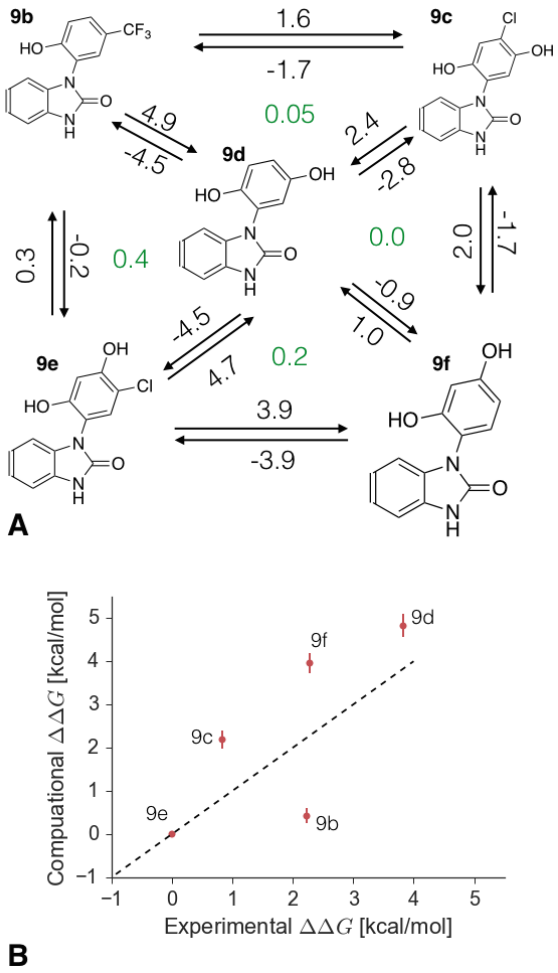

Figure 5. A) Perturbation network of the test dataset. Directional arrows indicated simulations carried out with computed relative free energies in [kcal/mol]. Green numbers indicate average cycle closure errors. **B)** Experimental versus computational relative binding free energies with respect to compound 9e. Dashed line represents optimal agreement between experiments and computations.

Set 2 – Amino pyrimidine scaffolds

Results of the dataset defined as set 2 by the D3R competition are shown in Figure 6. As before, the relative free energy values shown are obtained from the average network taken from two independent runs and a more detailed table summarising all relative binding free energies including error estimates can be found in table SI2 and SI3. Although crystallographic water molecules found in PDB ID: 2xdx were retained during alchemical free energy simulations, the results submitted to the D3R challenge included

data in which compound **100** was set up with only two interacting crystal water molecules in the binding site, due to a clash with the nitrile group of the compound with a third crystal water molecule as shown in figure 6C and figure 6D. The clashing water was carelessly deleted during the challenge phase when preparing the setup. However, this resulted in a very poor prediction for the relative binding free energies of all compounds in set 2. This is illustrated in figure 6B, where the red bars represent the experimental results and the blue bars computed relative binding free energies based on calculations, where simulations starting from compound **100** were simulated with two waters in the binding pocket, while simulations starting from all other compounds were simulated with three waters in the binding site. A scatter plot of the same data can be found in figure SI2. The resulting sampled R value of R = -0.6±0.2 shows a negative correlation and a MUE = 3.3±0.2 kcal/mol, clearly indicating that the computational prediction does not capture experimental trends. Organizers also computed the RMS error as a measure of performance and found this to be 2.0 kcal/mol, ranking this result 10^{th} out of 18 submissions. Upon further inspection, it was evident that no water molecule diffuses into the binding pocket during the 8 ns-per λ simulations starting from compound **100**. Effectively, this meant that compounds morphing from compound **100** structure were lacking a potentially essential water molecule for the ligand stabilisation in the binding pocket. Simulations were repeated after the challenge deadline with three crystal waters in the binding pocket with compound **100** bound, as seen in figure 6D. As a result, the trend in the relative binding free energy for all compounds improves significantly, with an R value of R = 0.5±0.2 and a MUE = 2.3±0.2 kcal/mol. Nevertheless, this is still a rather poor result for the overall accuracy of the prediction and, even though the improved results would rank the entry based on the R value as a top entry, it is unclear that it would classify as statistically significantly better than other submissions. However, these results highlight that influence of crystal waters in the binding pocket cannot be underestimated. While this has been shown elsewhere [73], and methodologies to predict water content and changes in water network energetics have been developed [74,75], it remains challenging to anticipate *a priori* whether changes in water structure will play a role for a given alchemical free energy calculation.

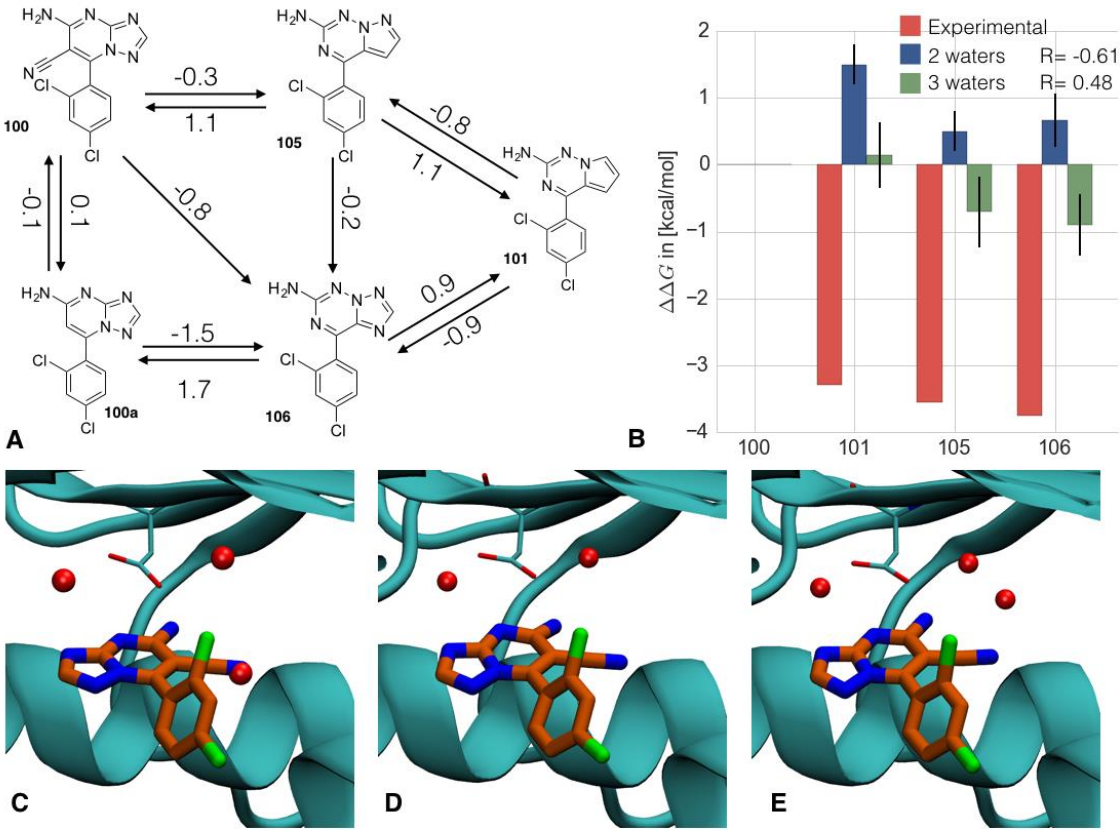

Figure 6. A) Alchemical perturbation network for set 2 with relative binding free energies indicated in kcal/mol. Compounds **100a** is an artificially introduced intermediate. **B)** results of the computed relative binding free energies with respect to compound **100**, with experimental results in red, and simulation results for **100** setup with 2 binding pocket waters in blue and 3 binding pocket waters in green. Errors are shown in terms of a standard deviation and R values for the two crystal and three crystal water setup are indicated. **C**) setup with crystal water resulting in clash with nitrile group **D)** binding pocket setup of compound **100** with two crystal waters, by simply removing clashing water. **E)** setup with three waters, by manually moving the clashing crystal water**.**

Set 3 – Benzimidazolone scaffolds

Set 3 consists of 10 compounds and the results of the free energy predictions can be found in figure 7. For the case of two compounds (**28** and **23**) two alternative binding modes, shown in panels B and C of figure 7, were simulated. In each case, the binding mode that gave the lower relative free energy was chosen for the submission. Figure 7A shows the correlation between the experimental and computational predictions. The perturbation network for the set 3 compounds is provided shown in figure SI2 and error

estimates for relative calculations in table SI3. During the challenge phase not all free energy computations converged well with some MBAR error estimates that were much larger than typically seen, hinting at a sampling problem. Furthermore, it was not possible to reliably incorporate compound **60** into the perturbation network, due to large perturbations required from small intermediates as seen in the perturbation network. Also not all cycles gave good estimates on the cycle closure error test: for example including two of the cycles that involve the intermediate **INT01** as seen in the perturbation network shown in figure SI2, with a closure error of >> 0.3 kcal/mol. With this information it was expected that the results would not give an excellent correlation with respect to the experimental values. Nonetheless a correlation of R = 0.43±0.03, τ =0.3±0.04 and a MUE = 1.3±0.04 kcal/mol was achieved. Overall this meant a top three ranking with respect to all 20 submissions for this dataset in terms of R, and a best ranking in terms of root mean square error, which was calculated by the organisers to be 1.43 kcal/mol. However, this result does not hold statistical scrutiny, and other top scoring entries fared similarly well.

Set 1 – Benzophenone scaffolds

Results for set 1 are summarized in figure 8. It is clear that no correlation between the experimental and computational data is observed, with R=-0.44±0.03, MUE=3.78±0.08 kcal/mol, and τ =-0.55±0.09. Therefore, it was rather surprising that, despite these poor results, this submission was ranked in 12th place out 44 submissions, according to the organizer computed RMS error of 2.67 kcal/mol. With the present simulation data it is hard to confirm the exact reason why the implemented alchemical protocol shows such poor performance for set 1. However, two different factors may contribute to this: first being the lack of available crystal structures for this data set and second its unsuitableness for a relative free energy calculation due to moving from a furanyl moiety (compound **80**) to a benzyl substituent (all other compounds). To the best of our knowledge, no crystallographic structure has been solved for any of the compounds in this set, hence pose prediction was obtained purely from docking calculations. Moreover, the high symmetry of the compounds poses an additional difficulty, since the most favourable binding mode will be determined by subtitle differences in the accommodation of very similar hydrophobic moieties that would be hardly captured by docking scoring functions. Therefore, similarly to set 3, this set faced the problem of multiple possible binding poses for each of the compounds. Thus, each compound had

to be simulated using three different binding poses. The full perturbation network for set 1 with 2D representations of the different binding poses is shown in figure SI3. In contrast with the previous set, the approach of choosing the binding pose that will give the overall most negative relative binding free energy does not yield a good correlation with experimental results. This observation is in line with a recent reported from Kaus et. al. [76]. The reason why perturbations from a five membered ring to a six membered ring are difficult is because they can currently only be achieved with the software *SOMD* via intermediates with neither of the rings present, such as structure **INT02**. Since two intermediates are needed, additional uncertainty in the relative free energies is introduced. This is especially relevant for this set since some of the paths to the reference compound include up to five intermediate calculations. Further analysis and possibly additional simulations for the dataset are desirable to establish why the protocol fared poorly. However, without further experimental evidence to validate the correctness of the predicted binding modes it is difficult to isolate errors introduced by the docking step and the scoring step of the competition.

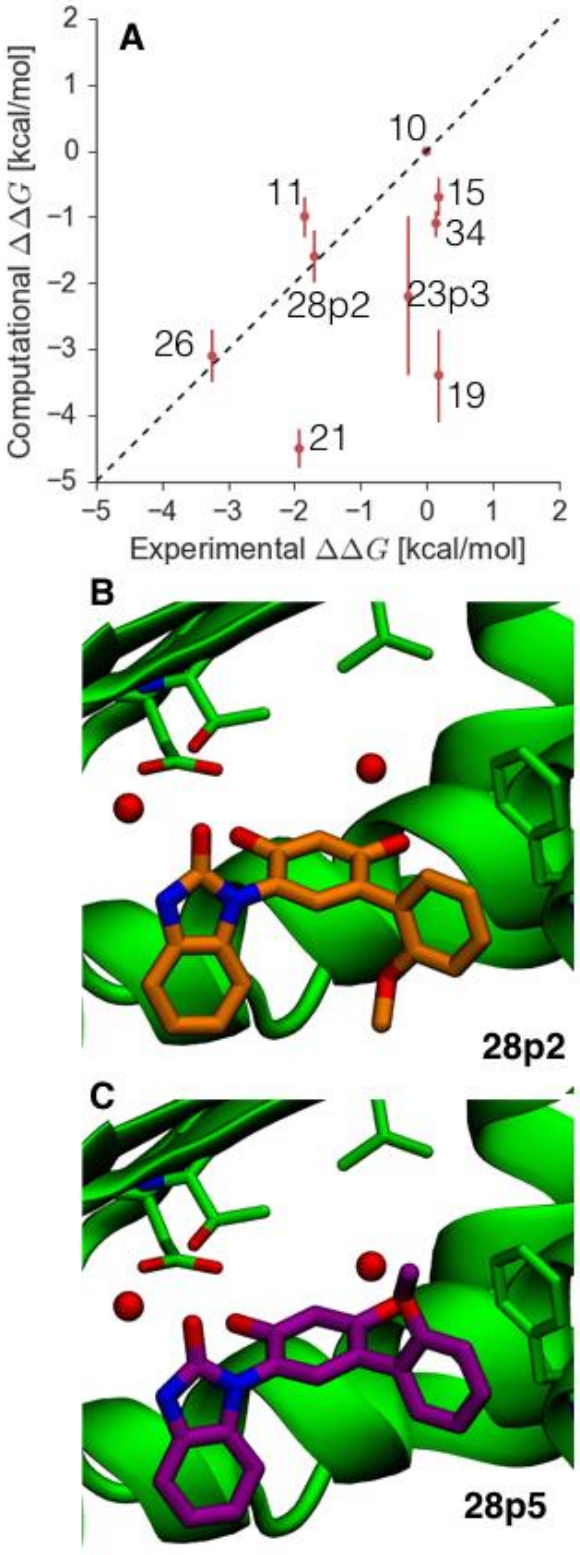

Figure 7. A) Comparison of experimental and computed relative binding free energies for set 3. Compounds numbers are indicated. **B)** Chosen binding pose for submission **C**) additional proposed binding pose which was also simulated, for a detailed perturbation map see figure 2 of the SI.

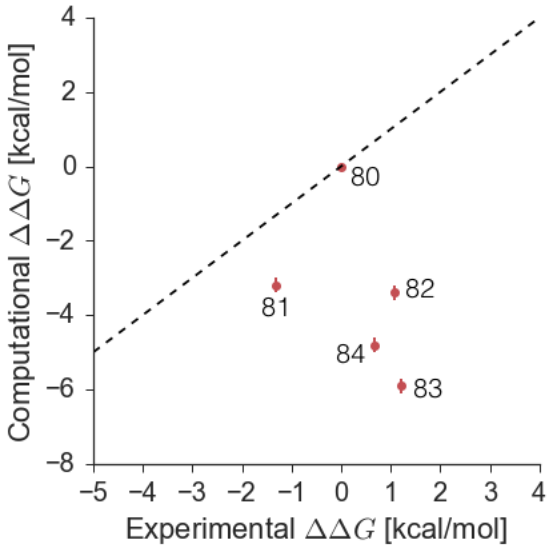

Figure 8. Comparison of experimental relative binding free energies with respect to computational binding free energies for set 1 data. Compounds are indicated.

5. Conclusions

Accounting for receptor flexibility is still one of the unsolved challenges in docking for drug design. Often time tight timelines or limited computational resources prevent undertaking an exhaustive conformational analysis of the receptor. When sufficient experimental data is available, it is possible to obtain valuable insights into protein flexibility by analysis of crystallographic structures. The visual inspection and classification of hundreds of structures, however, is a tedious and error prone task that can be quite time consuming. Here spectral clustering was employed to easily group together crystallographic structures with common features, and to aid in the identification of representative structures that capture most of the relevant conformational states of a protein binding site. Once a set of suitable structures has been identified, the second problem to overcome is to elucidate the binding mode of query compounds. The present results support the use of pharmacophoric restraints to speed up calculations and to yield more accurate predictions. Parsimonious use is wise since ill-chosen restraints will prevent docking of novel interesting scaffolds.

On the scoring side, some of the docking protocols reported here performed well relative to the rest of the competition, but the absolute performance was poor. The result mirrors the experience of many groups and warrants further research into more accurate

scoring methodologies, for instance alchemical free energy calculations. In general, retrospective analyses are commonly used to assess the performance of binding free energy prediction methods. An unexpected result was the low transferability of the performance of the binding free energy prediction protocol from the test set to the actual D3R sets. Such result may highlight the sensitivity of alchemical free energy calculations to factors such as the number of water molecules in the binding site, or deviations from the correct ligand binding mode. This information can be inferred easily when simulations are initiated from an X-ray structure of a protein-ligand complex. When predictions concern novel scaffolds, correct anticipation of the most likely binding mode may be much harder. Consequently, future expansion of the field should not only focus on the development of more accurate methods for free energy calculation, but also deliver new approaches for reliable pose predictions that account for water networks and protein flexibility.

The D3R grand challenge, together with the CSAR benchmark exercise [77–79], is one of the first attempts to promote blind predictions on datasets donated by the pharmaceutical industry. Since this situation more closely resemble the tasks faced by molecular modellers, blinded predictions should ideally be used more routinely to evaluate the performance of novel docking or scoring methodologies. However, it can be difficult to learn from a failure when sources of error are multiple and intricate. For instance, in this particular challenge it was difficult to determine whether failures in scoring the three compounds subsets with free energy methods was due to errors in binding mode predictions, or force fields employed. Another concern is the relatively small number of compounds present in the three subsets that were evaluated with free energy methods. Ideally datasets of at least 10-20 compounds spanning a few orders of magnitude in measured dissociation constants would be used as this would allow robust statistical analysis of the relevance of a prediction (i.e. it would be relatively unlikely that a given molecular modelling protocol would achieve good performance by chance if the dataset is sufficiently large). Datasets could be devised with the input of impartial experienced molecular modellers privy to the blinded data so the level of challenge posed by a given dataset may be anticipated. For instance the markedly different topology of compound **60** in set 3 could have been expected to pose considerable difficulties for a relative free energy calculation protocol. Arguably efforts are better spent validating initially methodologies on cases that are perceived as less challenging.

Ultimately molecular modelling protocols should handle robustly both docking and scoring aspects for reliable routine use; however there is merit in constructing datasets that stress-test specific methodological aspects, and incrementally validate the domain of applicability of a protocol. Careful design of futures challenges is a difficult but important task that should encourage the community to explore different source of molecular modelling errors in a more controlled fashion.

Acknowledgements

J.M. is supported by a Royal Society University Research Fellowship. This project has received funding from the Europeans Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 655677 awarded to J. J-J., and from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC grant agreement No. 336289. This work reflects author's views and the Research Executive Agency is not responsible for any use that may be made of the information it contains.

References

[1] W.L. Jorgensen, The many roles of computation in drug discovery., Science. 303 (2004) 1813–1818. doi:10.1126/science.1096361.

[2] B.C. Duffy, L. Zhu, H. Decornez, D.B. Kitchen, Early phase drug discovery: Cheminformatics and computational techniques in identifying lead series, Bioorganic Med. Chem. 20 (2012) 5324–5342. doi:10.1016/j.bmc.2012.04.062.

[3] J. Michel, Current and emerging opportunities for molecular simulations in structure-based drug design., Phys. Chem. Chem. Phys. 16 (2014) 4465–77. doi:10.1039/c3cp54164a.

[4] M. Bermudez, G. Wolber, Structure versus function - The impact of computational methods on the discovery of specific GPCR-ligands, Bioorganic Med. Chem. 23 (2015) 3907–3912. doi:10.1016/j.bmc.2015.03.026.

[5] B. Kuhn, W. Guba, J. Hert, D. Banner, C. Bissantz, S. Ceccarelli, et al., A Real-World Perspective on Molecular Design, J. Med. Chem. (2016) 4087–4102. doi:10.1021/acs.jmedchem.5b01875.

[6] J.D. Durrant, J.A. McCammon, Molecular dynamics simulations and drug discovery., BMC Biol. 9 (2011) 71. doi:10.1186/1741-7007-9-71.

[7] M.R. Shirts, D.L. Mobley, S.P. Brown, Free-energy calculations in structure-based drug design, Drug Des. (2010) 61–86. doi:http://dx.doi.org/10.1017/CBO9780511730412.007.

[8] M. De Vivo, M. Masetti, G. Bottegoni, A. Cavalli, Role of Molecular Dynamics and Related Methods in Drug Discovery., J. Med. Chem. (2016) acs.jmedchem.5b01684. doi:10.1021/acs.jmedchem.5b01684.

[9] X. Barril, F. Javier Luque, Molecular simulation methods in drug discovery: A prospective outlook, J. Comput. Aided. Mol. Des. 26 (2012) 81–86. doi:10.1007/s10822-011-9506-1.

[10] K.M. Merz, Limits of free energy computation for protein-ligand interactions, J. Chem. Theory Comput. 6 (2010) 1769–1776. doi:10.1021/ct100102q.

[11] J. Michel , N. Foloppe, J.W. Essex, Rigorous Free Energy Calculations in Structure-Based Drug Design, Mol. Inform. 29 (2010) 570–578. doi:10.1002/minf.201000051.

[12] N. Homeyer, F. Stoll, A. Hillisch, H. Gohlke, Binding free energy calculations for lead optimization: Assessment of their accuracy in an industrial drug design context, J. Chem. Theory Comput. 10 (2014) 3331–3344. doi:10.1021/ct5000296.

[13] N. Hansen, W.F. Van Gunsteren, Practical aspects of free-energy calculations: A review, J. Chem. Theory Comput. 10 (2014) 2632–2647. doi:10.1021/ct500161f.

[14] W. Sinko, S. Lindert, J.A. Mccammon, Accounting for Receptor Flexibility and Enhanced Sampling Methods in Computer-Aided Drug Design, Chem. Biol. Drug Des. 81 (2013) 41–49. doi:10.1111/cbdd.12051.

[15] M. Fischer, R.G. Coleman, J.S. Fraser, B.K. Shoichet, Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery, Nat. Chem. 6 (2014) 575–583. doi:10.1038/nchem.1954.

[16] H. Gohlke, G. Klebe, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, Angew. Chemie - Int. Ed. 41 (2002) 2644–2676. doi:10.1002/1521-

3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O.

[17] C.Y. Yang, H. Sun, J. Chen, Z. Nikolovska-Coleska, S. Wang, Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction, J. Am. Chem. Soc. 131 (2009) 13709–13721. doi:10.1021/ja9039373.

[18] F. Forti, C.N. Cavasotto, M. Orozco, X. Barril, F.J. Luque, A multilevel strategy for the exploration of the conformational flexibility of small molecules, J. Chem. Theory Comput. 8 (2012) 1808–1819. doi:10.1021/ct300097s.

[19] P. Pisani, P. Piro, S. Decherchi, G. Bottegoni, D. Sona, V. Murino, et al., Describing the conformational landscape of small organic molecules through Gaussian mixtures in dihedral space, J. Chem. Theory Comput. 10 (2014) 2557–2568. doi:10.1021/ct400947t.

[20] J. Trepel, M. Mollapour, G. Giaccone, L. Neckers, Targeting the dynamic HSP90 complex in cancer., Nat. Rev. Cancer. 10 (2010) 537–549. doi:10.1038/nrc2887.

[21] L. Whitesell, S.L. Lindquist, HSP90 and the chaperoning of cancer., Nat. Rev. Cancer. 5 (2005) 761–72. doi:10.1038/nrc1716.

[22] D. Mahalingam, R. Swords, J.S. Carew, S.T. Nawrocki, K. Bhalla, F.J. Giles, Targeting HSP90 for cancer therapy., Br. J. Cancer. 100 (2009) 1523–9. doi:10.1038/sj.bjc.6605066.

[23] R. Amaro, V. Feher, M.K. Gilson, S.K. Burley, Drug Design Data Resource: An Open Resource to Advance Computer-Aided Drug Design, (n.d.). https://drugdesigndata.org (accessed May 31, 2016).

[24] J.R. Huth, C. Park, A.M. Petros, A.R. Kunzer, M.D. Wendt, X. Wang, et al., Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies, Chem. Biol. Drug Des. 70 (2007) 1–12. doi:10.1111/j.1747-0285.2007.00535.x.

[25] L. Wright, X. Barril, B. Dymock, L. Sheridan, A. Surgenor, M. Beswick, et al., Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms, Chem. Biol. 11 (2004) 775–785. doi:10.1016/j.chembiol.2004.03.033.

[26] X. Barril, S.D. Morley, Unveiling the full potential of flexible receptor docking using multiple crystallographic structures, J. Med. Chem. 48 (2005) 4432–4443.

doi:10.1021/jm048972v.

[27] D. Bucher, B.J. Grant, P.R. Markwick, J.A. McCammon, Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics, PLoS Comput. Biol. 7 (2011). doi:10.1371/journal.pcbi.1002034.

[28] L.C.T. Pierce, R. Salomon-Ferrer, C. Augusto F. De Oliveira, J.A. McCammon, R.C. Walker, Routine access to millisecond time scale events with accelerated molecular dynamics, J. Chem. Theory Comput. 8 (2012) 2997–3002. doi:10.1021/ct300284c.

[29] A. Laio, M. Parrinello, Escaping Free-Energy Minima, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 12562–12566. doi:10.1073/pnas.202427399.

[30] R.L. Dunbrack, M. Karplus, Backbone-dependent rotamer library for proteins. Application to side-chain prediction., J. Mol. Biol. 230 (1993) 543–574. doi:10.1006/jmbi.1993.1170.

[31] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, et al., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, J. Med. Chem. 47 (2004) 1739–1749. doi:10.1021/jm0306430.

[32] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, A.B. Garmendia-Doval, S. Juhos, P. Schmidtke, et al., rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids, PLoS Comput. Biol. 10 (2014). doi:10.1371/journal.pcbi.1003571.

[33] H. Alonso, A.A. Bliznyuk, J.E. Gready, Combining docking and molecular dynamic simulations in drug design, Med. Res. Rev. 26 (2006) 531–568. doi:10.1002/med.20067.

[34] M.A. Lill, Efficient incorporation of protein flexibility and dynamics into molecular docking simulations, Biochemistry. 50 (2011) 6157–6169. doi:10.1021/bi2004558.

[35] R.R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (2006) 5–30. doi:10.1016/j.acha.2006.04.006.

[36] B. Nadler, S. Lafon, R.R. Coifman, I.G. Kevrekidis, Diffusion maps, spectral clustering and reaction coordinates of dynamical systems, Appl. Comput.

Harmon. Anal. 21 (2006) 113–127. doi:10.1016/j.acha.2005.07.004.

[37] S. Röblitz, M. Weber, Fuzzy spectral clustering by PCCA+: Application to Markov state models and data classification, Adv. Data Anal. Classif. 7 (2013) 147–179. doi:10.1007/s11634-013-0134-6.

[38] M.K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, et al., PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models, J. Chem. Theory Comput. 11 (2015) 5525–5542. doi:10.1021/acs.jctc.5b00743.

[39] G. Neudert, G. Klebe, fconv: Format conversion, manipulation and feature computation of molecular data, Bioinformatics. 27 (2011) 1021–1022. doi:10.1093/bioinformatics/btr055.

[40] C.R. Søndergaard, M.H.M. Olsson, M. Rostkowski, J.H. Jensen, Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values, J. Chem. Theory Comput. 7 (2011) 2284–2295. doi:10.1021/ct200133y.

[41] M.H.M. Olsson, C.R. SØndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: Consistent treatment of internal and surface residues in empirical p K a predictions, J. Chem. Theory Comput. 7 (2011) 525–537. doi:10.1021/ct100578z.

[42] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges, Tetrahedron. 36 (1980) 3219–3228. doi:10.1016/0040-4020(80)80168-2.

[43] X. Barril, P. Brough, M. Drysdale, R.E. Hubbard, A. Massey, A. Surgenor, et al., Structure-based discovery of a new class of Hsp90 inhibitors., Bioorg. Med. Chem. Lett. 15 (2005) 5187–91. doi:10.1016/j.bmcl.2005.08.092.

[44] D.L. Mobley, P. V. Klimovich, Perspective: Alchemical free energy calculations for drug discovery, J. Chem. Phys. 137 (2012). doi:10.1063/1.4769292.

[45] J.D. Chodera, D.L. Mobley, M.R. Shirts, R.W. Dixon, K. Branson, V.S. Pande, Alchemical free energy methods for drug discovery: Progress and challenges, Curr. Opin. Struct. Biol. 21 (2011) 150–160. doi:10.1016/j.sbi.2011.01.011.

[46] S.K. Mishra, G. Calabro, H.H. Loeffler, J. Michel, J. Koca, Evaluation of

Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate Complexes, J. Chem. Theory Comput. 11 (2015) 3333–3345. doi:10.1021/acs.jctc.5b00159.

[47] J. Michel, J.W. Essex, Prediction of protein-ligand binding affinity by free energy simulations: Assumptions, pitfalls and expectations, J. Comput. Aided. Mol. Des. 24 (2010) 639–658. doi:10.1007/s10822-010-9363-3.

[48] J.G. Zeevaart, L. Wang, V. V. Thakur, C.S. Leung, J. Tirado-Rives, C.M. Bailey, et al., Optimization of azoles as anti-human immunodeficiency virus agents guided by free-energy calculations, J. Am. Chem. Soc. 130 (2008) 9492–9499. doi:10.1021/ja8019214.

[49] T. Steinbrecher, A. Hrenn, K.L. Dormann, I. Merfort, A. Labahn, Bornyl (3,4,5-trihydroxy)-cinnamate--an optimized human neutrophil elastase inhibitor designed by free energy calculations., Bioorg. Med. Chem. 16 (2008) 2385–90. doi:10.1016/j.bmc.2007.11.070.

[50] J. Wang, Y. Deng, B. Roux, Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials., Biophys. J. 91 (2006) 2798–814. doi:10.1529/biophysj.106.084301.

[51] M.R. Shirts, J.D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states, J. Chem. Phys. 129 (2008). doi:10.1063/1.2978177.

[52] C. Woods, A.S.J. Mey, G. Calabro, J. Michel, Sire Molecular Simulations Framework, (n.d.). Http://siremol.org/ (accessed May 31, 2016).

[53] M. Bruncko, S.K. Tahir, X. Song, J. Chen, H. Ding, J.R. Huth, et al., N-aryl-benzimidazolones as novel small molecule HSP90 inhibitors., Bioorg. Med. Chem. Lett. 20 (2010) 7503–6. doi:10.1016/j.bmcl.2010.10.010.

[54] A. Caroli, F. Ballante, R.B. Wickersham, F. Corelli, R. Ragno, Hsp90 inhibitors, part 2: Combining ligand-based and structure-based approaches for virtual screening application, J. Chem. Inf. Model. 54 (2014) 970–977. doi:10.1021/ci400760a.

[55] C.W. Murray, M.G. Carr, O. Callaghan, G. Chessari, M. Congreve, S. Cowan, et al., Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency, J. Med. Chem. 53 (2010) 5942–5955.

doi:10.1021/jm100059d.

[56] H.H. Loeffler, J. Michel, C. Woods, FESetup: Automating Setup for Alchemical Free Energy Simulations, J. Chem. Inf. Model. 55 (2015) 2485–2490. doi:10.1021/acs.jcim.5b00368.

[57] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general Amber force field, J. Comput. Chem. 25 (2004) 1157–1174. doi:10.1002/jcc.20035.

[58] A. Jakalian, B.L. Bush, D.B. Jack, C.I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method, J. Comput. Chem. 21 (2000) 132–146. doi:10.1002/(sici)1096-987x(20000130)21:2<132::aid-jcc5>3.0.co;2-p.

[59] A. Jakalian, D.B. Jack, C.I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation, J. Comput. Chem. 23 (2002) 1623–1641. doi:10.1002/jcc.10128.

[60] C. Steffen, K. Thomas, U. Huniar, A. Hellweg, O. Rubner, A. Schroer, Accelerating Molecular Dynamic Simulation on Graphics Processing Units, J. Comput. Chem. 30 (2009) 9. doi:10.1002/jcc.

[61] C.W. Hopkins, S. Le Grand, R.C. Walker, A.E. Roitberg, Long-time-step molecular dynamics through hydrogen mass repartitioning, J. Chem. Theory Comput. 11 (2015) 1864–1874. doi:10.1021/ct5010406.

[62] P. Eastman, V. Pande, OpenMM: A Hardware-Independent Framework for Molecular Simulations, Comput. Sci. Eng. 12 (2010) 34–39. doi:10.1109/MCSE.2010.27.

[63] P. Eastman, M.S. Friedrichs, J.D. Chodera, R.J. Radmer, C.M. Bruns, J.P. Ku, et al., OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation, J. Chem. Theory Comput. 9 (2013) 461–469. doi:10.1021/ct300857j.

[64] L. Wang, T. Lin, R. Abel, Cycle Closure Estimation of Relative Binding Affinities and Errors, (2014). https://www.google.com/patents/US20140278295.

[65] A. a. Hagberg, D. a. Schult, P.J. Swart, Exploring network structure, dynamics, and function using NetworkX, Proc. 7th Python Sci. Conf. 836 (2008) 11–15.

http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=960616.

[66] J. Luccarelli, J. Michel, J. Tirado-Rives, W.L. Jorgensen, Effects of Water Placement on Predictions of Binding Affinities for p38α MAP Kinase Inhibitors., J. Chem. Theory Comput. 6 (2010) 3850–3856. doi:10.1021/ct100504h.

[67] X. Barril, M.C. Beswick, A. Collier, M.J. Drysdale, B.W. Dymock, A. Fink, et al., 4-Amino derivatives of the Hsp90 inhibitor CCT018159, Bioorganic Med. Chem. Lett. 16 (2006) 2543–2548. doi:10.1016/j.bmcl.2006.01.099.

[68] J. Ren, J. Li, Y. Wang, W. Chen, A. Shen, H. Liu, et al., Identification of a new series of potent diphenol HSP90 inhibitors by fragment merging and structure-based optimization, Bioorganic Med. Chem. Lett. 24 (2014) 2525–2529. doi:10.1016/j.bmcl.2014.03.100.

[69] R.M. Immormino, Y. Kang, G. Chiosis, D.T. Gewirth, Structural and quantum chemical studies of 8-aryl-sulfanyl adenine class Hsp90 inhibitors, J. Med. Chem. 49 (2006) 4953–4960. doi:10.1021/jm060297x.

[70] E. Casale, N. Amboldi, M.G. Brasca, D. Caronni, N. Colombo, C. Dalvit, et al., Fragment-based hit discovery and structure-based optimization of aminotriazoloquinazolines as novel Hsp90 inhibitors, Bioorganic Med. Chem. 22 (2014) 4135–4150. doi:10.1016/j.bmc.2014.05.056.

[71] C.M. McBride, B. Levine, Y. Xia, C. Bellamacina, T. Machajewski, Z. Gao, et al., Design, structure-activity relationship, and in vivo characterization of the development candidate NVP-HSP990, J. Med. Chem. 57 (2014) 9124–9129. doi:10.1021/jm501107q.

[72] W.M.J. Obermann, H. Sondermann, A.A. Russo, N.P. Pavletich, F.U. Hartl, In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis, J. Cell Biol. 143 (1998) 901–910. doi:10.1083/jcb.143.4.901.

[73] J. Michel, J. Tirado-Rives, W.L. Jorgensen, Energetics of displacing water molecules from protein binding sites: Consequences for ligand optimization, J. Am. Chem. Soc. 131 (2009) 15403–15411. doi:10.1021/ja906058w.

[74] G. Gerogiokas, M. Southey, M. Mazanetz, A. Heifetz, M. Bodkin, R. Law, et al., Evaluation of Water Displacement Energetics in Protein Binding Sites with Grid Cell Theory, Phys. Chem. Chem. Phys. (2015) -. doi:10.1039/C4CP05572A.

[75] J. Michel, J. Tirado-Rives, W.L. Jorgensen, Prediction of the water content in protein binding sites., J. Phys. Chem. B. 113 (2009) 13337–13346. doi:10.1021/jp9047456.

[76] J.W. Kaus, E. Harder, T. Lin, R. Abel, J.A. McCammon, L. Wang, How to deal with multiple binding poses in alchemical relative protein-ligand binding free energy calculations, J. Chem. Theory Comput. 11 (2015) 2670–2679. doi:10.1021/acs.jctc.5b00214.

[77] K.L. Damm-Ganamet, R.D. Smith, J.B. Dunbar, J.A. Stuckey, H.A. Carlson, CSAR benchmark exercise 2011-2012: Evaluation of results from docking and relative ranking of blinded congeneric series, J. Chem. Inf. Model. 53 (2013) 1853–1870. doi:10.1021/ci400025f.

[78] R.D. Smith, K.L. Damm-Ganamet, J.B. Dunbar, A. Ahmed, K. Chinnaswamy, J.E. Delproposto, et al., CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking and Scoring/Ranking Challenge, J. Chem. Inf. Model. (2015) 150929213517007. doi:10.1021/acs.jcim.5b00387.

[79] H.A. Carlson, R.D. Smith, K.L. Damm-Ganamet, J.A. Stuckey, A. Ahmed, M.A. Convery, et al., CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma, J. Chem. Inf. Model. (2016) acs.jcim.5b00523. doi:10.1021/acs.jcim.5b00523.