


DockBench as docking selector tool: the lesson learned from D3R Grand Challenge 2015

Veronica Salmaso¹ · Mattia Sturlese¹ · Alberto Cuzzolin¹ · Stefano Moro¹ 

Received: 2 June 2016 / Accepted: 7 September 2016 / Published online: 16 September 2016
© Springer International Publishing Switzerland 2016

Abstract Structure-based drug design (SBDD) has matured within the last two decades as a valuable tool for the optimization of low molecular weight lead compounds to highly potent drugs. The key step in SBDD requires knowledge of the three-dimensional structure of the target-ligand complex, which is usually determined by X-ray crystallography. In the absence of structural information for the complex, SBDD relies on the generation of plausible molecular docking models. However, molecular docking protocols suffer from inaccuracies in the description of the interaction energies between the ligand and the target molecule, and often fail in the prediction of the correct binding mode. In this context, the appropriate selection of the most accurate docking protocol is absolutely relevant for the final molecular docking result, even if addressing this point is absolutely not a trivial task. D3R Grand Challenge 2015 has represented a precious opportunity to test the performance of DockBench, an integrate informatics platform to automatically compare RMSD-based molecular docking performances of different docking/scoring methods. The overall performance resulted in the blind prediction are encouraging in particular for the

pose prediction task, in which several complex were predicted with a sufficient accuracy for medicinal chemistry purposes.

Keywords DockBench · D3R Grand Challenge 2015 · Blind prediction · Molecular docking · Docking benchmark

Introduction

Molecular docking is widely adopted SBDD approach and its impact is clearly demonstrated by the plethora of software developed until now. In the Click2Drug directory [1] more than 50 software are listed, while more than 60 are catalogued on Wikipedia [2]. Considering that several docking algorithms can be coupled to different scoring functions, the number of different docking/scoring combinations is extremely vast.

The primary issue all docking programs try to address is what combination of orientation and conformation (pose) is the most favorable relative to all the other combinations sampled. When applied to screening, the process also requires a comparison of the best pose (or top best poses) of a given ligand with those of the other ligands such that a final ranking (or ordering) can be obtained. However, molecular docking protocols suffer from inaccuracies in the description of the interaction energies between the ligand and the target molecule, and often fail in the prediction of the correct binding mode. In this context, the appropriate selection of the most accurate docking protocol is absolutely relevant for the final molecular docking performance, even if addressing this point is absolutely not a trivial task for several reasons: (a) each docking protocol has its peculiar input and output

Veronica Salmaso and Mattia Sturlese have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-016-9966-4) contains supplementary material, which is available to authorized users.

✉ Stefano Moro
stefano.moro@unipd.it

¹ Molecular Modeling Section (MMS), Department of Pharmaceutical and Pharmacological Sciences, University of Padova, Via Marzolo 5, Padua, Italy

file formats, making their managing really tedious when different software are used; (b) input docking parameters can be very different among diverse programs strongly limiting their use in parallel; (c) more and more frequently it is possible that a molecular target was crystallized in more than one form, and it is then necessary to determine which of these is the most suitable for the docking procedure, in particular, when applied to a virtual screening study; and d) last but not least, the fundamental role played by water molecules during the molecular docking simulations.

To overcome these critical issues, we recently developed a tool to support the molecular modeler in identifying the most accurate protocol by an automated and simultaneous comparison of 17 docking/scoring combinations using a self-docking benchmark procedure [3]. In particular, DockBench is an integrate informatics platform to automatically compare RMSD-based molecular docking performances of different docking/scoring methods. An intuitive graphical analysis can help docking users, including non-expert ones, to identify the best docking/scoring combination to perform a docking-based virtual screening campaign. In this contest, D3R Grand Challenge 2015 has represented a precious opportunity to test the performance of DockBench tool in a blind exercise and using high quality ligand–protein complex structures. In particular, D3R Grand Challenge 2015 was organized allowing participants to compete, in a two-stage process, in the prediction of ligand pose and ligand ranking using two very well known therapeutic targets: heat shock protein 90 (Hsp90) and Mitogen activated protein kinase kinase kinase 4 (MAP4K4). Hsp90 is a chaperone protein which has been deeply investigated over the past decades for its crucial role in cancer cells [4], and MAP4K4 is a serine/threonine kinase that has emerged as such a potential therapeutic target for several disorders, in particular for metabolic and cardiovascular diseases [5].

Considering the peculiarity of the DockBench tool in facilitating the prediction of the ligand poses, we decided to concentrate our efforts in determining the best docking method able to reproduce the most accurate pose geometries. The results obtained in the D3R Grand Challenge 2015 (GC2015) revealed a promising capability of our pipeline in pose prediction task. In particular, the mean RMSD obtained in the Hsp90-complexes was 0.86 Å, while for MAP4K4-complexes the mean RMSD showed less accurate value (3.34 Å). The complete pipeline of DockBench used during the two-stage process of the D3R Grand Challenge 2015 GC2015 and a retrospective analysis of its performance will be described in the present study.

Experimental section

Overview of the workflow

The key concept of the workflow adopted in the GC2015 was the identification of the best protocol available in our laboratory in reproducing the crystallographic poses of selected ligands. In detail, given the target and a set of blind ligands, the workflow was articulated into four steps:

- (1) Collection of a training set of complexes containing the target from the protein data bank;
- (2) Comparing the performance in a self-docking procedure of 17 different docking protocol on the training set;
- (3) Selection of one or more suitable protocols according the RMSD;
- (4) Evaluation of the similarity of the blind set and the training set of ligands. If significant similarity was found, it drove the selection of the protein conformation;
- (5) Docking of the blind ligands;
- (6) Selection of the poses using scoring procedures and visual inspection for ambiguous conformations.

In the ranking predictions the protocol was mostly derived from the pose prediction workflow with a further implementation of rescoring procedures.

The procedure of each pose and rank prediction follows all the points depicted above, but tailoring few of them according the set of blind ligands and the protein target (detailed workflow in Figs. 1, 5) and is commented along the results and discussions.

Hardware

All computational studies were performed on a 200 cores cluster based on Ubuntu operating system (distribution 14.04, 64 bit) under the network file system (NFS) service. MD simulations were carried out by using Acemd [6] on a GPU cluster of 20 NVIDIA GTX graphics cards.

Ligands preparation

All ligands were prepared following an in-house pipeline previously reported [7]. Briefly, Corina 3.4 was used to generate three-dimensional structures, as well as to neutralize and deprive them of potential counterions [8]. For each compound, the most favorable ionic state was selected by using the “Protonate” tool implemented in MOE suite and based on Generalized Born electrostatics model [9]. MOE was also used to generate the possible tautomeric

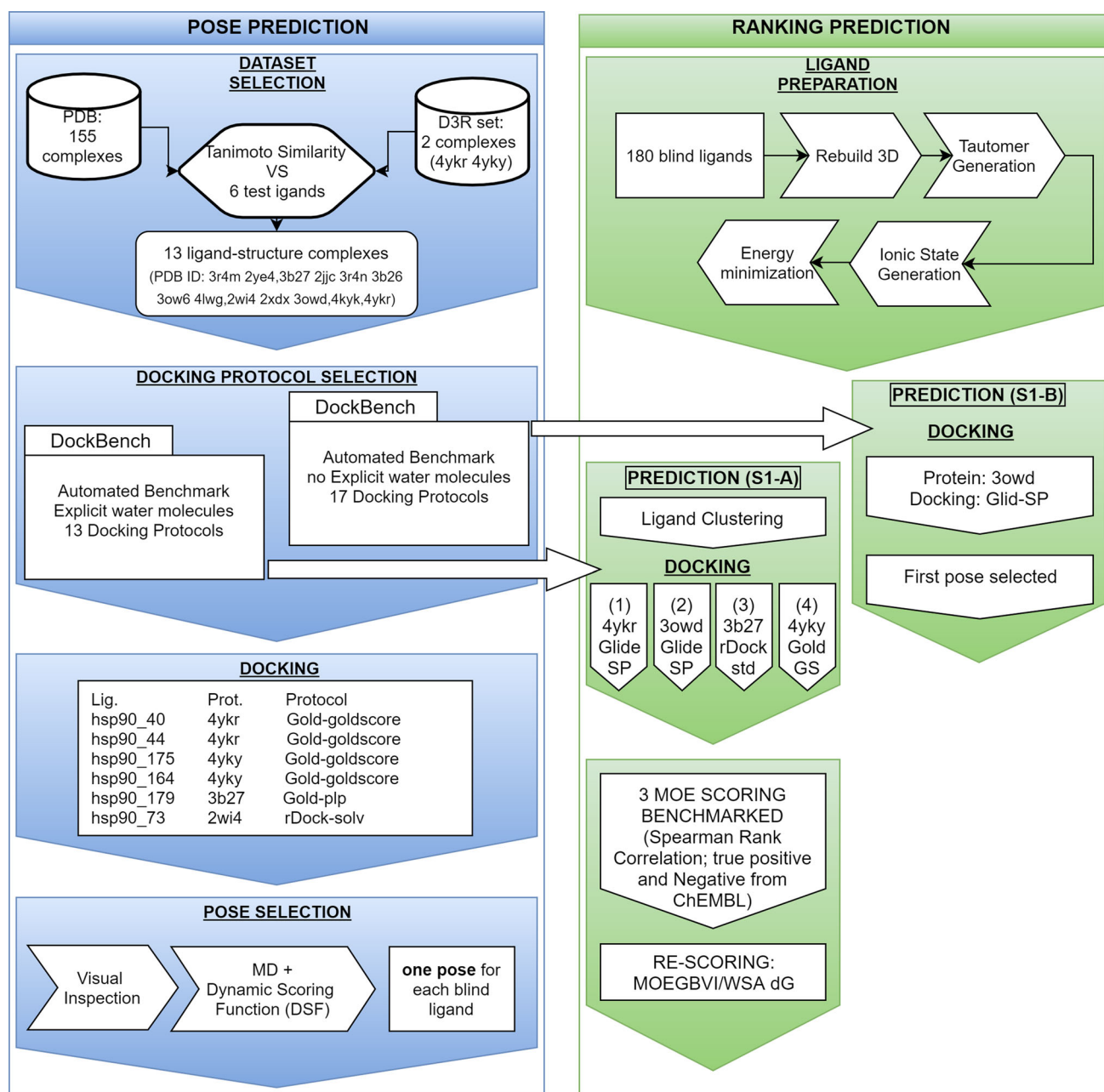


Fig. 1 Workflow for Posing and Scoring predictions designed for the challenge on Hsp90. In *blue panel* is reported the procedure used in the docking stage divided in four main tasks as reported in the discussion section: Database selection, Docking Protocol Selection,

Docking Calculation, and Pose Selection. The Scoring Prediction pipeline is schematized on the *green panel*. The Scoring Procedure consists in a first ligand preparation step, then two different prediction are sketched (Prediction S1-A and S1-B)

states, to energy minimize, and to assign the partial charges of each candidate using MMFF94x force field [10].

Preparation of ligand–protein complexes

The complexes provided by the organizers as well as those retrieved from the Protein Data Bank (PDB) [11] were subjected to the Structure Preparation and “Protonate-3D”

tools implemented in MOE2015.10 suite [9], including water molecules if present.

Molecular docking

Molecular docking calculations were carried out using the following software: AutoDock 4.2.5.1 [12], AutoDock Vina1.1.2 [13], Glide 6.5 [14, 15], GOLD 5.2 [16], MOE

2015.10 [17], PLANTS 1.2 [18], rDock [19]. DockBench 1.0 [3] was used to perform and analyze molecular docking benchmarks. DockBench default parameters have been set for all docking protocols. MOE 2015.10 was used for docking rescoring procedure, using the following scoring function: pKi, GBVI/WSA, Affinity dG [20].

Chemical similarity and docking analysis

In house bash or python scripts were used for determining Tanimoto's similarity using OpenBabel [21] and for calculating root mean square deviations (RMSD) using OpenEye [22], respectively. Visual inspection was performed on MOE 2015.10 and Chimera UCSF [23]. ChEMBL database [24] was queried to obtain experimental affinities using a substructure search tool as implemented in MOE.

Molecular dynamics simulations

Ligand-Hsp90 complexes selected among docking poses were prepared with AmberTool14 [25] for Molecular Dynamics (MD) simulations as follows.

Each system was solvated with explicit waters (TIP3P model) resulting in a box with boundaries at least 11Å far from any atom of the complex. The simulation box was neutralized with Na⁺/Cl⁻ ions to a final concentration of 0.1 M. Consequently, the prepared systems were simulated by using AMBER14 [26] Force Field [27] and periodic boundary conditions. General Amber Force Field (GAFF) [28] parameters were used for the ligands, along with RESP partial charges [29], which were obtained with Antechamber [25] by fitting electrostatic potential points calculated with Gaussian [30].

The system equilibration was performed through a stepwise procedure that begins with a conjugate-gradient minimization of 300 steps in order to reduce the steric clashes of the prepared system. The equilibration phase was performed through two consecutive steps, with different ensembles and atom positional restrains. In the first protocol, the MD simulation was performed in a NVE ensemble for 100 ps, with a force constant of 1 kcal mol⁻¹ Å⁻² applied to all protein atoms in order to allow the equilibration of the water molecules. Thereafter, a MD simulation of 500 ps in the NPT ensemble was performed by keeping the alpha-carbons of the protein restrained with the same force magnitude of the previous step. During this step, the temperature was maintained at 310 K by a Langevin thermostat and the pressure at 1 atm by a Berendsen barostat. Subsequently, all MD simulations were conducted in the NVT ensemble, maintaining the temperature at 310 K.

In all MD simulations, the non-bonded long-range Coulomb interactions were handled by using the particle mesh Ewald summation method (PME) [31] with a cutoff distance of 9 Å and a switching distance of 7.5 Å. All the poses were simultaneously compared in a knockout tournament framework.

Each MD simulation was carried out for 10 ns during which a modified dynamic scoring function (DSF) [32] was computed. This scoring is defined as the cumulative sum of the ligand–protein interaction energy (IE): it includes electrostatic (IEele) and van der Waals (IEvdw) contributions.

The wIE are plotted against the simulation time and linearly fitted to the collected data to obtain the slope coefficient that provides an estimation of the strength of the interaction and the stability of the binding mode.

Electrostatic energy fingerprints

Electrostatic interactions in MAP4K4-ligand complexes were studied by calculating the Electrostatic Energy Fingerprints (EEF). Amber99 partial charges were computed for the proteins and PM3 partial charges were computed for the ligands using MOE. Per residue electrostatic energy interactions were computed thanks to a in-house SVL script used in MOE. Interactions of the residues within 10 Å from each ligand were plotted in a heat map. This graph, reporting on the X-axis the protein residues of the binding site and on the Y-axis the ligands, attributes a color to the strength of the interactions: in particular, electrostatic energy diminishes going from red to blue. Gnuplot4.5 [33] was used to draw the plots.

Results and discussion

D3R Grand Challenge 2015 was organized as a two-stage process applied to Hsp90 and MAP4K4 datasets. In both cases, stage 1 was subdivided in two tasks: the first consisting of a “pose prediction” phase, and the second of a “ranking prediction” phase. Stage 2 had the same aim of stage 1 “ranking prediction” phase, with, as an advantage, the disclosure of the crystallographic structures object of phase 1 “pose prediction” phase.

As anticipated in the Introduction, our computational work was mainly devoted to “pose prediction” following the mantra concept in SBDD that the identification/selection of the most accurate docking protocol is the key step in the prediction of the correct binding mode. For this purpose, we have compared the ability of different docking/scoring combinations in reproducing crystallographic poses, taking advantage of the DockBench software.

Hsp90

Stage 1: Pose prediction phase

The challenge of Hsp90 Stage 1-“Pose prediction” phase was to predict the coordinates of six protein–ligand complexes and to rank the affinities of 180 compounds, referred to as “*ligand test set*” in this paper. The workflow used for the pose prediction is reported in Fig. 1 (on the left), and it is divided into four tasks: Hsp90 complexes selection, selection of docking protocol, docking calculations and, finally, best pose selection.

Hsp90 complexes selection Hsp90 is a well-known target in medicinal chemistry which has been deeply investigated in the last two decades by structural biology. At the time of the challenge, we identified in the Protein Data Bank [11] 155 Hsp90-ligand complexes, as listed in Supplementary Information. Two further complexes provided by the organizers (PDB ID: 4YKR, 4YKY) were added to the structures collected from the PDB. Due to the large amount of structural information, we decided to reduce the number of the crystallographic structures focusing our attention only to those complexes in which the co-crystallized ligands were structurally similar to those provided to us by the organizers. The selection was carried out using a filter based on Tanimoto’s similarity (FP2 fingerprints): in particular similarity was evaluated for each of the 6 ligands to be docked against the 157 crystallographic ligands. We selected the Hsp90 crystallographic complexes in which the co-crystallized ligand showed a similarity index greater than 0.5, resulting in 13 structures: 3R4 M [34], 2YE4 [35], 3B27 [36], 2JJC [37], 3R4 N [34], 3B26 [36], 3OW6 [38], 4LWG [39], 2WI4 [40], 2XDX [41], 3OWD [38], 4YKR [42], 4YKY [43] (referred to as their PDB ID). Co-crystallized ligands of these 13 complexes will be called “*ligand training set*” from here on out in this manuscript. Crystallographic structures and ligand training set were prepared for molecular docking study according to the pipeline reported in the Experimental section.

Selection of docking protocol In the case of Hsp90 the selection of the docking protocol has been carried out taking into account the possible presence of water molecules as mediators of interactions between the ligand and the residues of the binding cavity. The criterion that has been used for the selection of the water molecules is based both on the assessment of their direct interaction with the ligand and the protein, and of the similarity of their B-factor with the average B-factor of the heavy atoms of the backbone of the protein. A list of the water molecules taken into account for each crystal structure is reported in Table S11.

The *ligand training set* was subjected to two benchmark studies as reported in the workflow (Fig. 1): the first one, in which each ligand of the *training set* has been self-docked using 17 docking/scoring combinations in the absence of water molecules, and the second one in which the same ligands have been self-docked using 13 docking/scoring combinations protocols taking into account the selected water molecules.

In DockBench, to judge the performances of the different docking protocols, 20 poses were generated for each ligand of the *training set* and the RMSD values between predicted and crystallographic poses were calculated.

In order to evaluate the performances of the docking protocols, the lowest (RMSD_{min}) and average (RMSD_{ave}) RMSD values over the 20 poses, as well as the highest number of conformations with a RMSD value lower than the corresponding X-ray resolution (R), $N^{(RMSD < R)}$, were compared for all the docking protocols. For the specific purpose of the D3R Grand Challenge 2015, we have exploited DockBench ability in suggesting the most accurate docking protocols, that are the protocols able to predict the pose closest to the experimental one. For this reason, we have focalized our attention on the docking protocols showing lowest RMSD_{min} values. This resulted in different docking protocols in relation to different crystal structures, and sometimes in more than one successful docking protocol for the same crystal structure.

Interestingly, the results of DockBench indicated a significant improvement in reproducing the experimental crystallographic poses when the water molecules were included in the docking procedure, as reported in Fig. 2a. Indeed, including water molecules several protocols were able to reproduce the experimental coordinates with RMSD below 2.0 Å. Following these computational evidences, we decided to include the same crystallographic water molecules also during the docking simulations of the *ligand test set*.

Finally, for each ligand of the test set we have selected the crystallographic structure of Hsp90 in which the co-crystallized ligand was structurally more similar to the docked ligand.

Docking simulations and pose selection For each ligand of the *test set* we have chosen the crystallographic structures of Hsp90 whose co-crystallized ligand had higher Tanimoto similarity (calculated as FP2 fingerprints comparison) to it. In the case of compound 44 (a benzimidazol-2-one), we have decided to dock it to two structures: 3OWD, selected on the basis of highest similarity, and 4YKR, the benzimidazol-2-one derivative bound structure proposed by the challenge organizers. In the case of compound 73, Tanimoto index was not sufficient to discriminate structures, thus, besides evaluating chemical and

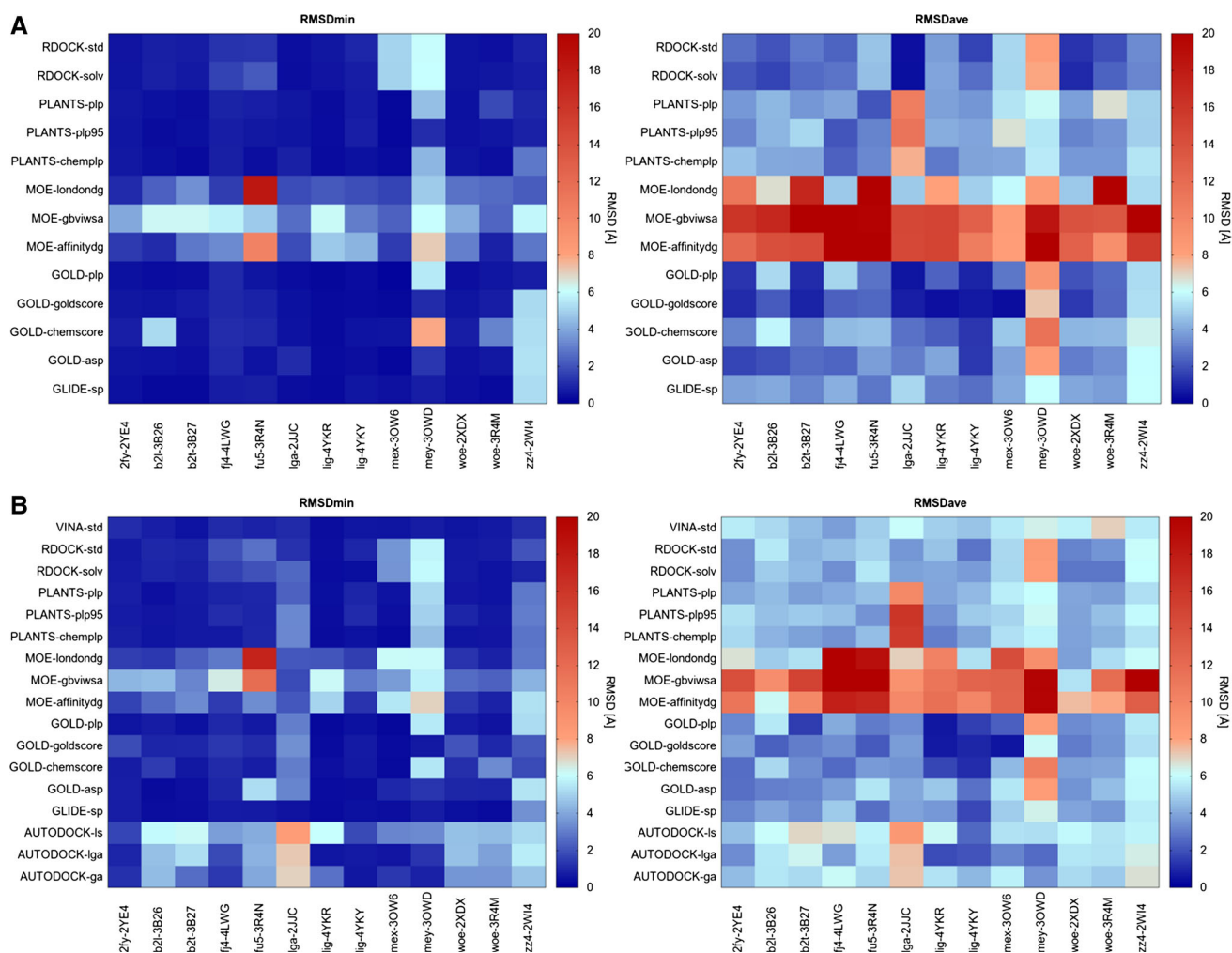


Fig. 2 Self-docking benchmark results obtained with DockBench. Two different benchmark are shown: in **a** a benchmark carried out including the most relevant water molecules, while in panel **B** the benchmark was performed on the same pool of complexes but removing all the crystallographic water molecules. For each panel two

heat map are reported: the minimum RMSD values (RMSD_{\min}) returned by the tested docking protocol (y -values) for the considered X-ray structures (x -values) and the Average RMSD values (RMSD_{ave}) for the 20 poses generated for each protocol considered. Values are color coded, blue spots identify the best obtained results

structural similarity of the co-crystallized ligands, we have chosen three protein structures showing different conformations of loop 104–114 near the binding site.

Finally, we have selected the docking protocols with best RMSD_{\min} performance for those crystallographic structures. The final report of the selected crystal structures and the relative docking protocols for each ligand of the *test set* is summarized in Table 1.

After the preliminary validation step using the *ligand training set*, the Virtual Screening Tool of DockBench was used to perform the *ligand test set* docking simulations using the same set of parameters adopted in the validation step. A summary of information used in the docking simulations of the *ligand test set* is collected in Table 1.

After docking, we selected one or more poses resulting from each docking simulation, according to electrostatic

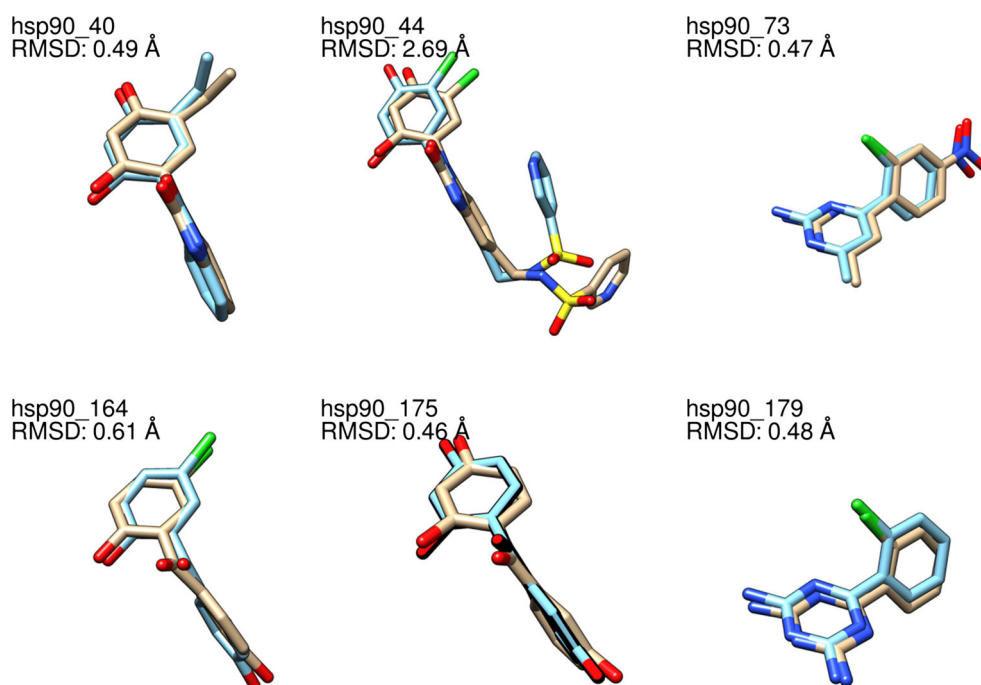
and van der Waals interaction energy evaluation and visual inspection. Finally, we used Molecular Dynamics (MD) simulation as post-docking tool to select a unique pose for the challenge submission [29]. For each pose a 10 ns simulation was performed and the dynamic scoring function (DSF) was evaluated. This scoring is computed along the trajectory with the aim to obtain the slope coefficient as an estimation of the binding strength and of the stability of the complex.

Results As anticipated, the proposed workflow was designed to produce a unique pose for each *ligand of the test set*. The superposition of the six predicted complexes on the corresponding crystallographic poses is reported on Fig. 3. The DockBench performance generally showed robust results (Table 2) with a mean RMSD of 0.86 Å

Table 1 List of *test set ligands* with relative docking protocols and protein crystallographic structures selected for docking in the pose prediction task

Ligand	Protein	Docking algorithm	Scoring function	Pose	Slope (DSF)
Hsp90_40	4YKR	Gold	Goldscore	1	−30.14
Hsp90_44	3OWD	Glide	SP	1	−12.94
	3OWD	Gold	Goldscore	2	−19.74; −19.81
	4YKR	Gold	Goldscore	3	−19.81; −22.57
				4	−17.48
Hsp90_73	3B27	Gold	PLP	1	−42.52
	3B26	Gold	Goldscore	2	−46.41; −29.07
	2WI4	rDock	Solv	3	−36.71; −36.02
	2WI4	rDock	STD	4	−8.37
Hsp90_164	4YKY	Gold	ASP	1	−29.85
				2	−37.86; −35.23
	4YKY	Gold	Goldscore	3	−48.70; −47.92
				4	−27.94
Hsp90_175	4YKY	Gold	ASP	1	−32.53; −30.17
				2	−30.25
	4YKY	Gold	Goldscore	3	−56.56; −54.96
				4	−20.75
Hsp90_179	3B27	Gold	PLP	1	−19.85
				2	−14.63

For each ligand the number of poses picked after docking and submitted to MD is reported, together with the slope of the DSF computed along the MD trajectory and used as final score (when two slope values are reported, the refer to the first and the second turn of the knockout tournament, respectively). In bold are indicated the poses selected on the basis of DSF slope and finally submitted to the challenge

Fig. 3 Superposition of the predicted poses (*light blue*) on the experimental ones (*tan*). RMSD values were calculated on the heavy atoms

considering only the heavy atoms of each docked ligand. Most notably, five complexes shown RMSD values under the 0.61 Å absolutely representative of crystallographic poses. Curiously, ligand Hsp90_44 showed a higher RMSD

value, 2.69 Å, mainly ascribable to the 3-pyridinesulfonamide moiety. This substituent in the crystal structure points out to the bulk water and is characterized by high B-factor values while in our prediction it is differently

Table 2 Summary of the results of all scoring and docking prediction

Receipt-ID	Pred. name	Target/stage	Scoring prediction				
			Num. ligands	τ err (Kendall)	τ (Kendall)	ρ err (Spearman)	ρ (Spearman)
564e3304a7724	S1-A	Hsp90 Stage1	180	0.052	0.11	0.08	0.16
564e330569871	S1-B	Hsp90 Stage1	180	0.054	0.16	0.08	0.23
56afca8517dc5	S2-A	Hsp90 Stage2	180	0.05	0.21	0.07	0.3
56afca9a3644d	S2-B	Hsp90 Stage2	180	0.05	0.24	0.07	0.35
56afca7f6927b	S2-C	Hsp90 - Stage2	180	0.056	0.12	0.08	0.18
5671ef9fdd7a3		MAP4K4 Stage1	18	0.15	0.32	0.2	0.46
56afc9e2ae8c8		MAP4K4 Stage2	18	0.203	-0.02	0.26	0.01
			Pose prediction				
			Num. Poses	RMSD (mean pose 1)	RMSD (mean all poses)	RMSD (mean best pose)	
564e43759677e	Hsp90 Stage1	5 ^a	0.5		0.5	0.5	
5671f1dac24a1	MAP4K4 Stage1	30	3.34		3.34	3.34	

The values reported correspond to those provided by the organizers. The Values of RMSD are indicated in angstrom (Å)

^a The final evaluation of Grand Challenge 2015 considers only 5 ligands; in the discussion the we included also the ligand Hsp90_44, resulting in a mean RMSD for 6 ligand of 0.86 Å

oriented establishing a pi stacking interaction with the benzimidazol-2-one scaffold (as shown in Fig. 4). Despite the shift of the 3-pyridinesulfonamide moiety, the key

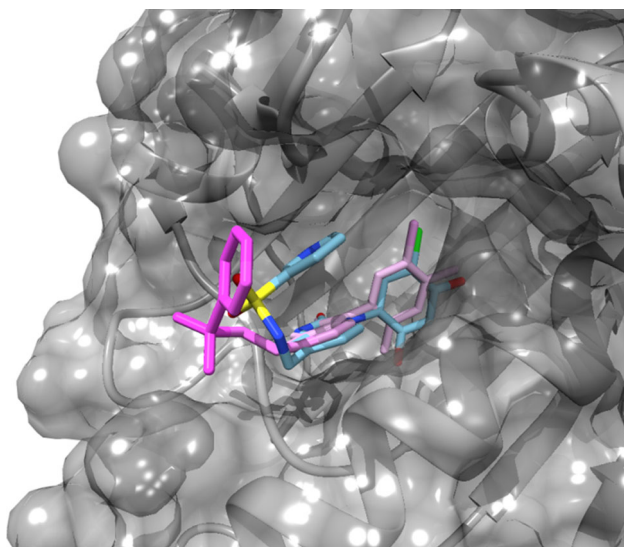


Fig. 4 Comparison of the predicted pose (*light blue*) and the experimentally derived complexes. The crystallographic ligand is colored according the B-factor in a light-to-dark pink palette corresponding to low-to-high values. While the benzimidazol-2-one scaffold is in nicely reproduced 0.62 Å the 3-pyridinesulfonamide moiety is placed out from the binding pocket is not well predicted resulting in a RMSD of 2.69 Å for the whole molecule. The binding mode of the portion establishing the key interaction is not affected by the different orientation. This observation is in agreement with the higher B-factor values of the 3-pyridinesulfonamide moiety (*dark pink*)

interactions of this scaffold are conserved as well as the orientation of the N-substituted benzimidazol-2-one portion as confirmed by the good RMSD (0.62 Å) calculated considering only this portion of the molecule.

Stage 1: ranking prediction step

The aim of Hsp90 Stage 1-“Ranking prediction” phase was to rank the affinities of 180 compounds, referred to as “*ligand test set*” in this paper. The workflow used for the ranking prediction is reported in Fig. 1 (on the right).

Scoring workflow As already anticipated in the Introduction, docking programs are usually successful in generating multiple poses that include binding modes similar to the crystallographically determined bound structure whereas scoring functions are much less successful at correctly ranking the “bioactive” binding mode. Aware of the current limitations of the scoring functions, however, we wanted to compare two ranking methodologies that represent on the one hand the most accurate ranking strategy available in our lab (S1-A) and the other the less expensive in terms of computational time (S1-B) (Fig. 1). This comparison was intriguing for us to establish the possible benefit-cost ratio of these two alternative strategies.

In the first pipeline (S1-A), we clustered the library of 180 compounds according to Tanimoto’s similarity exploiting the Fingerprint Database Clustering tool of MOE: briefly, Tanimoto’s similarity was computed for all

the 180 compounds against all of them, and each cluster was composed by molecules which were similar to the same set of molecules. Each cluster was screened by structural similarity (evaluated on the basis of common scaffold search, guided by user's chemical sensibility and experience) against the 13 ligands of the training set used in the previous benchmark. We selected the protein corresponding to the co-crystallized ligand with highest similarity to each cluster. The PDB ID of the 13 protein–ligand complexes subjected to DockBench were: 3R4 M, 2YE4, 3B27, 2JJC, 3R4 N, 3B26, 3OW6, 4LWG, 2WI4, 2XDX, 3OWD, 4YKR, AND 4YKY. After merging some of the clusters according to structural similarity of compounds scaffolds (evaluated by user's chemical sensitivity and experience), we identified 4 clusters (Table SI2) corresponding to 4 different protein–ligand complexes: 3OWD (2,3-dihydro-1H-benzimidazol-5-yl-methylsulfonamide scaffold), 4YKY (benzophenone scaffold), 4YKR (1,3-dihydro-2H-benzimidazol-2-one scaffold), and 3B27 (2-amino-1,3,5-triazine scaffold) as detailed in SI. Differently to the pose prediction challenge, here we selected the docking protocol for the four complexes using, in addition to DockBench results (RMSD_{min} and RMSD_{ave}), also the Spearman's and Kendall's correlations to evaluate the ability of the protocol to rank the near native pose at the top positions of the ranking list. Briefly, each protocol showing RMSD_{min} and RMSD_{ave} below 1 Å and 4 Å (Fig. 2), respectively were then compared according Spearman's and Kendall's coefficients (score vs RMSD). The final selection is reported in Table 3. We performed the docking calculation using the Virtual Screening Platform implemented in DockBench using the same parameters adopted in the previous benchmark and the first pose (best score) for each ligand was selected. Finally, to have a homogeneous scoring method, different scoring functions were evaluated for the rescoring procedure. Briefly, we picked a subset of compounds from ChEMBL with known activity (true positive and true negative) for each cluster by a substructure search. Only for cluster2 and cluster3 we identified a sufficient number of ligands, 14 and 17

Table 3 Combination of docking protocol, the PDB ID of protein conformation used for each cluster identify in the rank prediction stage 1 (Hsp90)

Cluster	Population	Protein	Protocol	$\rho; \tau$
1	44	4YKR	Glide-sp	0.83; 0.63
2	17	3OWD	Glide-sp	0.68; 0.47
3	62	3B27	rDock-std	0.73; 0.51
4	57	4YKY	Gold-chemscore	0.75; 0.56

The size of the cluster is indicated for each protocol (population). The Spearman's and Kendall's coefficients are reported for the selected protocols (RMSD vs score)

respectively (Table SI3), to have a raw indication of the classification ability of the tested scoring functions. For those clusters the Spearman and Kendall coefficient were calculated to identify the most performant scoring function (GBVI/WSA dG). The 180 compounds were finally ranked on the basis of the GBVI/WSA dG value of the selected pose.

In the “less than one hour approach” (S1-B), we selected Glide-sp, according to the metrics resulted by the benchmark without water molecules on the protein PDB ID 3OWD (RMSD_{min} : 0.67 Å), chosen on the basis of its wider binding pocket, suitable, at least in principle, to accommodate different classes of compounds. The screening was performed using Glide-sp from the DockBench Virtual Screening platform and for each ligand the best pose (lowest pseudo-energy) was selected. Glide score was used to rank the 180 compounds.

Scoring results As expected, both scoring strategies S1-A and S1-B showed their ineffectiveness in the ability to correctly rank ligands in terms of their binding affinities and, also, in discriminating between true positive and true negative active compounds. In fact, as reported in Table 2 their ranking performances measured by the Kendall correlation are 0.11 and 0.16 considering S1-A and S1-B ranking, respectively. These performances suggest that the apparently more accurate S1-A ranking strategy is not superior in terms of ranking accuracy respect the fast S1-B method.

Stage 2: ranking prediction step

The stage 2 of the D3R Grand Challenge 2015 was characterized by the release, from the organizers, of the Hsp90 crystallographic structures used as test set in the pose prediction phase of stage 1. As in the stage 1, also here it was compared the two previously ranking methodologies (S1-A and S1-B) with the aim to rank the same 180 ligands analyzed in the stage 1 but taking into account the additional available crystallographic information.

Scoring workflow The applied workflow in stage 2 retraced the pipeline described for stage 1, and reported in Fig. 1, with few exceptions. In fact, the stage 2 of the D3R Grand Challenge 2015 was characterized by the release, from the organizers, of the Hsp90 crystallographic structures used as test set in the stage 1. Consequently, we re-performed the docking benchmark study of stage 1 using Hsp90 crystallographic structures (PDB ID: 2XDX, 4YKW, 4YKY, 4YKQ, 2YE4, 4YKT, 3R4 N, 2JJC, 2WI4, 3B26, 4YKZ, 3OW6, 3B27, 3OWD, 4YKR, 4YKU, 4YKX, 4LWG, 3R4 M). As previously described, also in this case all docking simulations have been carried out

including the more crucial water molecules (Table S11). The new benchmark was also interesting to retrospectively analyze the ability of the docking/scoring combinations in reproducing the new crystallographic poses and, therefore, to evaluate the goodness of our protocol selection in the stage 1. The results of the new benchmark are reported in Figure S11 (panel A). Interestingly, the protocols selected in the stage 1 showed low RMSD_{\min} also in the self-docking exercise confirming, again, the goodness in the identification of the docking protocol.

Moreover, the ranking prediction (S2-A) also retraced the S1-A pipeline. Again, we clustered the 180 ligands according Tanimoto's similarity to the ligands co-crystallized as included in the benchmark (in presence of the most relevant water). In this way we obtained 7 clusters as listed in SI (Table S14). For each of them, we carried out the docking calculation selecting the protocol according the RMSD_{\min} and RMSD_{ave} performances but also considering the ability in discriminating the near native conformation within the family of conformations generated in the benchmark. To highlight this, we used the Spearman index correlating the RMSD versus the score. The resulting combination of cluster, protein and protocol is detailed in Table 4.

Then, we extracted the best scoring pose for each ligand according to the scoring method proper of the protocol. Finally, to be able to rank ligands conformations originated from different docking protocols, we sorted all the best conformations by using a rescoring procedure with MOE-pKi function.

As previously mentioned, the second ranking submission (S2-B) is strictly link to the first. It was designed to highlight the effect in considering the Spearman's correlation in the protocol selection.

In more detail, all steps of this pipeline were exactly the same of the prediction S2-A except in the selection of the protocol that in this case was merely based on the RMSD_{\min} and RMSD_{ave} performances obtained in the

benchmark. In Table 4 is reported which protocol was assigned for each clusters.

The third and last submission in the stage 2 (S2-C) was based on the submission S1-B in stage 1 and follow the same philosophy: simplest and fastest. The workflow adopted was exactly the same of stage 1. Briefly, we performed a new benchmark on the 19 complexes (13 complexes already known at stage 1 plus the new 6 unveil complexes) removing all the water molecules. Also in this case the benchmark outputs indicated Glide-sp, as the protocol more suitable in generate the near native conformations (Figure S11, panel B). We decided to use the same protein conformation used in S1-B (PDB ID: 30WD), which is characterized by a wider binding pocket able in principle to host different classes of compounds. The screening was carried out using the Virtual Screening tool of DockBenck selection for each ligand its more stable pose.

Scoring results Unexpectedly, the scoring performances of the stage 2 have been significantly different from those observed in phase 1. The three approaches appreciably differ in terms of ranking and classification capability (Table 2). The more articulated methods (S2-A and S2-B) outperformed the basic approach (S2-C); in the Kendall rank correlation the three predictions S2-A, S2-B, and, S2-C scored 0.24, 0.21, and, 0.12 respectively. Whereas the score of S2-C was expected due to the fact that is has been performed with the same methodology of in stage 1, the score of S2-A is doubled. The introduction of a more suitable protein conformation has improved the performance; however, the value is still far from a desirable value. Also considering the Spearman's rank correlation, the S2-A and S2-B outperformed S2-C with a coefficient of 0.30, 0.35, and 0.18 respectively. From the performance comparison of S2-A and S2-B is interesting to note that the use of the Spearman's correlation in the protocol selection has slightly improved the quality in the rank classification as partially expected.

Table 4 Combination of docking protocol and PDB ID of protein conformation used for each cluster identify in the rank prediction stage 2 (Hsp90)

Cluster	Population	Protein	S2-A protocol	S2-B protocol
1	31	4YKU	rDock-std	Gold-goldscore
2	11	4LWG	Gold-chemscore	Plants-chemplp
3	7	3B26	Gold-asp	rDock-std
4	21	4YKZ	rDock-std	Gold-goldscore
5	63	4YKR	Gold-asp	Gold-goldscore
6	13	3R4 N	Gold-plp	Gold-goldscore
7	34	4YKW	rDock-std	Gold-plp

The size of the cluster is indicated for each protocol (population). The S2-A and S2-B differ only for the protocol selected

MAP4K4

Stage 1: pose prediction step

The challenge on MAP4K4 Stage 1-“Pose prediction” step was to predict the coordinates of 30 protein–ligand complexes and to rank the affinity of 18 of these 30 compounds referred, also in these case, as “*ligand test set*”. The workflow used for the pose prediction is reported in Fig. 5 (on the left), and it is divided again into four tasks: MAP4K4 complexes selection, selection of docking protocol, docking calculations and, finally, best pose selection.

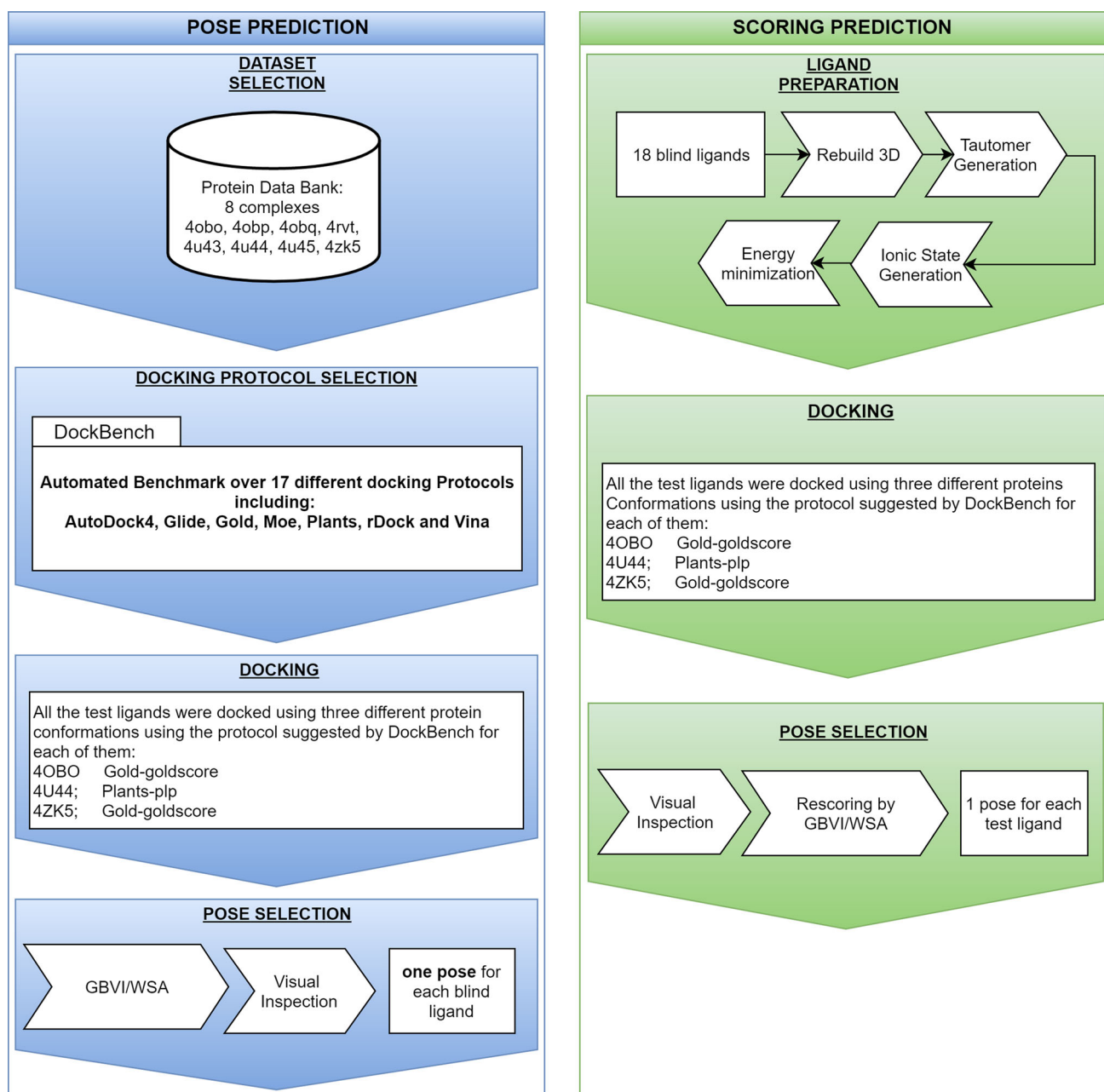


Fig. 5 Workflow for posing and scoring predictions designed for the challenge on MAP4K4. In *blue panel* is reported the procedure used in the docking stage divided in four main tasks as reported in the discussion section: Database selection, Docking Protocol Selection, Docking Calculation, and Pose Selection. The Scoring Prediction

pipeline is schematized on the *green panel*. The Scoring Procedure consists in three tasks strictly correlated to the posing challenge: a first ligand preparation step, docking calculation and finally the re-scoring and pose selection

MAP4K4 complexes selection Similar to what was done in stage 1 for Hsp90, we retrieved all eight ligand-MAP4K4 complexes present in the PDB (PDB ID: 4OBO [44], 4OBP [44], 4OBQ [44], 4RVT [45], 4U43 [46], 4U44 [46], 4U45 [46], and 4ZK5 [47]) in which the co-crystallized ligand will be referred again as “*ligand training set*”. Crystallographic structures and ligand training set were

prepared for molecular docking study according to the pipeline reported in the Experimental section.

Selection of docking protocol All the 8 known complexes were submitted to a self docking benchmark within DockBench using all the 17 different docking protocols available in the tool. Unlike what has been observed for

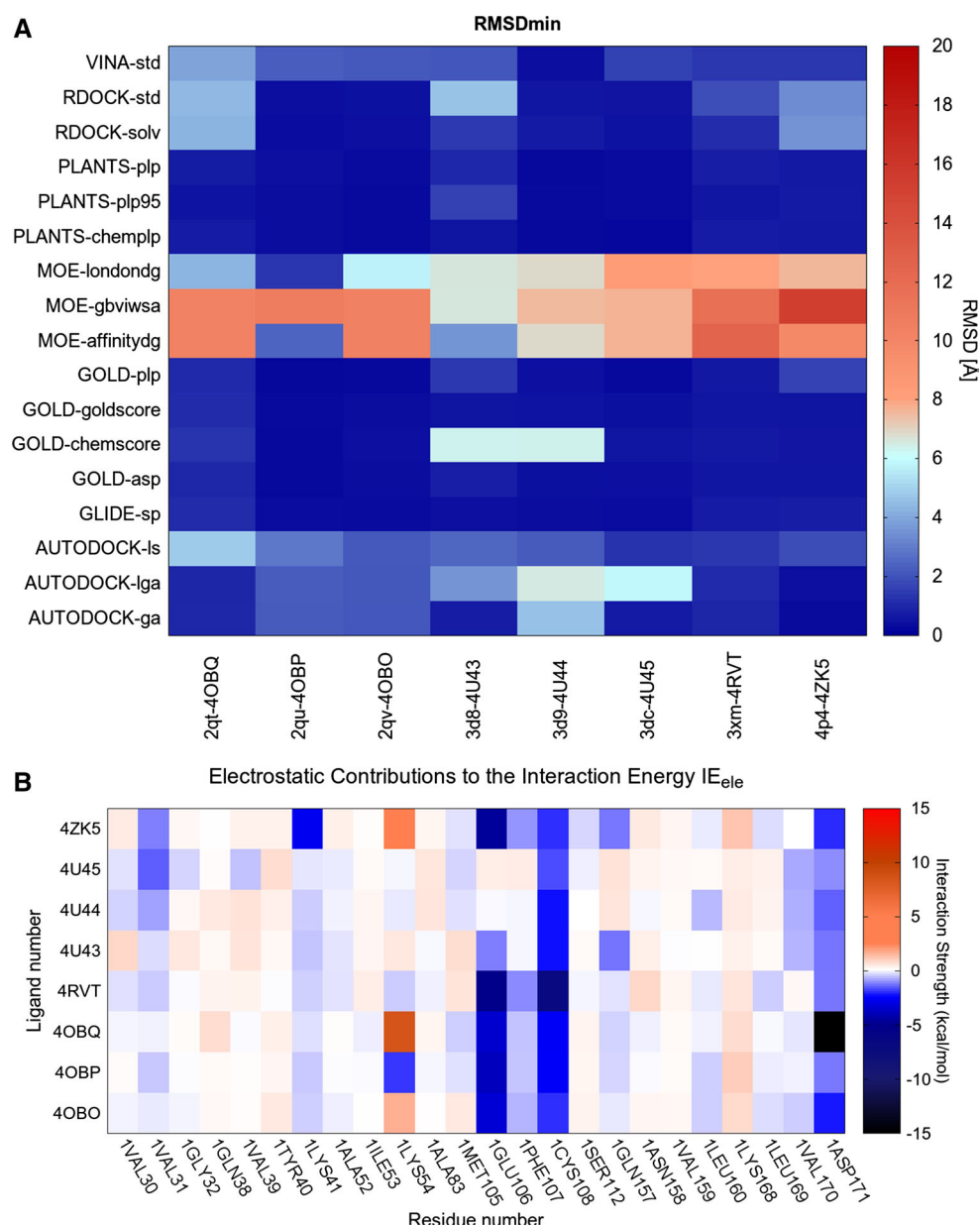
Hsp90, in this case have not been highlighted water molecules that may play a crucial role in the recognition crystallized ligands. As reported in Fig. 6a, several protocols showed good results with $RMSD_{min}$ values below 2 Å. In particular, Gold and Plants software were able to reproduce the crystal pose in the majority of cases except when Gold was coupled with chemscore function. Following these preliminary information, we selected “Gold-goldscore” and “Plants-plp” as best docking/scoring combinations.

Unfortunately, in this case the chemical variability of the 30 ligands of the test set didn't give us the opportunity to cluster them according to chemical similarity to the ligands of the training set, as in the case of Hsp90.

Therefore, we adopted a different strategy to select MAP4K4 structures for docking: in particular, we took into consideration the interaction network of co-crystallized ligands in the PDB complexes, and selected those structures that conserved the same pattern for the docked poses of the test ligands. We used EEF to estimate the residues mainly involved in electrostatic interactions with the ligands.

Docking calculation and best pose selection The EEF of the MAP4K4 complexes suggested E106 and C108 as key residues in ligand binding; in fact, those residues are involved in strong electrostatic interactions in almost all ligand of the training set (Fig. 6b).

Fig. 6 a Self-docking Benchmark results obtained with DockBench on 8 complexes containing MAP4K4. The minimum RMSD values ($RMSD_{min}$) returned by the tested docking protocol (y-values) for the considered X-ray structures (x-values) for the 20 poses generated for each protocol considered. Values are color coded, blue spots identify the best obtained results. **b** Electrostatic Energy Fingerprints representing per-residue electrostatic contribution to interaction energy. This term was calculated for the eight training set complexes subjected to the benchmark. The interaction strength is coded in the heatmap using a red to blue palette going from a highly positive to a deeply negative potential. The calculation was performed for the most relevant residues for the binding. The blue bars corresponding to E106 and C108 highlight the relevance of this residues



From this, we decided to pick the structure with lowest crystallographic resolution (PDB ID 4OBO) presenting the P-loop in a “closed” conformation. Among all docking/scoring combinations, Gold-goldscore was selected as docking protocol due to its good performance in reproducing the 4OBO ligand pose, as indicated by the corresponding RMSD_{\min} value in Fig. 6. The Virtual Screening Tool of DockBench was used to dock the 30 ligands of the test set. Using this strategy, it was possible to select a pose showing interactions with E106 or C108 for the following ligands: MAP01, MAP02, MAP03, MAP04, MAP08, MAP09, MAP14, MAP15, MAP16, MAP18, MAP19, MAP20, MAP21, MAP23, MAP26, MAP27, MAP28, MAP32 (Figure SI2).

For the remaining ligands of the test set, alternative selection strategies have been used in the selection of both MAP4K4 crystallographic structures and docking/scoring protocols. The first important alternative was to change the protein structure in which the P-loop was in an “open” conformation and the crystallographic structure coded as 4U44 was selected as the best compromise between its crystallographic resolution and its DockBench performance.

Moreover, Plants-plp combination was selected as docking/scoring protocol adopted for the 12 remaining ligands (see Fig. 6) and acceptable poses interacting with either E106 or C108 were selected for MAP05, MAP06, MAP07, MAP11, MAP22, MAP25, MAP29, MAP30, MAP31 as shown in Figure SI3. At the end, were only three exceptions: MAP12, MAP13 and MAP17. Since those ligands are voluminous, for these three ligands we chose the MAP4K4 crystallographic structure coded as 4ZK5, which performed well in the benchmark and whose co-crystallized ligand is the bulkiest among the training set (Figure SI4). In this specific case, we carried out the docking simulation using Gold-goldscore combination. Unfortunately, even with these changes, we were not able to find ligand poses directly interacting with E106 and C108 and, consequently, we decided to select the best poses by visual inspection.

Results The superposition of the 30 predicted complexes on the corresponding X-ray crystal structures is reported in Fig. 7. In general, the proposed workflow has shown encouraging results: the pose of several ligands were appropriately predicted but, understandably, there are a certain number of exceptions. In particular, 14 ligands were predicted with a RMSD below 2 Å and notably 11 of them below 1.5 Å. These values fall below the resolution of the crystal structures, which range from 1.59 to 3.04 Å. 4 ligands were in the range between 2 and 3 Å, whereas 12 showed a RMSD bigger than 3 Å. However, four ligands were poorly predicted (with an RMSD values >8Å). In

particular, the poses of the three ligands containing the dehydro-oxepin ring were completely wrong. The poor predictions are mainly due to the erroneous pose selection performed by visual inspection. In fact, a retrospective analysis of the docking result revealed the presence of a native like poses in the ensemble of the generated conformations. Not surprisingly, a subset of small ligands with molecular weight lower than 300 Da (MAP04, MAP20, MAP22, MAP26, MAP29, MAP30, MAP31) resulted in inaccurate poses confirming the difficulties of docking protocols with fragments in particular when docked in wide binding side and in absence of a clear shape complementarity between the ligand and the docking site. In addition, the experimental structure of four of them revealed the presence of molecules of water stabilizing their conformation (MAP04, MAP20, MAP22, MAP29).

Stage 1 ranking prediction step

The aim of MAP4K4 Stage 1-“Ranking prediction” phase was to rank the affinities of 18 of the 30 compounds docked in the previous phase. The workflow used for the ranking prediction is reported in Fig. 5 (on the right).

Scoring workflow The selected poses for MAP01, MAP02, MAP03, MAP04, MAP05, MAP06, MAP07, MAP08, MAP09, MAP11, MAP12, MAP13, MAP14, MAP15, MAP16, MAP17, MAP18, and MAP19 were rescored with MOE using GBVI/WSA method, in order to have a homogeneous scoring method.

Scoring results Also in this case and as expected, the scoring strategy showed its ineffectiveness in the ability to correctly rank ligands in terms of their binding affinities (Table 2). Pearson and Kendall coefficients values (0.46 and 0.32, respectively) show a modest positive correlation between affinities and GBVI/WSA scores.

Stage 2 ranking prediction step

Also in this case, the stage 2 of the D3R Grand Challenge 2015 was characterized by the release, from the organizers, of the MAP4K4 crystallographic structures used as test set in the pose prediction phase of stage 1. As in the ranking prediction phase of stage 1, the aim of this stage was the ranking of the same 18 ligands but taking into account the additional available crystallographic information.

Scoring workflow With the release of the new 30 MAP4K4 crystallographic structures, we re-performed to DockBench analysis (see Figure SI5). Also in this case, for each of the 18 compounds which were to be analyzed we selected the pose corresponding to the best value of

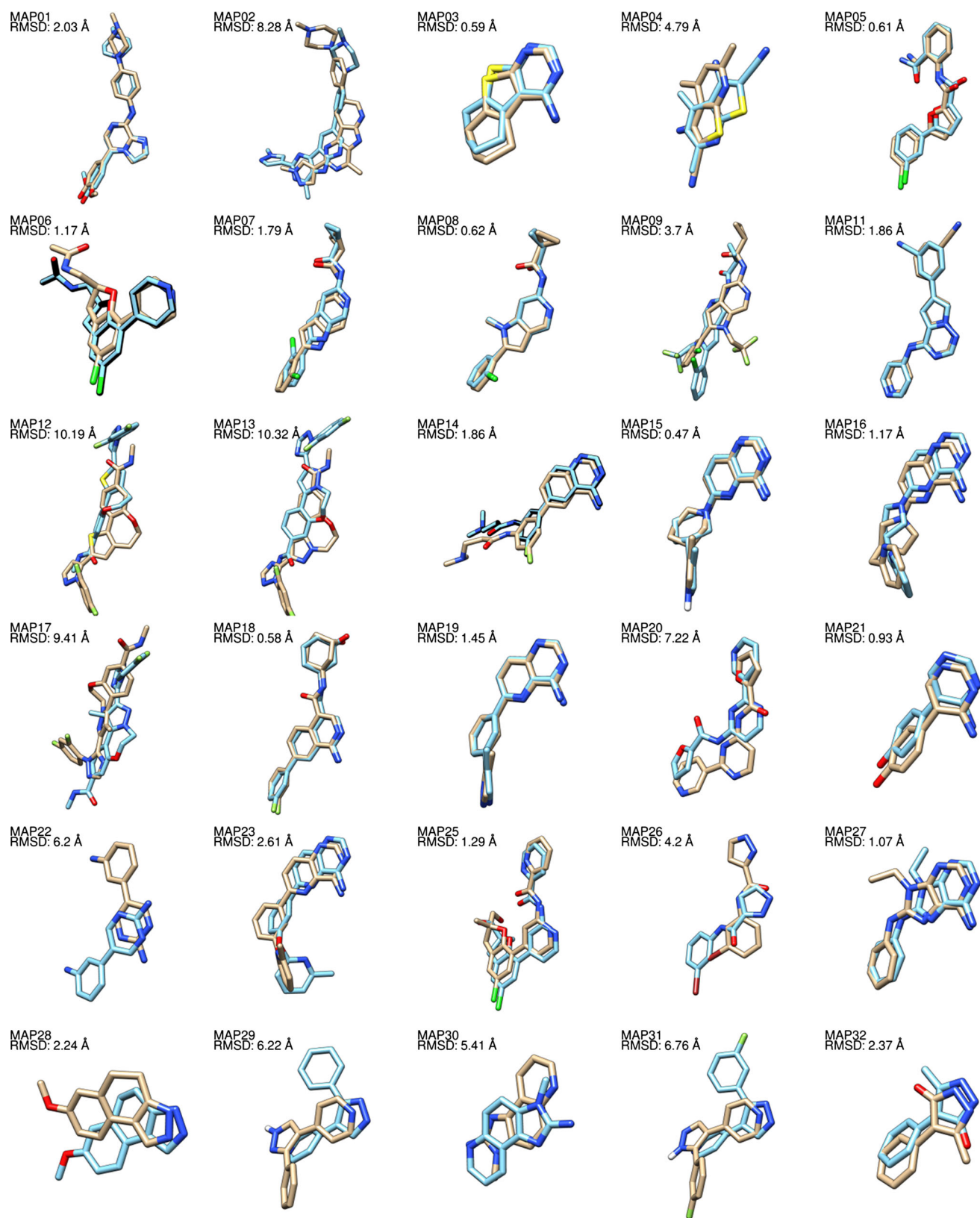


Fig. 7 Superposition of the predicted poses (*light blue*) on the experimental ones (*tan*). RMSD values were calculated on the heavy atoms

RMSD_{min} obtained in the benchmark (see Table 5). Finally, the complexes were rescored, and sorted, using MOE dock_pKi scoring function.

Scoring stage results Also in this case, as expected, the ranking performance in the second stage was even less accurate than that obtained in the first stage. As reported in Table 2, Spearman's rank coefficient (0.01) showed absence of correlation between affinities and dock_pKi scores, and, even worse, the Kendall coefficient (−0.02) showed a tendency to negative correlation.

Conclusions and consideration

Our sincere feeling is that D3R Grand Challenge represented an important moment of scientific and methodological reflection regarding the real robustness of docking/scoring methodologies currently available to our scientific community. Molecular docking is certainly one of the most popular and used tools in computational medicinal chemistry and beyond. For this reason, we believe that our community must pay particular attention to point out what are the intrinsic limitations of this tool and to appropriately describe the best practice for its correct use.

In this contest, we could evaluate the predictive ability of a docking selection tool recently developed in our laboratory and called DockBench. Considering the peculiarity of the DockBench tool in facilitating the prediction of the ligand poses, we decided to concentrate our efforts in determining the best docking method able to reproduce the most accurate pose geometries.

The take home message learned from the GC2015 is that an accurate selection of both the docking protocol and protein conformation may lead in remarkable improvement of the prediction. In addition, the differences emerged in the accuracy between the two targets reveal two interesting

points. First, when more data is already available as in the case of Hsp90 of which a notable number of complexes are available in the PDB, lead to better results if the similarity of between the ligand is taken into account. In particular, is not always straightforward the definition of similarity in this context and the selection of which kind of similarity can be the more appropriated (e.g. fingerprint similarity, shape similarity, substructure matching, etc.). The second point is that the role of the water molecule that improved the quality of the ligand-Hsp90 prediction. The significance of this two points are convincing us to introduce these aspects in our software also considering that the automation of these tasks into the docking pipeline would reduce the time needed to the user.

Even if the overall performance of DockBench is encouraging, from this assessment have emerged still delicate issues which limit the performance of docking/scoring algorithms and, consequently, their positive impact in the design of new drugs. Some of these are briefly summarized below:

- with the increasing number of docking programs (docking/scoring combinations), it becomes progressively more complex and risky to determine a priori which of these will be more accurate in reproducing a realistic poses of a ligand in its binding cavity;
- With the increasing number of crystal structures available in the PDB for a single protein, it becomes increasingly hazardous to determine a priori which crystallographic structure will be more appropriate to use to obtain a realistic pose of a ligand in its binding cavity;
- Nowadays, it is clear the crucial role of the water molecules, eventually present in the binding cavity, in determining the performance of the docking algorithms;
- Scoring functions are very often useless in realistically ranking a set of ligands.

As this D3R Grand Challenge has demonstrated, each docking run can be considered a singularity in a mathematical sense, or rather, a point in which a function is undefined. In fact, considering the degree of theoretical simplification of the problem we are dealing with docking and the large number of variables that define the problem itself, it is extremely difficult to determine a priori the degree of accuracy of the solution of our problem (realistic pose).

To paraphrase Albert Einstein, our take-home message may be summarized as follows: “Docking should be made as simple as possible, but not simpler.”

Acknowledgments The computational work coordinated by S.M. has been supported with financial support from the University of Padova,

Table 5 Combination of docking protocol, the PDB ID of protein conformation used for each cluster identify in the rank prediction stage 2 (Hsp90)

Complex	Protocol	RMSDmin
lig-MAP01	Plants-chemplp	0.49
lig-MAP02	Gold-asp	0.31
lig-MAP03	Gold-goldscore	0.22
lig-MAP04	Gold-chemscore	0.28
lig-MAP05	Plants-plp95	0.26
lig-MAP06	Gold-plp	0.19
lig-MAP07	Gold-plp	0.27
lig-MAP08	Gold-chemscore	0.44
lig-MAP09	Gold-asp	0.19

Italy. MMS lab is also very grateful to Chemical Computing Group, OpenEye and Acellera for the scientific and technical partnership. S.M. participates in the European COST Action CM1207 (GLISTEN). The work of M.S. has been supported by University of Padova, Italy (UNIPD, Progetto Giovani Studiosi 2013: Protocol number 79122). Finally, MMS lab is extremely grateful to the organizers of the D3R Grand Challenge for the perfect organization of the event.

Compliance with ethical standards

Conflicts of interest The authors declare no conflict of interest.

References

1. Directory of in silico Drug Design tools—Docking. http://www.click2drug.org/directory_Docking.html. Accessed 25 May 2016
2. Docking (molecular)—Wikipedia. [https://en.wikipedia.org/wiki/Docking_\(molecular\)](https://en.wikipedia.org/wiki/Docking_(molecular)). Accessed 25 May 2016
3. Cuzzolin A, Sturlese M, Malvacio I, Ciancetta A, Moro S (2015) DockBench: an integrated informatic platform bridging the gap between the robust validation of docking protocols and virtual screening simulations. *Molecules* 20:9977–9993. doi:10.3390/molecules20069977
4. Solit DB, Rosen N (2006) Hsp90: a novel target for cancer therapy. *Curr Top Med Chem* 6:1205–1214. doi:10.2174/156802606777812068
5. Virbasius JV, Czech MP (2016) Map4k4 signaling nodes in metabolic and cardiovascular diseases. *Trends Endocrinol Metab*. doi:10.1016/j.tem.2016.04.006
6. Harvey MJ, Giupponi G, Fabritiis GD (2009) ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J Chem Theory Comput* 5:1632–1639. doi:10.1021/ct9000685
7. Masciocchi J, Frau G, Fanton M, Sturlese M, Floris M, Pireddu L, Palla P, Cedrati F, Rodriguez-Tomé P, Moro S (2009) MMsINC: a large-scale cheminformatics database. *Nucleic Acids Res* 37:D284–D290. doi:10.1093/nar/gkn727
8. Sadowski J, Gasteiger J, Klebe G (1994) Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J Chem Inf Comput Sci* 34:1000–1008
9. Labute P (2009) Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* 75:187–205. doi:10.1002/prot.22234
10. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* 17:490–519. doi:10.1002/(SICI)1096-987X(199604)17:5/6<490:AID-JCC1>3.0.CO;2-P
11. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242. doi:10.1093/nar/28.1.235
12. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791. doi:10.1002/jcc.21256
13. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461. doi:10.1002/jcc.21334
14. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749. doi:10.1021/jm0306430
15. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 47:1750–1759. doi:10.1021/jm030644s
16. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. *Proteins* 52:609–623. doi:10.1002/prot.10465
17. Chemical Computing Group Inc. (2015) Molecular Operating Environment (MOE), 2014.09, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7
18. Korb O, Stützle T, Exner TE (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model* 49:84–96. doi:10.1021/ci800298z
19. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, Barril X, Hubbard RE, Morley SD (2014) rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol* 10:e1003571. doi:10.1371/journal.pcbi.1003571
20. Corbeil CR, Williams CI, Labute P (2012) Variability in docking success rates due to dataset preparation. *J Comput Aided Mol Des* 26:775–786. doi:10.1007/s10822-012-9570-1
21. O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. *J Cheminform* 3:33. doi:10.1186/1758-2946-3-33
22. OpenEye Scientific Software Inc. (2016) OEChem. OpenEye Scientific Software Inc., Santa Fe
23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. doi:10.1002/jcc.20084
24. Davies M, Nowotka M, Papadatos G, Dedman N, Gaulton A, Atkinson F, Bellis L, Overington JP (2015) ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res* 43:W612–W620. doi:10.1093/nar/gkv352
25. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 25:247–260. doi:10.1016/j.jmgm.2005.12.005
26. Case D, Babin V, Berryman J, Betz R, Cai Q, Cerutti D, Cheatham Iii T, Darden T, Duke R, Gohlke H (2014) Amber14, version AMBER14; <http://ambermd.org/> (accessed Oct 2015). University of California, San Francisco
27. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65:712–725. doi:10.1002/prot.21123
28. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174. doi:10.1002/jcc.20035
29. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 97:10269–10280. doi:10.1021/j100142a004
30. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09,

- Revision B.01. <http://gaussian.com/>. Accessed Oct 2015. Gaussian, Inc.: Wallingford, CT
31. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577. doi:[10.1063/1.470117](https://doi.org/10.1063/1.470117)
 32. Sabbadin D, Ciancetta A, Moro S (2014) Bridging molecular docking to membrane molecular dynamics to investigate GPCR-ligand recognition: the human A2A adenosine receptor as a key study. *J Chem Inf Model* 54:169–183. doi:[10.1021/ci400532b](https://doi.org/10.1021/ci400532b)
 33. Williams T, Kelley C GnuPlot 4.5: an interactive plotting program, version 4.5; <http://gnuplot.info> (Accessed Oct 2015)
 34. Zehnder L, Bennett M, Meng J, Huang B, Ninkovic S, Wang F, Braganza J, Tatlock J, Jewell T, Zhou JZ, Burke B, Wang J, Maegley K, Mehta PP, Yin MJ, Gajiwala KS, Hickey MJ, Yamazaki S, Smith E, Kang P, Sistla A, Dovalsantos E, Gehring MR, Kania R, Wythes M, Kung PP (2011) Optimization of potent, selective, and orally bioavailable pyrrolidinopyrimidine-containing inhibitors of heat shock protein 90. Identification of development candidate 2-amino-4-{4-chloro-2-[2-(4-fluoro-1H-pyrazol-1-yl)ethoxy]-6-methylphenyl]-N-(2,2-difluoropropyl)-5,7-dihydro-6H-pyrrolo[3,4-d]pyrimidine-6-carboxamide. *J Med Chem* 54:3368–3385. doi:[10.1021/jm200128m](https://doi.org/10.1021/jm200128m)
 35. Roughley SD, Hubbard RE (2011) How well can fragments explore accessed chemical space? A case study from heat shock protein 90. *J Med Chem* 54:3989–4005. doi:[10.1021/jm200350g](https://doi.org/10.1021/jm200350g)
 36. Miura T, Fukami TA, Hasegawa K, Ono N, Suda A, Shindo H, Yoon DO, Kim SJ, Na YJ, Aoki Y, Shimma N, Tsukuda T, Shiratori Y (2011) Lead generation of heat shock protein 90 inhibitors by a combination of fragment-based approach, virtual screening, and structure-based drug design. *Bioorg Med Chem Lett* 21:5778–5783. doi:[10.1016/j.bmcl.2011.08.001](https://doi.org/10.1016/j.bmcl.2011.08.001)
 37. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. *J Med Chem* 51:3661–3680. doi:[10.1021/jm8000373](https://doi.org/10.1021/jm8000373)
 38. Bruncko M, Tahir SK, Song X, Chen J, Ding H, Huth JR, Jin S, Judge RA, Madar DJ, Park CH, Park CM, Petros AM, Tse C, Rosenberg SH, Elmore SW (2010) N-aryl-benzimidazolones as novel small molecule HSP90 inhibitors. *Bioorg Med Chem Lett* 20:7503–7506. doi:[10.1016/j.bmcl.2010.10.010](https://doi.org/10.1016/j.bmcl.2010.10.010)
 39. Li J, Shi F, Xiong B, He J (2014) 4LWI, crystal structure of the human Hsp90- α N-domain bound to the hsp90 inhibitor FJ6. doi:[10.2210/pdb4lwi/pdb](https://doi.org/10.2210/pdb4lwi/pdb)
 40. Brough PA, Barril X, Borgognoni J, Chene P, Davies NG, Davis B, Drysdale MJ, Dymock B, Eccles SA, Garcia-Echeverria C, Fromont C, Hayes A, Hubbard RE, Jordan AM, Jensen MR, Massey A, Merrett A, Padfield A, Parsons R, Radimerski T, Raynaud FI, Robertson A, Roughley SD, Schoepfer J, Simmonite H, Sharp SY, Surgenor A, Valenti M, Walls S, Webb P, Wood M, Workman P, Wright L (2009) Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. *J Med Chem* 52:4794–4809. doi:[10.1021/jm900357y](https://doi.org/10.1021/jm900357y)
 41. Murray CW, Carr MG, Callaghan O, Chessari G, Congreve M, Cowan S, Coyle JE, Downham R, Figueroa E, Frederickson M, Graham B, McMenemy R, O'Brien MA, Patel S, Phillips TR, Williams G, Woodhead AJ, Woolford AJ (2010) Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency. *J Med Chem* 53:5942–5955. doi:[10.1021/jm100059d](https://doi.org/10.1021/jm100059d)
 42. Kang YN, Stuckey JA (2016) 4YKR, heat shock protein 90 bound to CS302. doi:[10.2210/pdb4ykr/pdb](https://doi.org/10.2210/pdb4ykr/pdb)
 43. Kang YN, Stuckey JA (2016) 4YKY, heat shock protein 90 bound to CS319. doi:[10.2210/pdb4yky/pdb](https://doi.org/10.2210/pdb4yky/pdb)
 44. Crawford TD, Ndubaku CO, Chen H, Boggs JW, Bravo BJ, Delatorre K, Giannetti AM, Gould SE, Harris SF, Magnuson SR, McNamara E, Murray LJ, Nonomiya J, Sambroline A, Schmidt S, Smyczek T, Stanley M, Vitorino P, Wang L, West K, Wu P, Ye W (2014) Discovery of selective 4-Amino-pyridopyrimidine inhibitors of MAP4K4 using fragment-based lead identification and optimization. *J Med Chem* 57:3484–3493. doi:[10.1021/jm500155b](https://doi.org/10.1021/jm500155b)
 45. Schröder P, Förster T, Kleine S, Becker C, Richters A, Ziegler S, Rauh D, Kumar K, Waldmann H (2015) Neuritogenic militarinone-inspired 4-hydroxypyridones target the stress pathway kinase MAP4K4. *Angew Chem Int Ed Engl* 54:12398–12403. doi:[10.1002/anie.201501515](https://doi.org/10.1002/anie.201501515)
 46. Wang L, Stanley M, Boggs JW, Crawford TD, Bravo BJ, Giannetti AM, Harris SF, Magnuson SR, Nonomiya J, Schmidt S, Wu P, Ye W, Gould SE, Murray LJ, Ndubaku CO, Chen H (2014) Fragment-based identification and optimization of a class of potent pyrrolo[2,1-f][1,2,4]triazine MAP4K4 inhibitors. *Bioorg Med Chem Lett* 24:4546–4552. doi:[10.1016/j.bmcl.2014.07.071](https://doi.org/10.1016/j.bmcl.2014.07.071)
 47. Ndubaku CO, Crawford TD, Chen H, Boggs JW, Drobnick J, Harris SF, Jesudason R, McNamara E, Nonomiya J, Sambroline A, Schmidt S, Smyczek T, Vitorino P, Wang L, Wu P, Yeung S, Chen J, Chen K, Ding CZ, Wang T, Xu Z, Gould SE, Murray LJ, Ye W (2015) Structure-Based Design of GNE-495, a Potent and Selective MAP4K4 Inhibitor with Efficacy in Retinal Angiogenesis. *ACS Med Chem Lett* 6:913–918. doi:[10.1021/acsmchemlett.5b00174](https://doi.org/10.1021/acsmchemlett.5b00174)