Molecular docking performance evaluated on the D3R Grand Challenge 2015 drug-like ligand datasets

Edithe Selwa, Virginie y Martiny, Bogdan Iorga

Molecular docking performance evaluated on the D3R Grand Challenge 2015 drug-like ligand datasets

Edithe Selwa,[1] *Virginie Y. Martiny,*[1,2] *Bogdan I. Iorga*[1,*]

[1] Institut de Chimie des Substances Naturelles, CNRS UPR 2301, LabEx LERMIT, 91198 Gif-sur-Yvette, France; [2] Department of Nephrology and Dialysis, AP-HP, Tenon Hospital, INSERM UMR_S 1155, 75020 Paris, France

Corresponding Author

[*] Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

KEYWORDS: docking ; scoring function ; Gold ; Glide ; Autodock ; Heat Shock Protein 90 ; HSP90 ; Mitogen-Activated Protein Kinase Kinase Kinase Kinase 4 ; MAP4K4 ; D3R Grand Challenge 2015

ABSTRACT

The D3R Grand Challenge 2015 was focused on two protein targets: Heat Shock Protein 90 (HSP90) and Mitogen-Activated Protein Kinase Kinase Kinase Kinase 4 (MAP4K4). We used a protocol involving a preliminary analysis of the available data in PDB and PubChem BioAssay, and then a docking/scoring step using more computationally demanding parameters that were required to provide more reliable predictions. We could evidence that different docking software and scoring functions can behave differently on individual ligand datasets, and that the flexibility of specific binding site residues is a crucial element to provide good predictions.

INTRODUCTION

The blinded prediction challenges organized by the Drug Design Data Resource (D3R) represent unique occasions for our community to evaluate, in "blind" conditions, the current state of computer-aided drug discovery technology and the performance of the currently available tools and protocols, with a special emphasis on docking and scoring, through the interchange of high quality protein-ligand datasets and workflows.

The D3R Grand Challenge 2015 was focused on two protein targets (Figure 1): Heat Shock Protein 90 (HSP90) [1-12] and Mitogen-Activated Protein Kinase Kinase Kinase Kinase 4 (MAP4K4) [13-22].

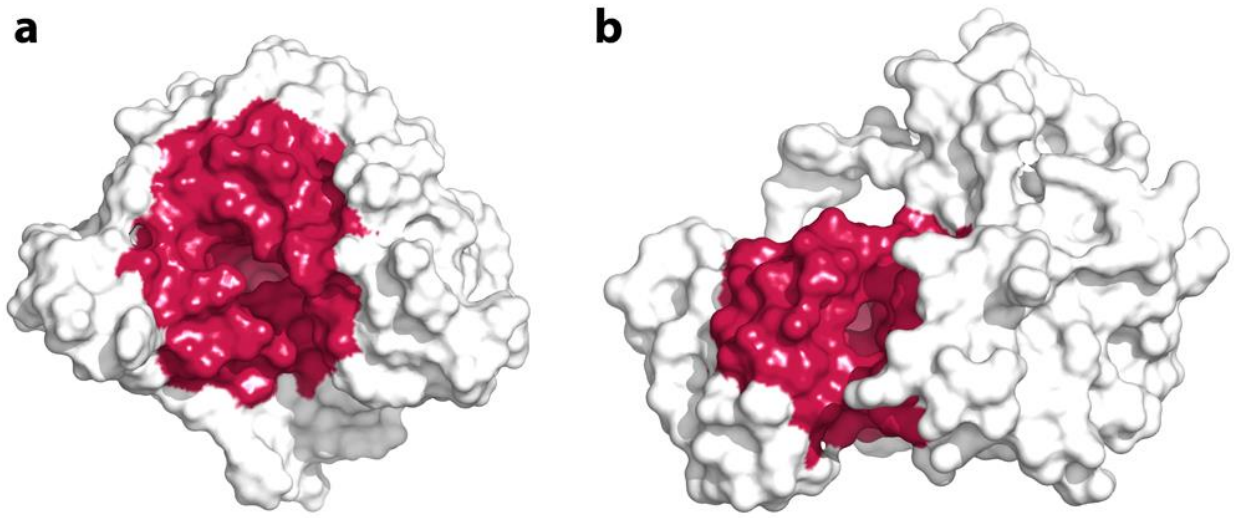


Figure 1. Surface representation of the two protein targets included in the D3R Grand Challenge 2015: (a) HSP90 (PDB code 2JJC) and (b) MAP4K4 (PDB code 4OBO). The binding site, as defined for our docking studies, is colored in red.

In Phase 1 the participants were asked to provide affinity predictions for 180 HSP90 ligands and pose prediction for 6 of them, as well as pose prediction for 30 MAP4K4 ligands and affinity predictions for 18 of them. In Phase 2 the participants were required to provide the same affinity predictions as in Phase 1, taking into account the additional structural data released at the end of Phase 1.

Figure 2 shows the chemical structures of compounds from HSP90 dataset for which both affinity predictions and pose predictions were required. Their chemical structures are representative for the whole dataset, containing three main classes: benzimidazolones [23], aminopyrimidines [24], and benzophenone-like.

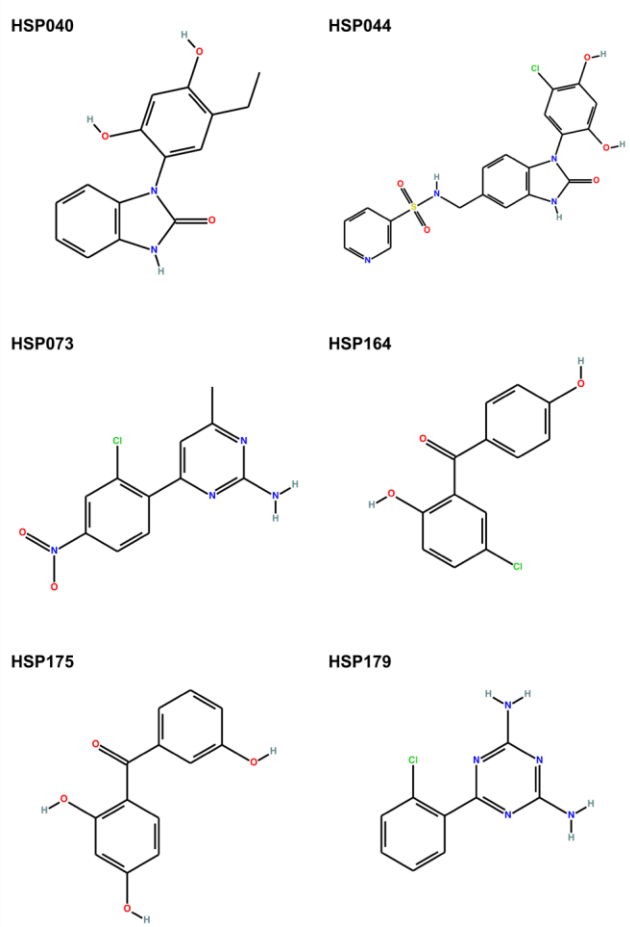


Figure 2. Chemical structures of the 6 HSP90 ligands included in Phase 1 for both pose prediction and ranking (the compound **HSP044** was ultimately retired from the pose prediction analysis). The entire HSP90 dataset, containing 180 ligands used for ranking prediction, is depicted in the Electronic Supplementary Material (Figure S1).

The MAP4K4 dataset for which both affinity predictions and pose predictions were required is shown in Figure 3. It contains a relatively diverse distribution of chemical groups, some of them already described in literature [25-28].

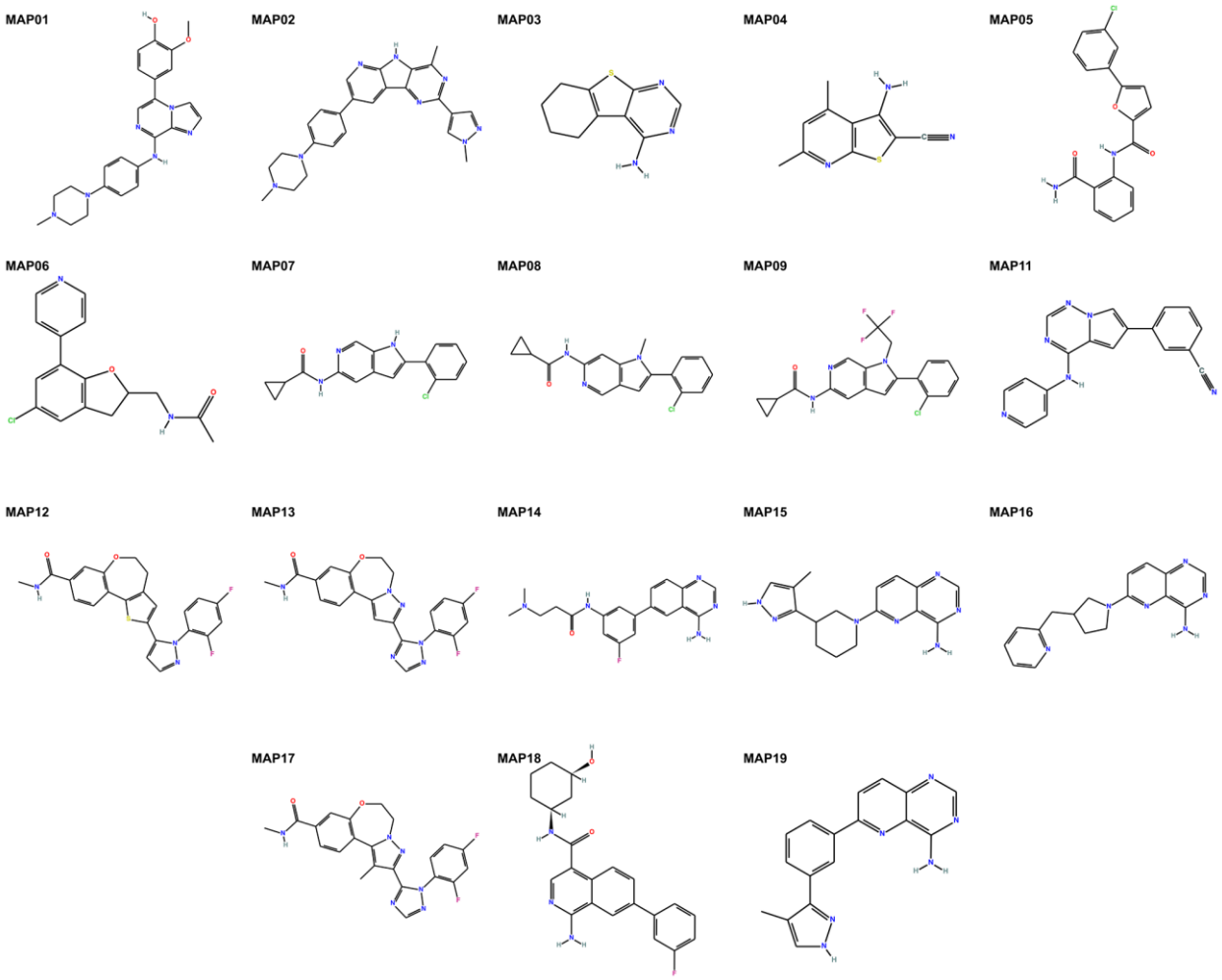


Figure 3. Chemical structures of the 18 MAP4K4 ligands included in Phase 1 for both pose prediction and ranking. The entire MAP4K4 dataset, containing 30 ligands used for pose prediction, is depicted in the Electronic Supplementary Material (Figure S2).

METHODS

Protein structures. We found 191 and 8 crystal structures available in the Protein Data Bank (PDB) [29] for human HSP90 and MAP4K4, respectively (see the Electronic Supplementary Material for the complete list). All ligands, ions and solvent molecules that were present were manually removed, then the structures were superimposed on the reference structures provided

by the D3R Grand Challenge 2015 organizers (PDB codes 2JJC and 4OBO, respectively) in order to conserve the same coordinate system through the whole process. Then, a common center of mass of all ligands from these structures was calculated, and defined as the center of the binding site (coordinates 1.0, 10.1, 24.9 for HSP90 and -16.1, -5.4, 30.4 for MAP4K4 in the coordinate system of the structure provided). The binding sites were considered as spheres with a 20 Å radius around these points. In the case of MAP4K4, the same eight ligands were used to generate a common pharmacophore using MOE v2013.0802 (http://www.chemcomp.com/), which was used later for filtering the docking poses. Missing residues in the structure 4AWQ (HSP90) were added using Modeller 9v12 [30]. Hydrogen atoms were added using Hermes, the graphical interface of Gold v5.2.2 [31] software, prior to docking.

Ligands. Ligand structures of the HSP90 and MAP4K4 datasets were provided in SMILES format and they were converted into three-dimensional MOL2 files using CORINA v3.60 (http://www.molecular-networks.com/). The stereochemistry of chiral centers was missing in several ligands from the MAP4K4 dataset and in this case all possible stereoisomers were considered. The protonation state for all compounds was adjusted at physiological pH using LigPrep (Schrödinger, http://www.schrodinger.com/).

Docking. In the preliminary analysis step, several docking software and scoring functions have been tested for their ability to reproduce the protein-ligand complexes from the evaluation datasets (11 representative structures of HSP90 and 8 structures of MAP4K4, see Electronic Supplementary Material for the complete list of PDB structures): Glide (Schrödinger, http://www.schrodinger.com/) with the standard precision (SP) scoring function, Gold [31] with the GoldScore, ChemScore, ChemPLP and ASP scoring functions, Vina [32] and Autodock [33]. As a result of preliminary analysis, Gold with the GoldScore scoring function was used in Phase

1 for evaluating both datasets, whereas Glide was used only for the HSP90 dataset. Default parameters were used in all cases for docking, except with Gold, where a search efficiency of 200% was used in order to better explore the conformational space, as well as a limited side-chain flexibility: Lys58 (HSP90) and Lys54 (MAP4K4) were flexible, and the two flipped conformations of Asn51 (HSP90) were considered. In Phase 2, the rescoring of the MAP4K4 complexes was carried out using Gold with the GoldScore scoring function.

Graphics. Chemical structures were depicted using CACTVS Chemoinformatics Toolkit v3.409 (Xemistry, http://www.xemistry.com/), images for protein structures were generated using PyMol 1.8.1 (Schrödinger, http://www.pymol.org/) and histograms were obtained using the R package (http://www.r-project.org).

RESULTS AND DISCUSSION

In this study we followed a general approach specially designed for predictions without prior knowledge of the results, which are the conditions generally encountered in real-life projects. Therefore, we took advantage of any structural and biochemical data publicly available for the target proteins and for the ligands, and we tried to avoid potential problems by taking into account all protein conformations described to date and the flexibility of key residues in the binding sites. This protocol involves a preliminary analysis of information available in literature (structural and enzymatic data), which is used for the evaluation of the best docking software and associated scoring function that are adapted for the system to be studied (protein targets and ligand datasets). This combination of docking software and scoring function is then used for the actual prediction. This approach proved to be highly successful during our participation to the

SAMPL3 (2011) [34], SAMPL4 (2013) [35] and CSAR (2014) [36] docking and virtual screening challenges. Noteworthy, as for our previous studies [34-36], we use normal docking (not virtual screening) parameters, and a search efficiency of 200%. Considering the size and conformational flexibility of some ligands included in the D3R Grand Challenge 2015 datasets, these parameters are key points required for providing reliable results, especially in the pose prediction step, and for a better conformational sampling of docking conformations.

Preliminary analysis

We started by identifying the structural data available for the two protein targets (HSP90 and MAP4K4) in the Protein Data Bank (PDB) [29] and PubChem BioAssay [37]. This information was further used to evaluate several different docking software and scoring functions and to identify those that are the most adapted for the given targets in positioning the ligand in the binding site and in scoring (or rescoring) the docking poses.

HSP90. A number of 191 crystal structures of human HSP90 were identified in the PDB, containing 225 unique ligands. The three-dimensional structure of protein in these structures is well conserved, with the exception of the fragment 99-129 which is very flexible [38] (Figure 4). According to the conformation of this fragment, we could cluster these structures in 11 groups (see the Electronic Supplementary Material for the complete list of structures included in each group). From each group a representative structure was chosen and the ensemble of these 11 structures was used for further calculations.

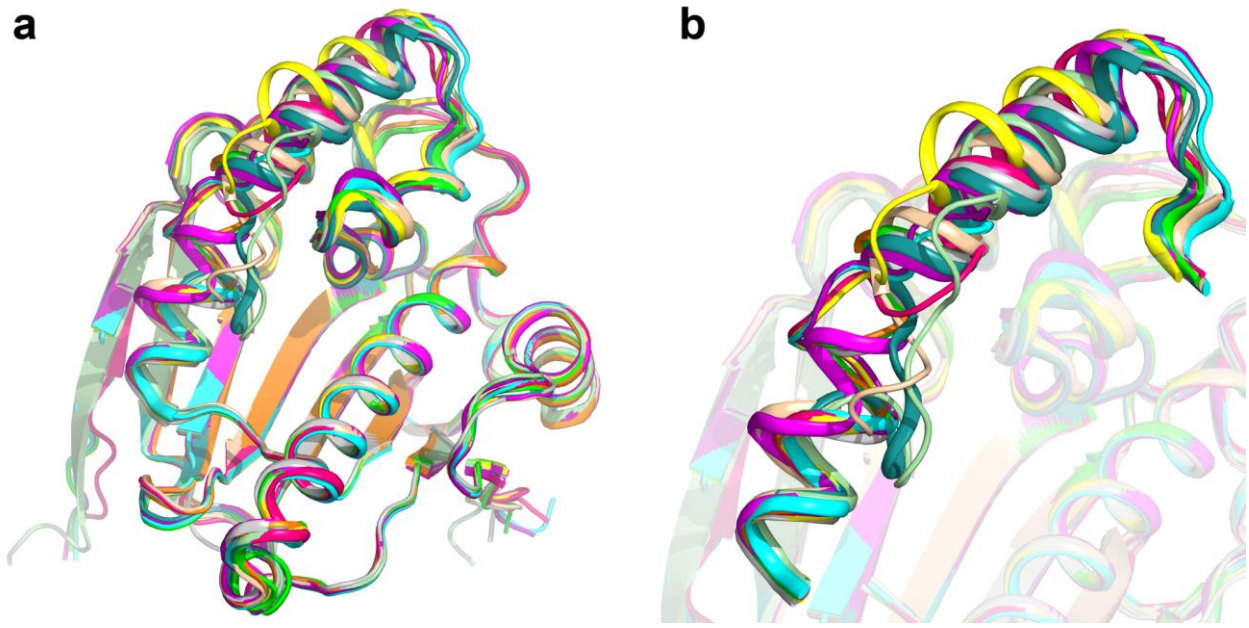


Figure 4. Representative 11 HSP90 PDB structures superimposed: a) general view, showing a very good global conservation of structural features; b) zoom on residues 99-129, highlighting the conformational flexibility of this fragment. The structures are represented as follows: 3BMY (green), 3VHA (cyan), 2YK9 (magenta), 3OWD (yellow), 4AWQ (wheat), 3WHA (gray), 3T10 (purple), 2QF6 (orange), 3K99 (pale green), 3T0H (deep teal), 4BQJ (pink).

On the other hand, we could find in PubChem BioAssay 740 compounds with enzymatic data for human HSP90. Among them, 670 compounds were active and 70 inactive. The structure comparison of the 225 unique ligands from the PDB structures and the 670 active compounds from the PubChem BioAssay afforded 50 structures that were common between the two datasets.

These 50 ligand structures, for which both binding modes and enzymatic data are available, were docked into the 11 representative HSP90 structures selected previously using several combinations of docking software and scoring functions: Glide with the standard precision (SP) scoring function, Gold with the GoldScore, ChemScore, ChemPLP and ASP scoring functions,

Vina and Autodock. RMSD values compared with the native ligands from the crystallographic structures were calculated for all docking poses. For each combination protein-ligand-(docking software)-(scoring function) we have considered the lowest RMSD value and the RMSD value of the best ranking pose, in order to evaluate the accuracy of docking and scoring. In these conditions, all docking software and scoring functions provided good results for 70% of the dataset, whereas Gold with GoldScore scoring function and to a lesser extent Glide performed better for the remaining structures from the dataset (Figure 5, left).

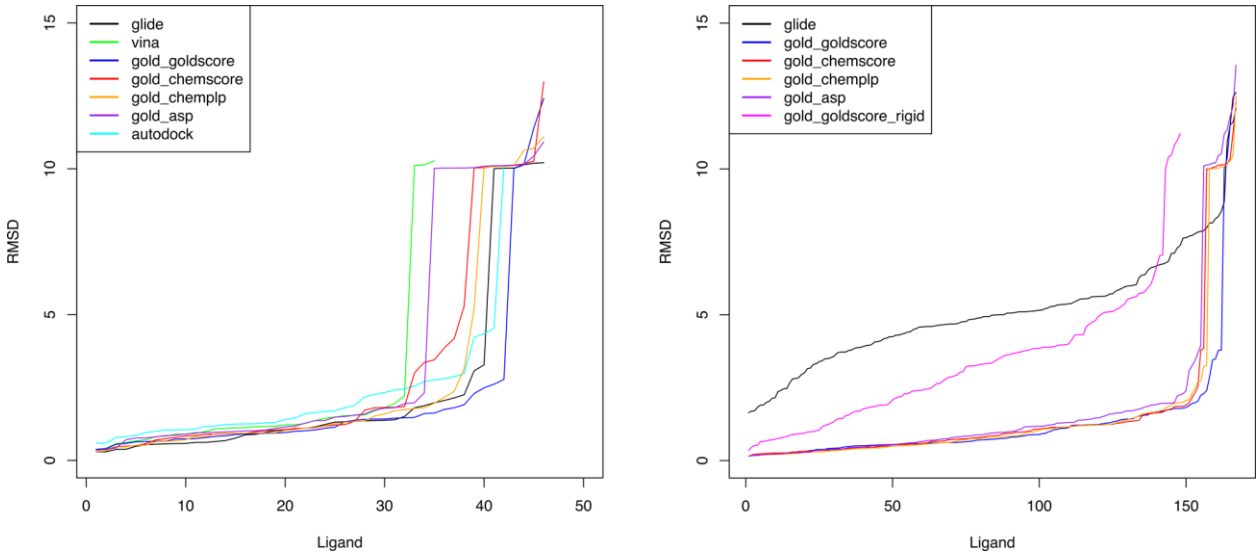


Figure 5. Performance evaluation of different combinations of docking software and scoring functions using the dataset of 50 HSP90 ligands common between PDB and PubChem BioAssay (left) and the D3R Grand Challenge 2015 HSP90 dataset containing 180 ligands (right).

MAP4K4. Eight X-ray structures of human MAP4K4 were available in the PDB, showing a good conservation of the three-dimensional structure with the exception of a region containing the P-loop, which is known to adopt either a "closed" or an "open" conformation (Figure 6).

Given the limited availability of structural data and the diversity of structures in the D3R Grand Challenge 2015 MAP4K4 dataset, we have generated a pharmacophore using MOE and the superposed ligands from the 8 cystal structures (see Figure S4 in the Electronic Supplementary Material for the structures of these ligands). This pharmacophore, which contained a single pharmacophoric point – a hydrogen bond acceptor able to interact with the backbone NH group of Cys108, was further used in Phase 1 to filter the docking poses.

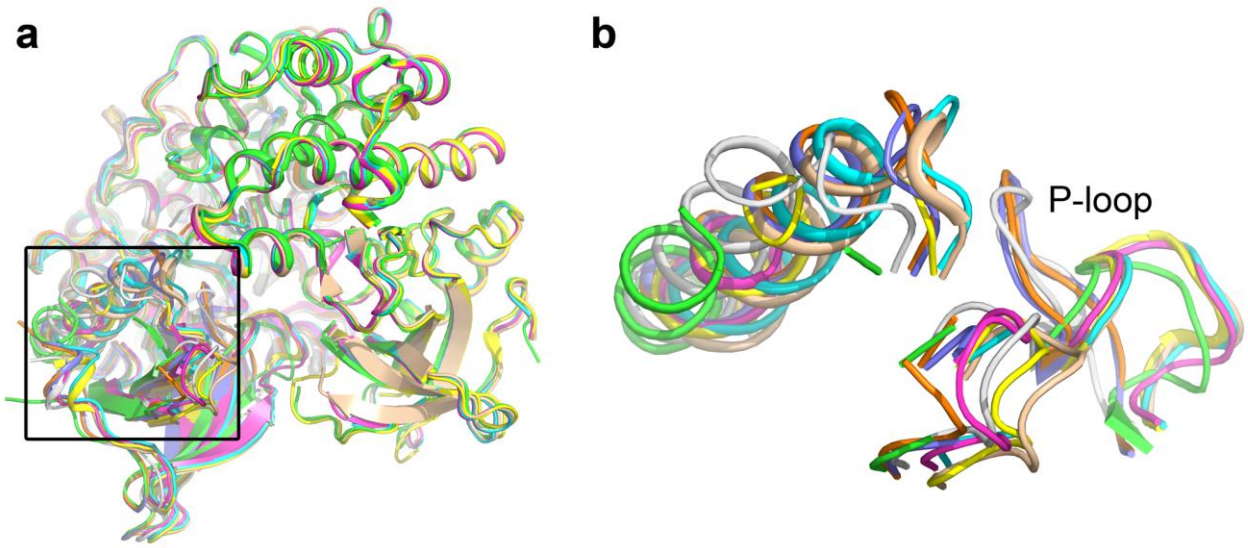


Figure 6. The 8 MAP4K4 PDB structures superimposed: a) general view, showing a very good global conservation of structural features; b) zoom on the flexible region, highlighting the conformational flexibility of the P-loop. The structures are represented as follows: 4OBO (green), 4OBP (cyan), 4OBQ (magenta), 4RVT (yellow), 4U43 (gray), 4U44 (purple), 4U45 (orange), 4ZK5 (wheat).

Phase 1

HSP90. The D3R Grand Challenge 2015 HSP90 dataset containing 180 ligands was docked on the 11 representative HSP90 structures using Glide with SP scoring function and Gold with the four scoring functions mentioned above, in the preliminary analysis step. As the 180 ligands can be easily classified according to their chemical structures (aminopyrimidines, benzimidazolones and benzophenone-like – see the Electronic Supplementary Material, page S9, for more details) and representative binding modes are known for all these three chemical moieties (Figure 7), we have used the common substructures between the three ligands from crystallographic structures (A13, MEX and 4EU) and the 180 ligands from the HSP90 dataset in order to compute RMSD values, using the same approach as in the Preliminary analysis step. For each ligand, we have selected the pose displaying the best RMSD (close to zero), regardless the protein structure on which docking was performed, then all the ligands were ranked based on their Goldscore.

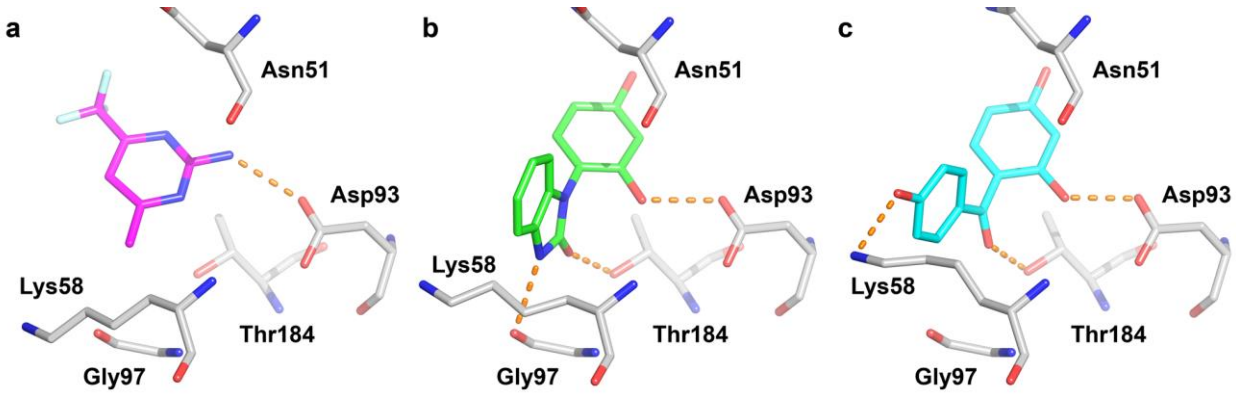


Figure 7. Representative interactions between the protein HSP90 and ligands A13 (a, PDB code 2QFO), MEX (b, PDB code 3OW6) and 4EU (c, PDB code 4YKY). Ligands are colored in magenta, green and cyan, respectively. Hydrogen atoms were omitted for clarity.

The results are plotted in Figure 5 (right), showing a very good behavior of Gold with all four scoring functions on more than 90% of the dataset, whereas Glide provides very deceiving

results on the whole dataset. It is noteworthy the big difference in behavior with Glide between the Preliminary analysis dataset and the D3R Grand Challenge 2015 dataset, showing the sensitivity of this software on the input ligand dataset. This difference is not observed for Gold, which produces equally good predictions on the testing and on the D3R Grand Challenge 2015 datasets. The results obtained with Gold and GoldScore scoring function were submitted as the first prediction and those obtained with Glide as the second prediction (the last one being expected to behave not very well), in order to evaluate the accuracy of the two extremes.

This difference between the docking programs might be due, at least in part, to the flexibility of the Lys58 residue in HSP90, which was taken into account with Gold and treated as rigid with Glide, Vina and Autodock. The flexibility of this residue seems to be of crucial importance during the docking process, since Gold with rigid side chain of Lys58 provides very bad RMSD values with the D3R Grand Challenge 2015 HSP90 dataset, comparable with those obtained for Glide (Figure 5, right).

In the pose prediction section of the challenge were included six HSP90 ligands (**HSP044**, **HSP044**, **HSP073**, **HSP164**, **HSP175** and **HSP179**), for which we have submitted the docking conformations corresponding to the best score from the calculations presented above. For four of them (**HSP044**, **HSP044**, **HSP175** and **HSP179**) we have also submitted a second conformation, which presented a significantly different binding mode. At the end of Phase 1 were released the crystallographic structures of HSP90 complexes containing the six ligands proposed for pose prediction. A comparison of our docking poses with the best score and the crystallographic conformations is provided in Figure 8. We had very good predictions for the first four ligands, with the exception of the pyridylsulfonyl group in **HSP044**, and a good overall ligand orientation, but some different interactions with the binding site for **HSP175** and **HSP179**. This

resulted into a mean RMSD of 1.48 Å for the best score conformations, and a mean RMSD of 1.20 Å for the lowest RMSD poses.

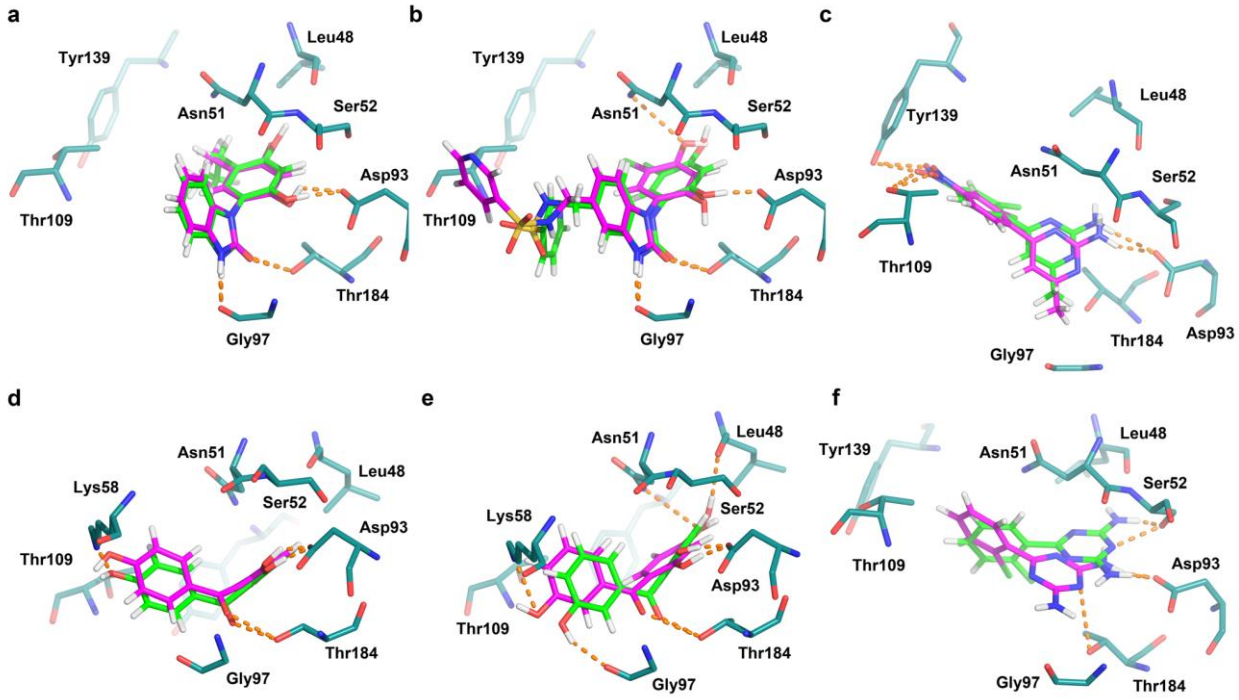


Figure 8. Comparison of our docking poses (green) with crystal structure conformations of HSP90 ligands (magenta) **HSP040** (a), **HSP044** (b), **HSP073** (c), **HSP164** (d), **HSP175** (e) and **HSP179** (f).

MAP4K4. The protocol used for MAP4K4 was identical with those used for HSP90 (docking and scoring), with the exception of an additional pharmacophore-based post-docking filtering step. All 8 MAP4K4 crystal structures available were used for docking, using Gold with the GoldScore scoring function. The analysis of the docked ligand poses was carried in two steps. The first filtering was done using the pharmacophore generated in a previous step, in order to retain those poses that establish the interactions that are essential for the biological activity. When several different poses were compatible with the pharmacophore, the pose(s) that establish the maximum number of favorable interactions with the rest of the binding site were selected for submission. The GoldScore value for the best pose of each ligand selected above was then

retrieved and used for ranking. The analysis of the Phase 1 submission showed that our prediction reproduced moderately well the experimental affinities data, with values of 0.46 and 0.33 for Pearson R and Kendall Tau, respectively, which ranks our Phase 1 MAPK4K scoring submission *ex-aequo* on the 3^{th}-4^{th} positions out of 76 submissions (see Figure S5 in the Electronic Supplementary Material).

For the pose prediction challenge we have submitted the docking pose corresponding to the best score for each ligand. The crystallographic structures of the 30 MAP4K4 ligands included in the pose prediction challenge that were released at the end of Phase 1 showed two binding patterns for which representative examples are shown in Figure 9. In the first case (representative for 11 ligands), the binding mode was already known and generally it was correctly predicted. In the second (representative for 17 ligands), a previously unknown binding mode is present and this one was generally not well predicted. There are also 2 ligands (**MAP04** and **MAP17**) that interact with Cys108 only through crystallographic water molecules and for which the docking poses were also incorrect. We can therefore assume that in the last two cases our analysis of docking results was biased by the limited availability of structural information regarding the interaction of this protein with different families of ligands. The direct consequence of this is the relatively high overall mean RMSD values of our predictions compared with the experimental coordinates (4.64 Å and 4.32 Å, for the best scoring poses and for lowest-RMSD poses, respectively, see Figure 10).

A tentative explanation for the good ranking prediction of the MAP4K4 ligands while the positioning of these ligands in the binding site is relatively poor might be related to the fact that the key interactions of the ligand with the binding site are reproduced in the predictions, but not with the same atoms as in the crystal structure. In these conditions, the energy of the protein-

ligand interaction is relatively well evaluated, in spite of important differences in binding geometries.

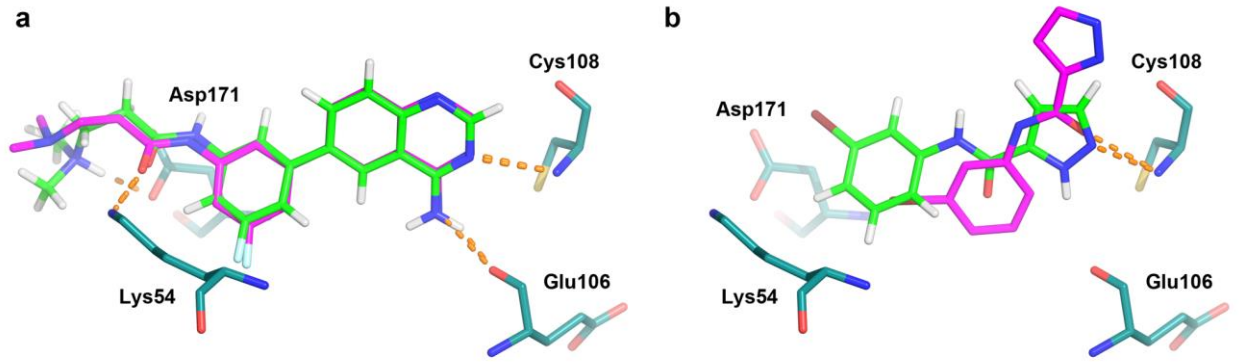


Figure 9. Comparison of our docking poses (green) with crystal structure conformations of representative MAP4K4 ligands (magenta) **MAP14** (a) and **MAP26** (b).

Phase 2

HSP90. As we have already taken into account the protein flexibility in the Preliminary analysis and Phase 1 steps, the release of the 6 new HSP90 crystal structures did not bring any new structural information, and therefore our submission for Phase 2 (ligand ranking prediction) was identical with the one from Phase 1.

MAP4K4. In Phase 2 we have only rescored the 30 protein-ligand MAP4K4 complexes that were released at the end of Phase 1, using Gold with GoldScore scoring function. All ligands were then ranked according to their best GoldScore value, from the highest to the lowest. The performance of scoring using crystal structures of the complexes (Phase 2) was reasonably good, with values of 0.45 and 0.34 for Pearson R and Kendall Tau, respectively, which are very similar with those obtained using docking poses (Phase 1).

Figure 10 shows a comparison of mean RMSD values for all pose prediction submissions of the D3R Grand Challenge 2015. Our submissions (for which the results were obtained in both cases using Gold with GoldScore scoring function) were ranked in the 7th position out of 42 for HSP90 and in the 5th position out of 27 for MAP4K4.

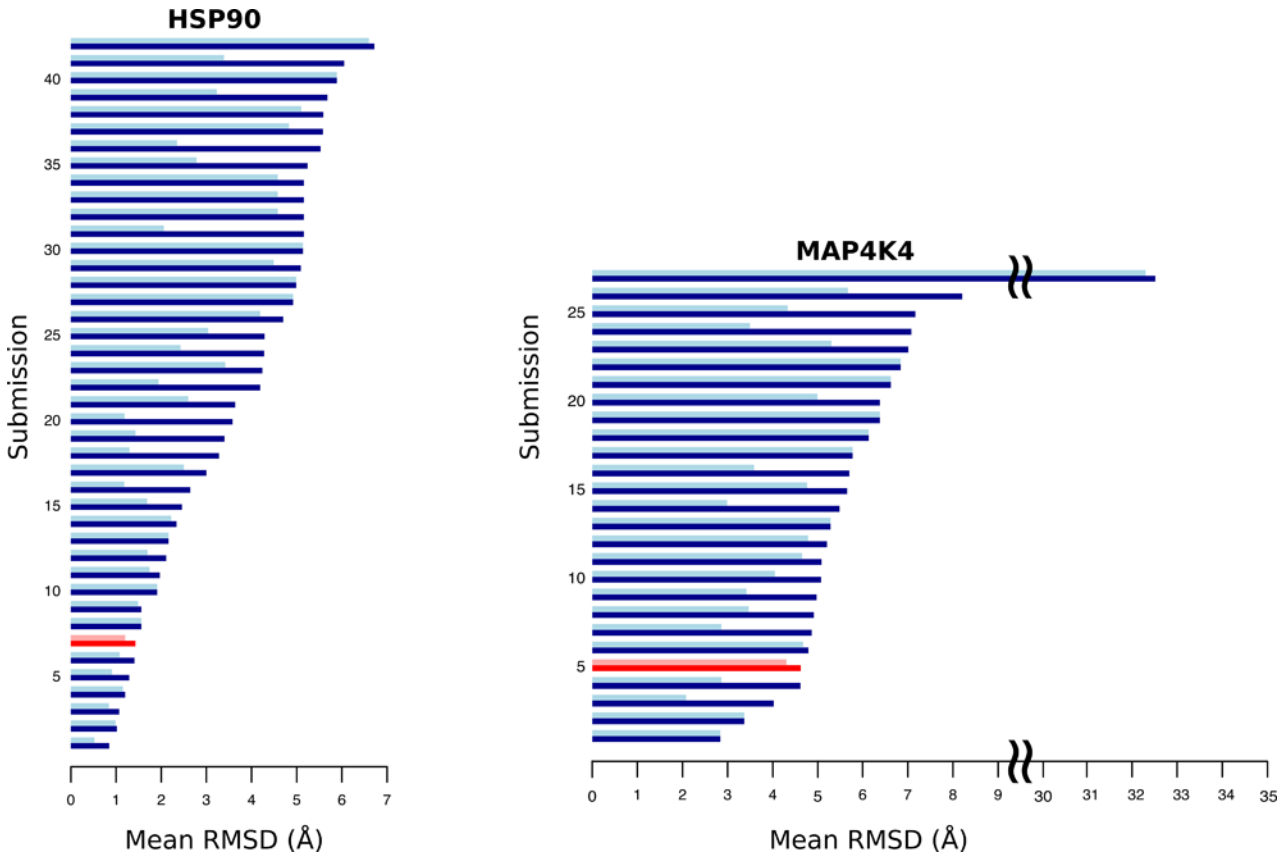


Figure 10. Mean RMSD of our pose prediction (red) for HSP90 (left) and MAP4K4 (right) compared to the other submissions (blue) from the D3R Grand Challenge 2015. The red and blue values represent the mean RMSD for the first ranked pose, whereas the light red and light blue values correspond to the mean RMSD of the lowest-RMSD pose. A pose prediction section was included only in the Phase 1 of the challenge.

An overview of the ligand scoring submissions from Phase 2 is presented in Figure 11 (see Figure S5 in the Electronic Supplementary Material for the corresponding Phase 1 results). Our HSP90 and MAP4K4 submissions prepared with Gold and GoldScore scoring function were ranked 29th out of 59 and 2nd out of 46, whereas, as expected, the HSP90 submission prepared with Glide was ranked on the last, 59th position.

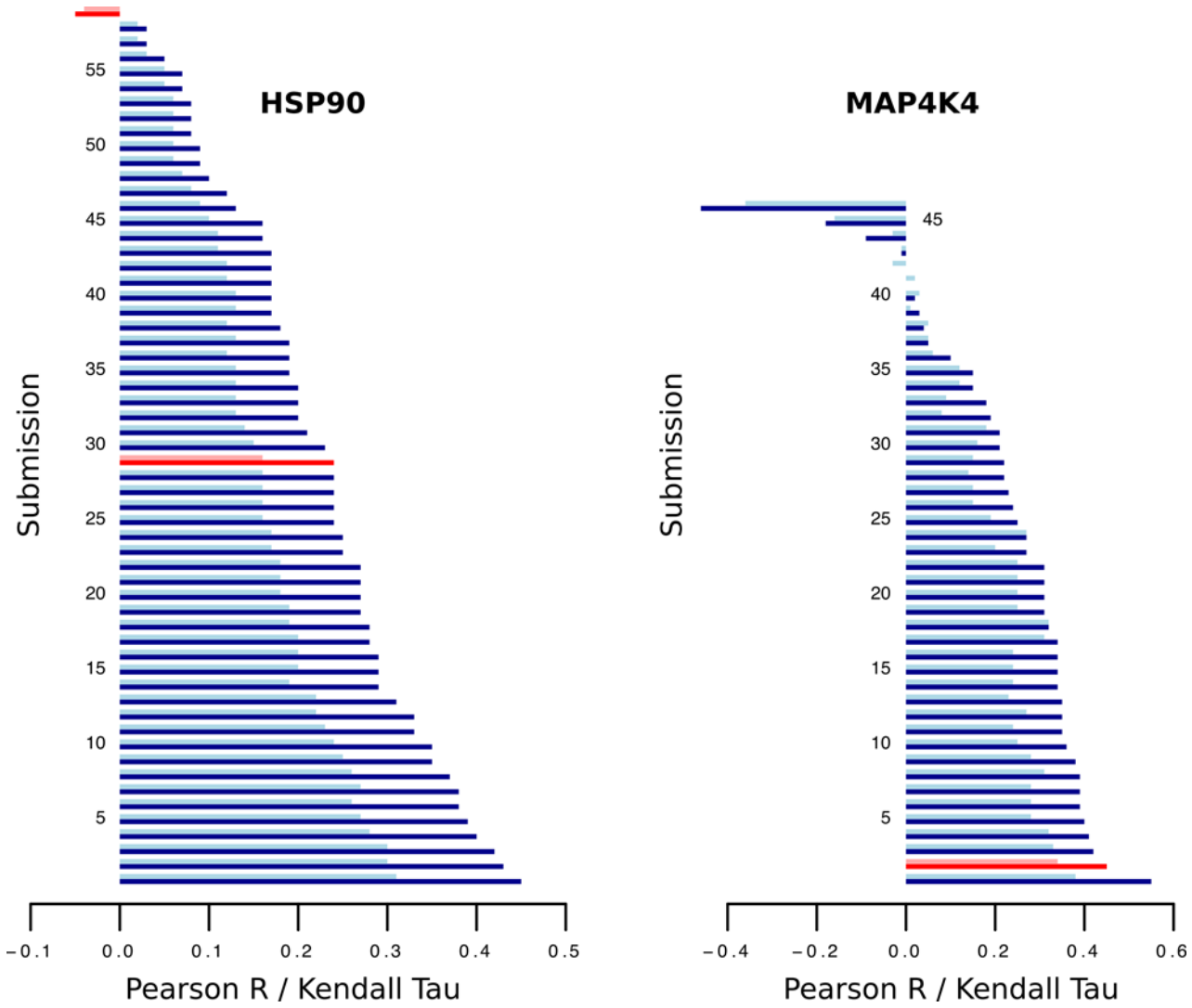


Figure 11. Performance of Phase 2 ranking submissions (Pearson R in blue and Kendall Tau in light blue) for the HSP90 (left) and MAP4K4 (right) D3R Grand Challenge 2015 datasets. The

best predictions are those with values closest to 1. Our submissions are colored in red and light red, respectively.

Looking retrospectively, the quality of our predictions, at least for the MAP4K4 target, suffered from the use of a pharmacophore filtering that was not fuzzy enough, therefore missing conformations that might have been correct. All the MAP4K4 ligands interacting with Cys108 in the protein-ligand complexes available at the time of the D3R Grand Challenge 2015, which were used for building the pharmacophore, had the hydrogen bond acceptor in exactly the same region of space, whereas the crystal structures released during the challenge showed a wider spatial distribution for the partners involved in this interaction.

We consider that the ranking of ligands can be greatly improved by using post-docking processing, especially by using free energy calculations. In addition to a more reliable evaluation of the affinity between protein and ligand, this post-docking processing can also identify and correct small deviations in the docking poses, take into account the influence of water molecules and of fully flexible protein, etc. Although these techniques require more important computational resources, they are fully justified by the potential benefits in providing better predictions and we intend to pursue our future research in this direction.

CONCLUSIONS

In this study we used a protocol involving a preliminary analysis of the available data in PDB and PubChem BioAssay, and then a docking/scoring step using more computationally demanding parameters that were required to provide more reliable predictions. We could evidence that different docking software and scoring functions can behave differently on

individual ligand datasets, and that the flexibility of specific binding site residues is a crucial element to provide good predictions.

Electronic Supplementary Material. The Electronic Supplementary Material contains the chemical structures of different datasets mentioned in the main text, as well as the performance of ligand scoring predictions from Phase 1 submissions.

Acknowledgments. Fellowships from the LabEx LERMIT (E.S.) and the ERC Advanced Grant OSAI (V.M.) are gratefully acknowledged. B.I.I.'s laboratory is member of the Laboratory of Excellence in Research on Medication and Innovative Therapeutics (LabEx LERMIT), supported by a grant from French National Research Agency (ANR-10-LABX-33). We would like to thank the D3R Grand Challenge 2015 organizers for their help throughout the submission and evaluation process.

REFERENCES

1. Holzbeierlein JM, Windsperger A, Vielhauer G (2010) Hsp90: a drug target? Current oncology reports 12 (2):95-101. doi:10.1007/s11912-010-0086-3

2. Whitesell L, Santagata S, Lin NU (2012) Inhibiting HSP90 to treat cancer: a strategy in evolution. Current molecular medicine 12 (9):1108-1124. doi:10.2174/156652412803306657

3. Barrott JJ, Haystead TA (2013) Hsp90, an unlikely ally in the war on cancer. The FEBS journal 280 (6):1381-1396. doi:10.1111/febs.12147

4. Jackson SE (2013) Hsp90: structure and function. Topics in current chemistry 328:155-240. doi:10.1007/128_2012_356

5. Li J, Buchner J (2013) Structure, function and regulation of the hsp90 machinery. Biomedical journal 36 (3):106-117. doi:10.4103/2319-4170.113230

6. Patki JM, Pawar SS (2013) HSP90: chaperone-me-not. Pathology oncology research : POR 19 (4):631-640. doi:10.1007/s12253-013-9675-4

7. Karagoz GE, Rudiger SG (2015) Hsp90 interaction with clients. Trends in biochemical sciences 40 (2):117-125. doi:10.1016/j.tibs.2014.12.002

8. Khurana N, Bhattacharyya S (2015) Hsp90, the concertmaster: tuning transcription. Frontiers in oncology 5:100. doi:10.3389/fonc.2015.00100

9. Kumalo HM, Bhakat S, Soliman ME (2015) Heat-shock protein 90 (Hsp90) as anticancer target for drug discovery: an ample computational perspective. Chemical biology & drug design 86 (5):1131-1160. doi:10.1111/cbdd.12582

10. Mayer MP, Le Breton L (2015) Hsp90: breaking the symmetry. Molecular cell 58 (1):8-20. doi:10.1016/j.molcel.2015.02.022

11. Pennisi R, Ascenzi P, di Masi A (2015) Hsp90: A New Player in DNA Repair? Biomolecules 5 (4):2589-2618. doi:10.3390/biom5042589

12. Pearl LH (2016) The HSP90 molecular chaperone-an enigmatic ATPase. Biopolymers 105 (8):594-607. doi:10.1002/bip.22835

13. Danai LV, Guilherme A, Guntur KV, Straubhaar J, Nicoloro SM, Czech MP (2013) Map4k4 suppresses Srebp-1 and adipocyte lipogenesis independent of JNK signaling. Journal of lipid research 54 (10):2697-2707. doi:10.1194/jlr.M038802

14. Haas DA, Bala K, Busche G, Weidner-Glunde M, Santag S, Kati S, Gramolelli S, Damas M, Dittrich-Breiholz O, Kracht M, Ruckert J, Varga Z, Keri G, Schulz TF (2013) The inflammatory kinase MAP4K4 promotes reactivation of Kaposi's sarcoma herpesvirus and enhances the invasiveness of infected endothelial cells. PLoS pathogens 9 (11):e1003737. doi:10.1371/journal.ppat.1003737

15. Wang M, Amano SU, Flach RJ, Chawla A, Aouadi M, Czech MP (2013) Identification of Map4k4 as a novel suppressor of skeletal muscle differentiation. Molecular and cellular biology 33 (4):678-687. doi:10.1128/mcb.00618-12

16. Ammirati M, Bagley SW, Bhattacharya SK, Buckbinder L, Carlo AA, Conrad R, Cortes C, Dow RL, Dowling MS, El-Kattan A, Ford K, Guimaraes CR, Hepworth D, Jiao W, LaPerle J, Liu S, Londregan A, Loria PM, Mathiowetz AM, Munchhof M, Orr ST, Petersen DN, Price DA, Skoura A, Smith AC, Wang J (2015) Discovery of an in Vivo Tool to Establish Proof-of-Concept for MAP4K4-Based Antidiabetic Treatment. ACS medicinal chemistry letters 6 (11):1128-1133. doi:10.1021/acsmedchemlett.5b00215

17. Roth Flach RJ, Skoura A, Matevossian A, Danai LV, Zheng W, Cortes C, Bhattacharya SK, Aouadi M, Hagan N, Yawe JC, Vangala P, Menendez LG, Cooper MP, Fitzgibbons TP, Buckbinder L, Czech MP (2015) Endothelial protein kinase MAP4K4 promotes vascular inflammation and atherosclerosis. Nature communications 6:8995. doi:10.1038/ncomms9995

18. Schwaid AG, Su C, Loos P, Wu J, Nguyen C, Stone KL, Kanyo J, Geoghegan KF, Bhattacharya SK, Dow RL, Buckbinder L, Carpino PA (2015) MAP4K4 Is a Threonine Kinase That Phosphorylates FARP1. ACS chemical biology 10 (12):2667-2671. doi:10.1021/acschembio.5b00679

19. Vitorino P, Yeung S, Crow A, Bakke J, Smyczek T, West K, McNamara E, Eastham-Anderson J, Gould S, Harris SF, Ndubaku C, Ye W (2015) MAP4K4 regulates integrin-FERM binding to control endothelial cell motility. Nature 519 (7544):425-430. doi:10.1038/nature14323

20. Feng XJ, Pan Q, Wang SM, Pan YC, Wang Q, Zhang HH, Zhu MH, Zhang SH (2016) MAP4K4 promotes epithelial-mesenchymal transition and metastasis in hepatocellular

carcinoma. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. doi:10.1007/s13277-016-5022-1

21. Roth Flach RJ, Danai LV, DiStefano MT, Kelly M, Garcia Menendez L, Jurczyk A, Sharma RB, Jung DY, Kim JH, Kim JK, Bortell R, Alonso LC, Czech MP (2016) Protein Kinase Mitogen Activated Protein Kinase Kinase Kinase Kinase 4 (MAP4K4) Promotes Obesity-Induced Hyperinsulinemia. The Journal of biological chemistry. doi:10.1074/jbc.M116.718932

22. Virbasius JV, Czech MP (2016) Map4k4 Signaling Nodes in Metabolic and Cardiovascular Diseases. Trends in endocrinology and metabolism: TEM. doi:10.1016/j.tem.2016.04.006

23. 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. Bioorganic & medicinal chemistry letters 20 (24):7503-7506. doi:10.1016/j.bmcl.2010.10.010

24. Barker JJ, Barker O, Boggio R, Chauhan V, Cheng RK, Corden V, Courtney SM, Edwards N, Falque VM, Fusar F, Gardiner M, Hamelin EM, Hesterkamp T, Ichihara O, Jones RS, Mather O, Mercurio C, Minucci S, Montalbetti CA, Muller A, Patel D, Phillips BG, Varasi M, Whittaker M, Winkler D, Yarnold CJ (2009) Fragment-based identification of Hsp90 inhibitors. ChemMedChem 4 (6):963-966. doi:10.1002/cmdc.200900011

25. 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, Sambrone 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 (8):3484-3493. doi:10.1021/jm500155b

26. 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. Bioorganic & medicinal chemistry letters 24 (18):4546-4552. doi:10.1016/j.bmcl.2014.07.071

27. Gobbi A, Giannetti AM, Chen H, Lee ML (2015) Atom-Atom-Path similarity and Sphere Exclusion clustering: tools for prioritizing fragment hits. Journal of cheminformatics 7:11. doi:10.1186/s13321-015-0056-8

28. Ndubaku CO, Crawford TD, Chen H, Boggs JW, Drobnick J, Harris SF, Jesudason R, McNamara E, Nonomiya J, Sambrone 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 medicinal chemistry letters 6 (8):913-918. doi:10.1021/acsmedchemlett.5b00174

29. 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 (1):235-242

30. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234 (3):779-815. doi:10.1006/jmbi.1993.1626

31. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein–ligand docking using GOLD. Proteins Struct Funct Bioinf 52 (4):609-623. doi:10.1002/prot.10465

32. 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 (2):455-461. doi:10.1002/jcc.21334

33. 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 (16):2785-2791. doi:10.1002/jcc.21256

34. Surpateanu G, Iorga BI (2012) Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors. J Comput Aided Mol Des 26 (5):595-601. doi:10.1007/s10822-011-9526-x

35. Colas C, Iorga BI (2014) Virtual screening of the SAMPL4 blinded HIV integrase inhibitors dataset. J Comput Aided Mol Des 28 (4):455-462. doi:10.1007/s10822-014-9707-5

36. Martiny VY, Martz F, Selwa E, Iorga BI (2016) Blind pose prediction, scoring, and affinity ranking of the CSAR 2014 dataset. J Chem Inf Model 56 (6):996-1003. doi:10.1021/acs.jcim.5b00337

37. Wang Y, Suzek T, Zhang J, Wang J, He S, Cheng T, Shoemaker BA, Gindulyte A, Bryant SH (2014) PubChem BioAssay: 2014 update. Nucleic Acids Res 42 (Database issue):D1075-1082. doi:10.1093/nar/gkt978

38. Huth JR, Park C, Petros AM, Kunzer AR, Wendt MD, Wang X, Lynch CL, Mack JC, Swift KM, Judge RA, Chen J, Richardson PL, Jin S, Tahir SK, Matayoshi ED, Dorwin SA, Ladror US, Severin JM, Walter KA, Bartley DM, Fesik SW, Elmore SW, Hajduk PJ (2007) Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies. Chemical biology & drug design 70 (1):1-12. doi:10.1111/j.1747-0285.2007.00535.x

Electronic Supplementary Material

Molecular docking performance evaluated on the D3R Grand Challenge 2015 drug-like ligand datasets

Edithe Selwa,[1] Virginie Y. Martiny,[1,2] Bogdan I. Iorga[1,]*

[1] Institut de Chimie des Substances Naturelles, CNRS UPR 2301, LabEx LERMIT, 91198 Gif-sur-Yvette, France; [2] Department of Nephrology and Dialysis, AP-HP, Tenon Hospital, INSERM UMR_S 1155, 75020 Paris, France

Corresponding Author

[*] Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Table of contents

Protein Data Bank (PDB) structures used for the preliminary analysis

HSP90:

191 structures were available in the PDB for human HSP90 at the moment when the D3R Grand Challenge 2015 took place. They were organized in 11 distinct groups, according to the conformation of the fragment 99-129. The representative structure for each group (based on the crystal structure resolution and lack of missing residues) is colored in red.

Group 1: 1uy6, 1uy7, 1uy8, 1uy9, 1uyc, 1uyd, 1uye, 1uyf, 1uyg, 1uyh, 1uyi, 1uyk, 1uym, 2fwy, 2fwz, 2qg2, 2wi4, 2wi7, 2xds, 2xdu, 2ye7, 2ye8, 2yee, 2yei, 2yej, 2yjx, 2yk2, **2yk9**, 2ykb, 2ykc, 2yke, 2yki, 2ykj, 3b25, 3d0b, 3ft8, 3hyy, 3hz1, 3hz5, 3inw, 3inx, 3mnr, 3nmq, 3o0i, 3qdd, 3qtf, 3r91, 3r92, 3rkz, 3wq9, 4cwf, 4cwn, 4cwo, 4cwp, 4cwq, 4cwr, 4cws, 4cwt, 4eft, 4efu, 4hy6, 4l8z, 4l90, 4l91, 4lwe, 4nh7, 4nh8, 4o04, 4o05, 4o07, 4o09, 4o0b, 4r3m, 4u93, 4xip, 4xiq, 4xir, 4xit

Group 2: 1uyl, 1yer, 2bsm, 2bt0, 2ccs, 2cct, 2ccu, 2jjc, 2uwd, 2vci, 2vcj, 2wi1, 2wi2, 2wi3, 2wi5, 2xab, 2xdl, 2ye2, 2ye3, 2ye4, 2ye5, 2ye6, 2ye9, 2yea, 2yeb, 2yec, 2yed, 2yeg, 2yeh, 2yi0, 2yi6, 2yi7, 2yjw, 3b24, 3bm9, 3ft5, 3hhu, 3owb, **3t0h**, 4eeh, 4egh, 4fcp, 4fcq, 4l94, 4lwf, 4lwg, 4lwh, 4lwi

Group 3: 1byq, 1yc3, 1yes, 1yet, 2byh, 2byi, 2xdx, 2xht, 2xhx, 2xjg, 2xjx, 2xk2, 3b27, 3r4m, 3r4o, 3t0z, **3vha**, 3vhc, 4b7p, 4egi

Group 4: 1osf, 2xdk, 2xjj, 2yef, 3eko, 3ekr, 3hek, 3k97, 3k98, 3r4n, 3r4p, 3rlp, 3rlq, 3rlr, **3t10**, 3t1k, 3t2s, 3vhd, 4egk, 4jql

Group 5: 1yc1, 1yc4, 2bz5, 2qfo, 2qg0, 2yi5, 3b26, 3b28, 3hyz, 3ow6, 3tuh, **3wha**, 4fcr, 4l93, 4w7t

Group 6: 2wi6, 2xhr, **3bmy**

Group 7: 4awo, 4awp, **4awq**

Group 8: **4bqj**

Group 9: **3owd**

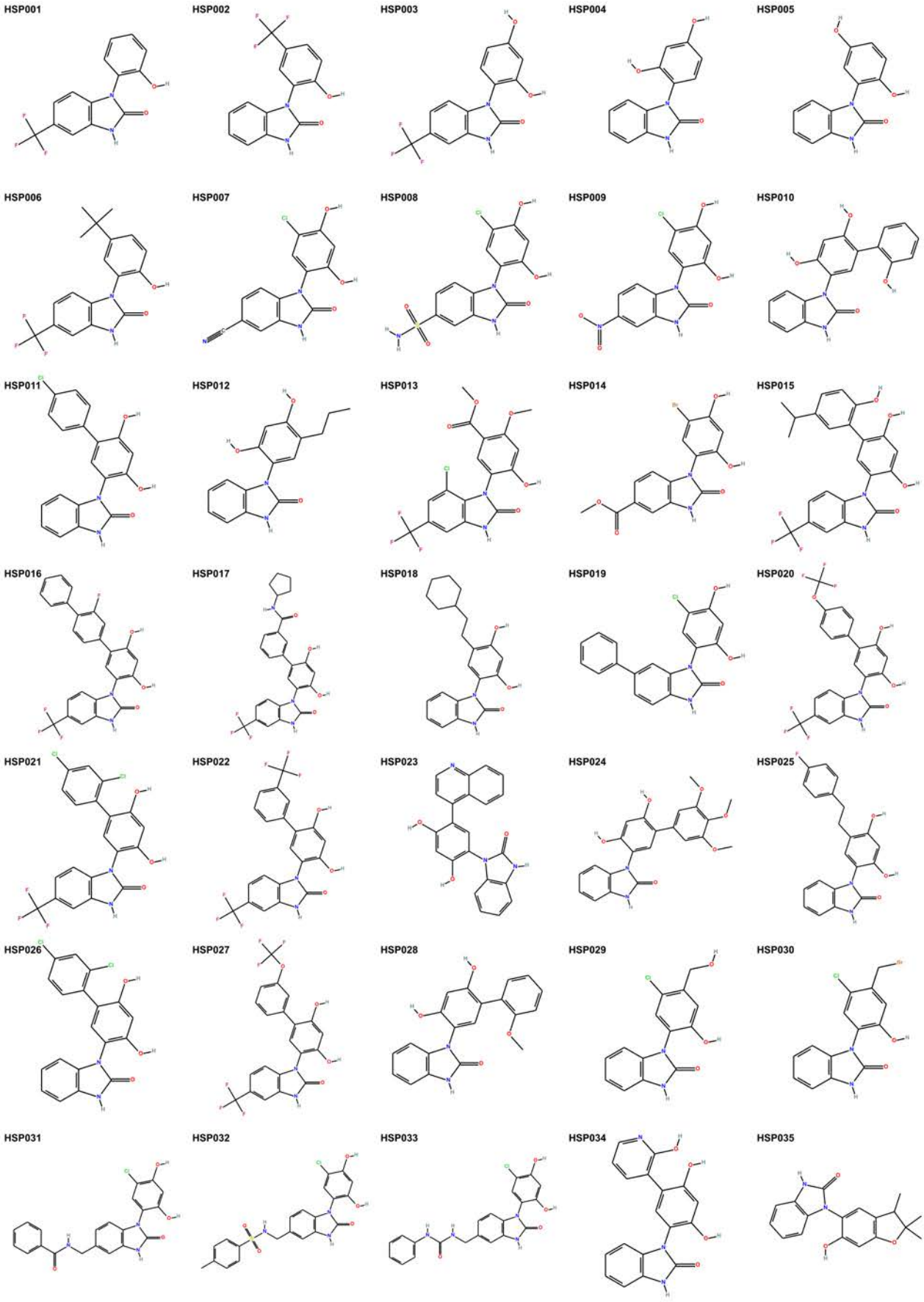
Group 10: **3k99**

Group 11: **2qf6**

MAP4K4:

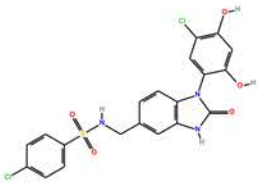
8 structures were available in the PDB for human MAP4K4 at the moment when the D3R Grand Challenge 2015 took place: 4obo, 4obp, 4obq, 4rvt, 4u43, 4u44, 4u45, 4zk5.

Figure S1. Chemical structures of the entire HSP90 dataset, containing 180 ligands used for ranking prediction.

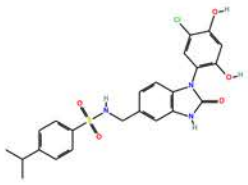


HSP036

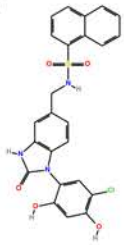
HSP037



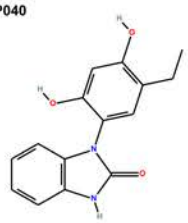
HSP038



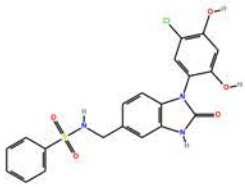
HSP039



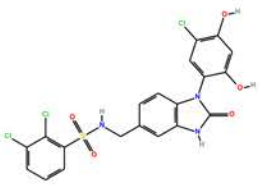
HSP040



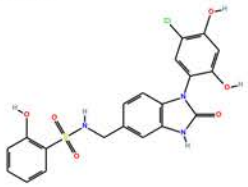
HSP041



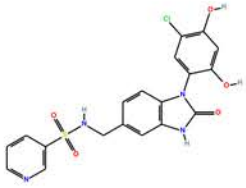
HSP042



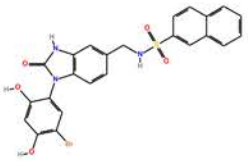
HSP043



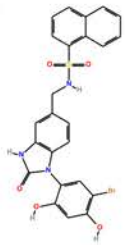
HSP044



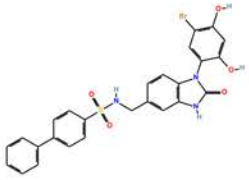
HSP045



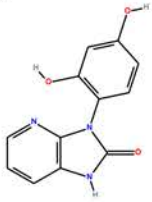
HSP046



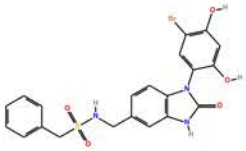
HSP047



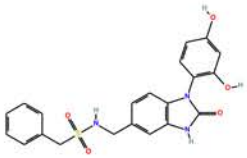
HSP048



HSP049

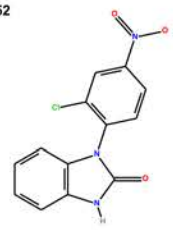


HSP050



HSP051

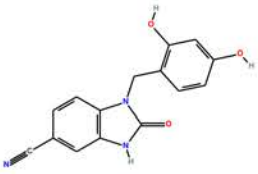
HSP052



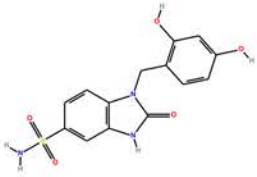
HSP053

HSP054

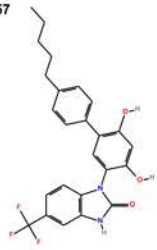
HSP055



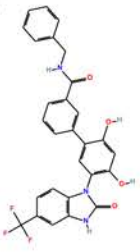
HSP056



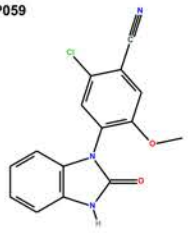
HSP057



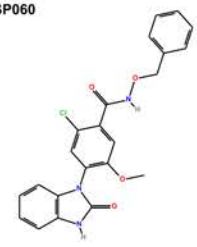
HSP058



HSP059

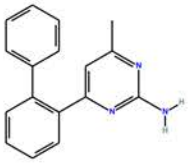


HSP060

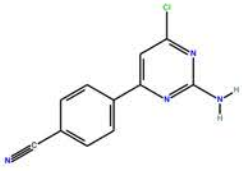


HSP061

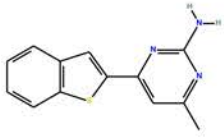
HSP063



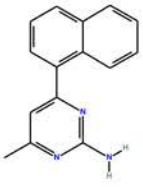
HSP064



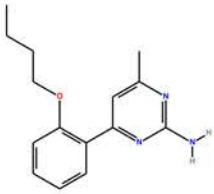
HSP065



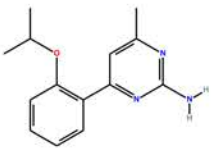
HSP066



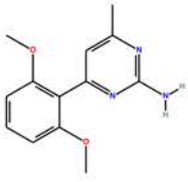
HSP068



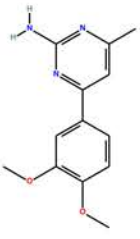
HSP069



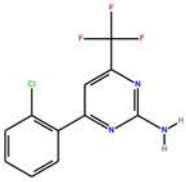
HSP070



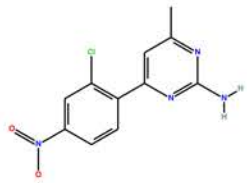
HSP071



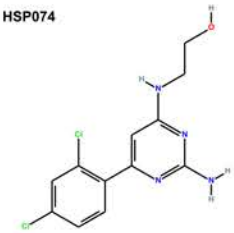
HSP072



HSP073

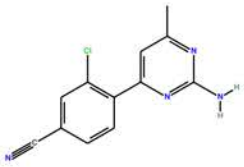


HSP074

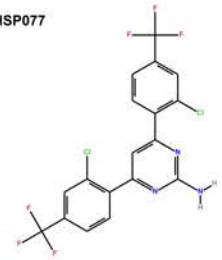


HSP075

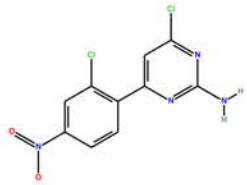
HSP076



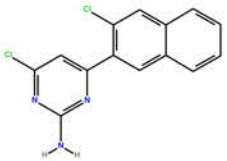
HSP077



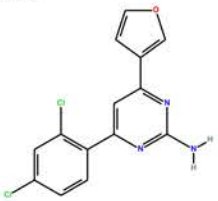
HSP078



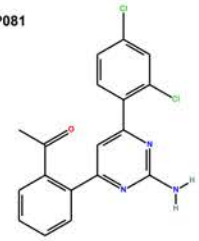
HSP079



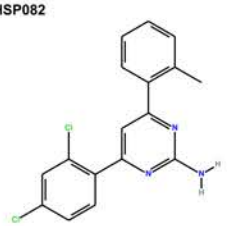
HSP080



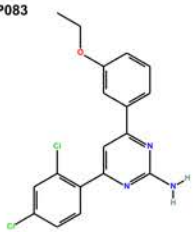
HSP081



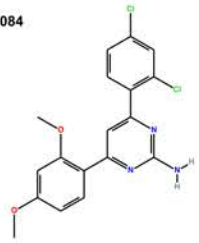
HSP082



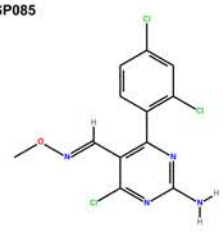
HSP083



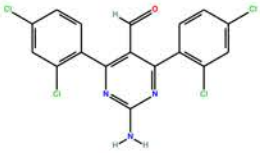
HSP084



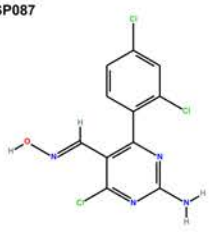
HSP085



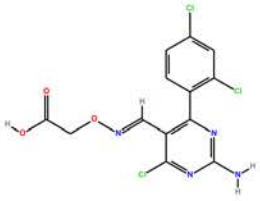
HSP086



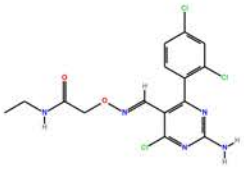
HSP087



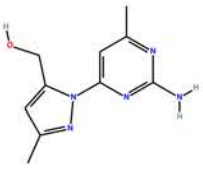
HSP088



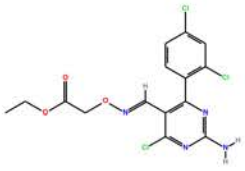
HSP089



HSP090



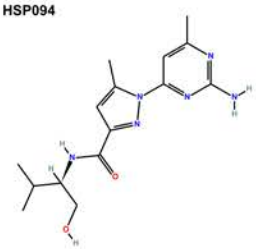
HSP091



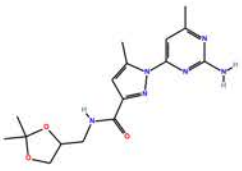
HSP092

HSP093

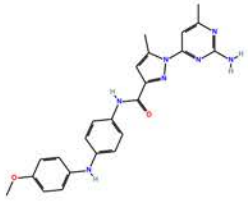
HSP094



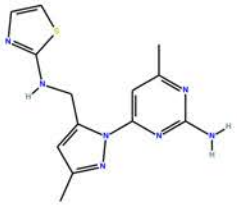
HSP095



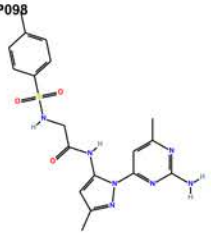
HSP096



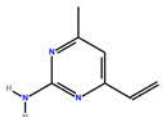
HSP097



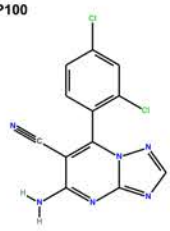
HSP098



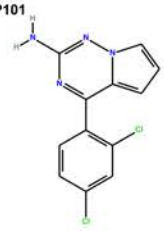
HSP099



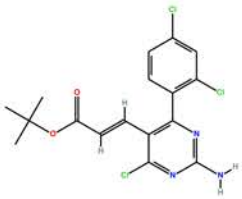
HSP100



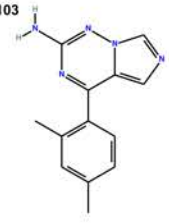
HSP101



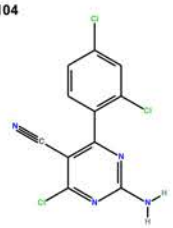
HSP102



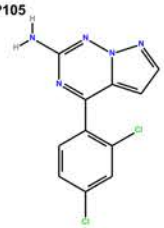
HSP103



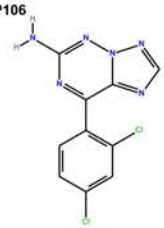
HSP104



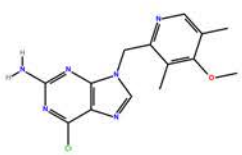
HSP105



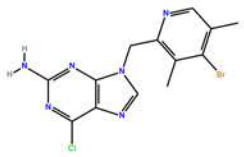
HSP106



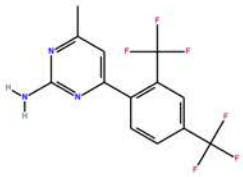
HSP107



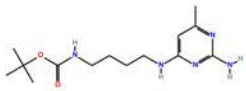
HSP108



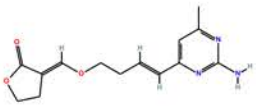
HSP110



HSP111

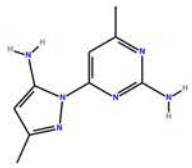


HSP112

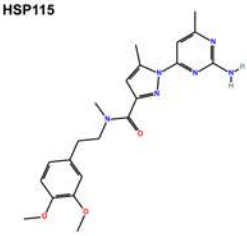


HSP113

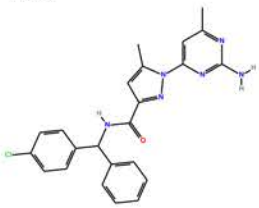
HSP114



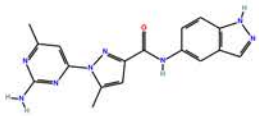
HSP115



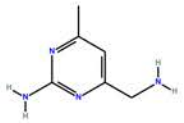
HSP116



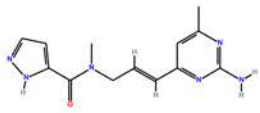
HSP117



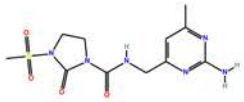
HSP118



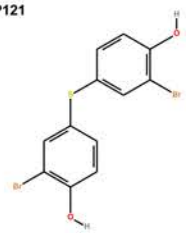
HSP119



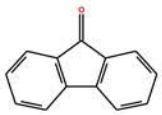
HSP120



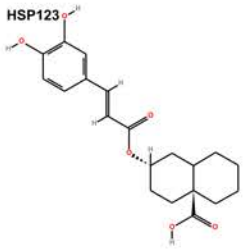
HSP121



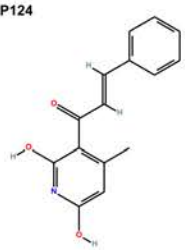
HSP122



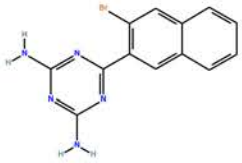
HSP123



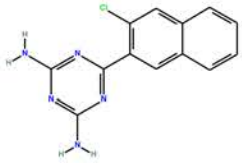
HSP124



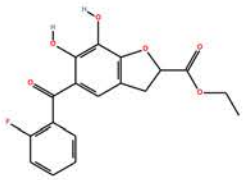
HSP125



HSP126

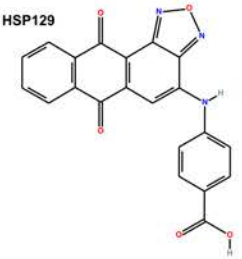


HSP127



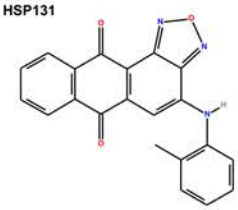
HSP128

HSP129

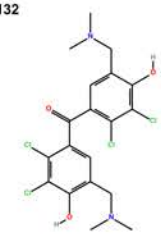


HSP130

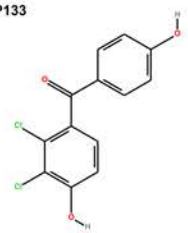
HSP131



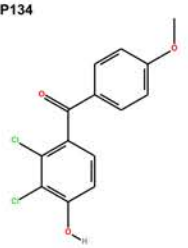
HSP132



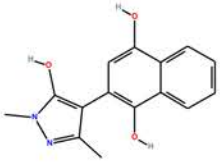
HSP133



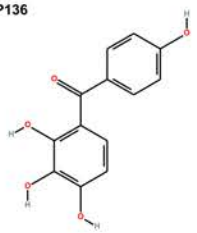
HSP134



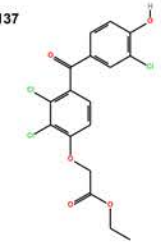
HSP135



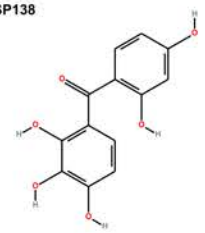
HSP136



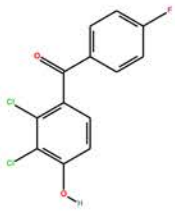
HSP137



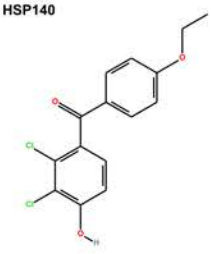
HSP138



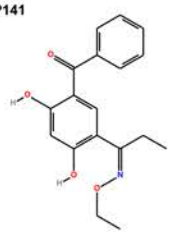
HSP139



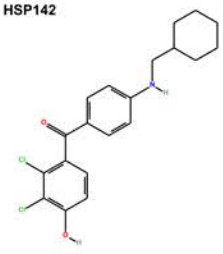
HSP140



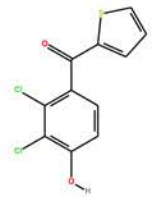
HSP141



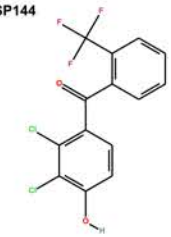
HSP142



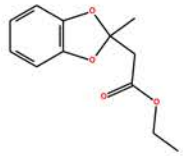
HSP143



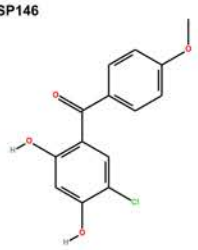
HSP144



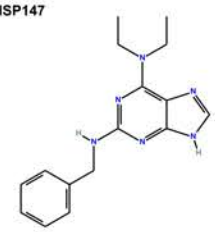
HSP145



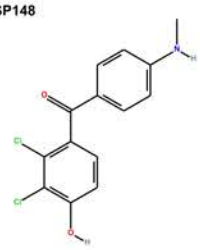
HSP146



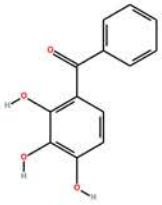
HSP147



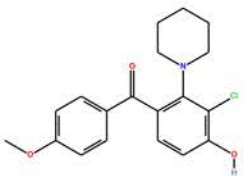
HSP148



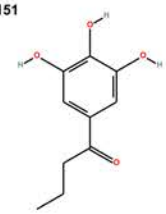
HSP149



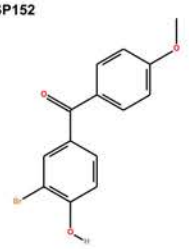
HSP150



HSP151

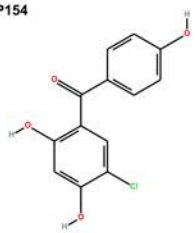


HSP152

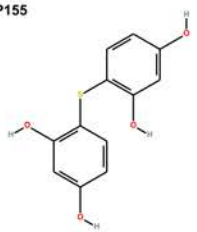


HSP153

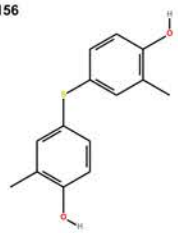
HSP154



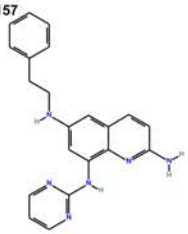
HSP155



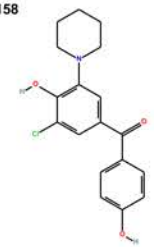
HSP156



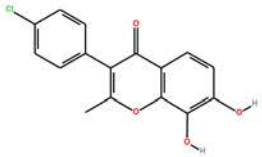
HSP157



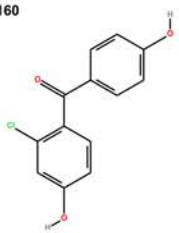
HSP158



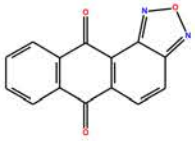
HSP159



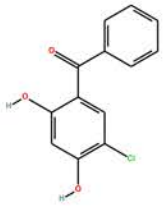
HSP160



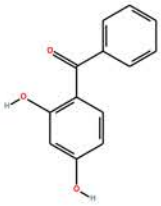
HSP161



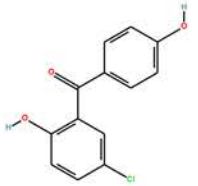
HSP162



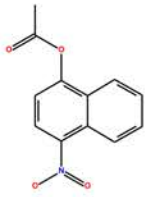
HSP163



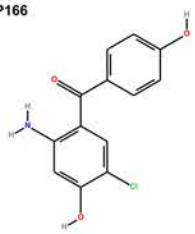
HSP164



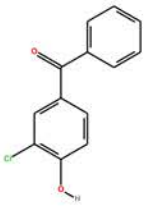
HSP165



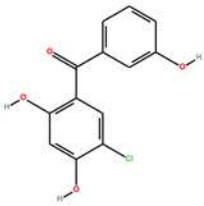
HSP166



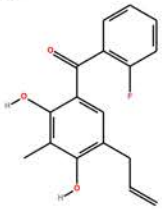
HSP167



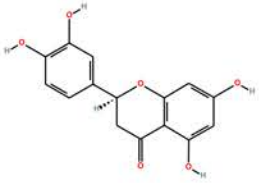
HSP168



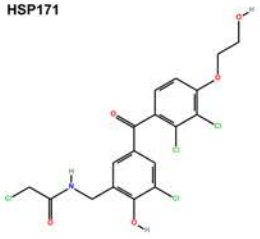
HSP169



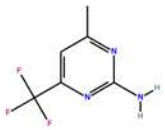
HSP170



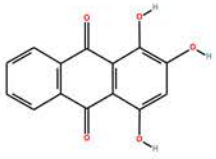
HSP171



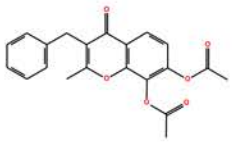
HSP172



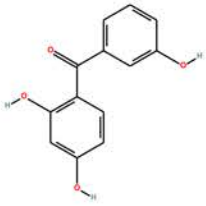
HSP173



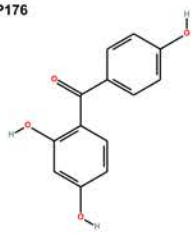
HSP174



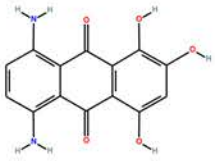
HSP175



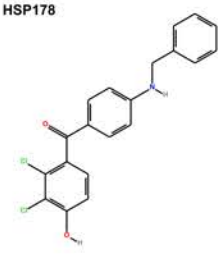
HSP176

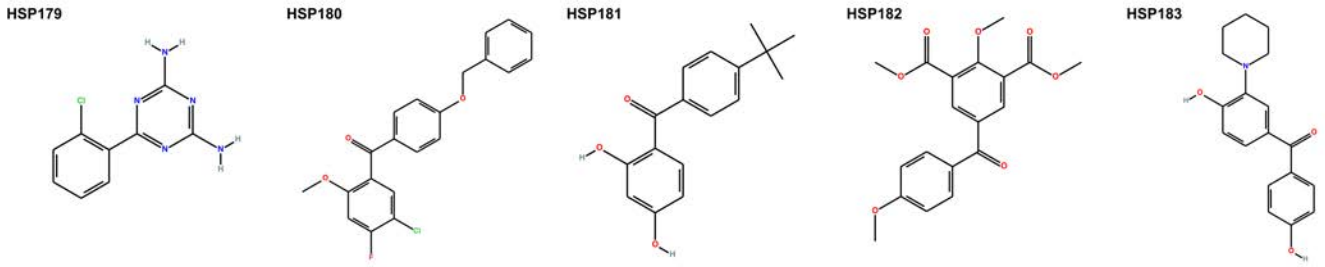


HSP177



HSP178





HSP90 ligands regroupment according to their chemical structure:

Aminopyrimidines: HSP063, HSP064, HSP065, HSP066, HSP068, HSP069, HSP070, HSP071, HSP072, HSP073, HSP074, HSP075, HSP076, HSP077, HSP078, HSP079, HSP080, HSP081, HSP082, HSP083, HSP084, HSP085, HSP086, HSP087, HSP088, HSP089, HSP090, HSP091, HSP092, HSP093, HSP094, HSP095, HSP096, HSP097, HSP098, HSP099, HSP100, HSP101, HSP102, HSP103, HSP104, HSP105, HSP106, HSP107, HSP108, HSP110, HSP111, HSP112, HSP113, HSP114, HSP115, HSP116, HSP117, HSP118, HSP119, HSP120, HSP125, HSP126, HSP147, HSP157, HSP172, HSP179

Benzimidazolones: HSP001, HSP002, HSP003, HSP004, HSP005, HSP006, HSP007, HSP008, HSP009, HSP010, HSP011, HSP012, HSP013, HSP014, HSP015, HSP016, HSP017, HSP018, HSP019, HSP020, HSP021, HSP022, HSP023, HSP024, HSP025, HSP026, HSP027, HSP028, HSP029, HSP030, HSP031, HSP032, HSP033, HSP034, HSP035, HSP036, HSP037, HSP038, HSP039, HSP040, HSP041, HSP042, HSP043, HSP044, HSP045, HSP046, HSP047, HSP048, HSP049, HSP050, HSP051, HSP052, HSP053, HSP054, HSP055, HSP056, HSP057, HSP058, HSP059, HSP060, HSP061

Benzophenone-like: HSP122, HSP127, HSP128, HSP129, HSP130, HSP131, HSP132, HSP133, HSP134, HSP136, HSP137, HSP138, HSP139, HSP140, HSP141, HSP142, HSP143, HSP144, HSP146, HSP148, HSP149, HSP150, HSP152, HSP154, HSP158, HSP159, HSP160, HSP161, HSP162, HSP163, HSP164, HSP166, HSP167, HSP168, HSP169, HSP170, HSP171, HSP173, HSP174, HSP175, HSP176, HSP177, HSP178, HSP180, HSP181, HSP182, HSP183

Miscellaneous: HSP121, HSP123, HSP124, HSP135, HSP145, HSP151, HSP153, HSP155, HSP156, HSP165

Figure S2. Chemical structures of the entire MAP4K4 dataset, containing 30 ligands used for pose prediction.

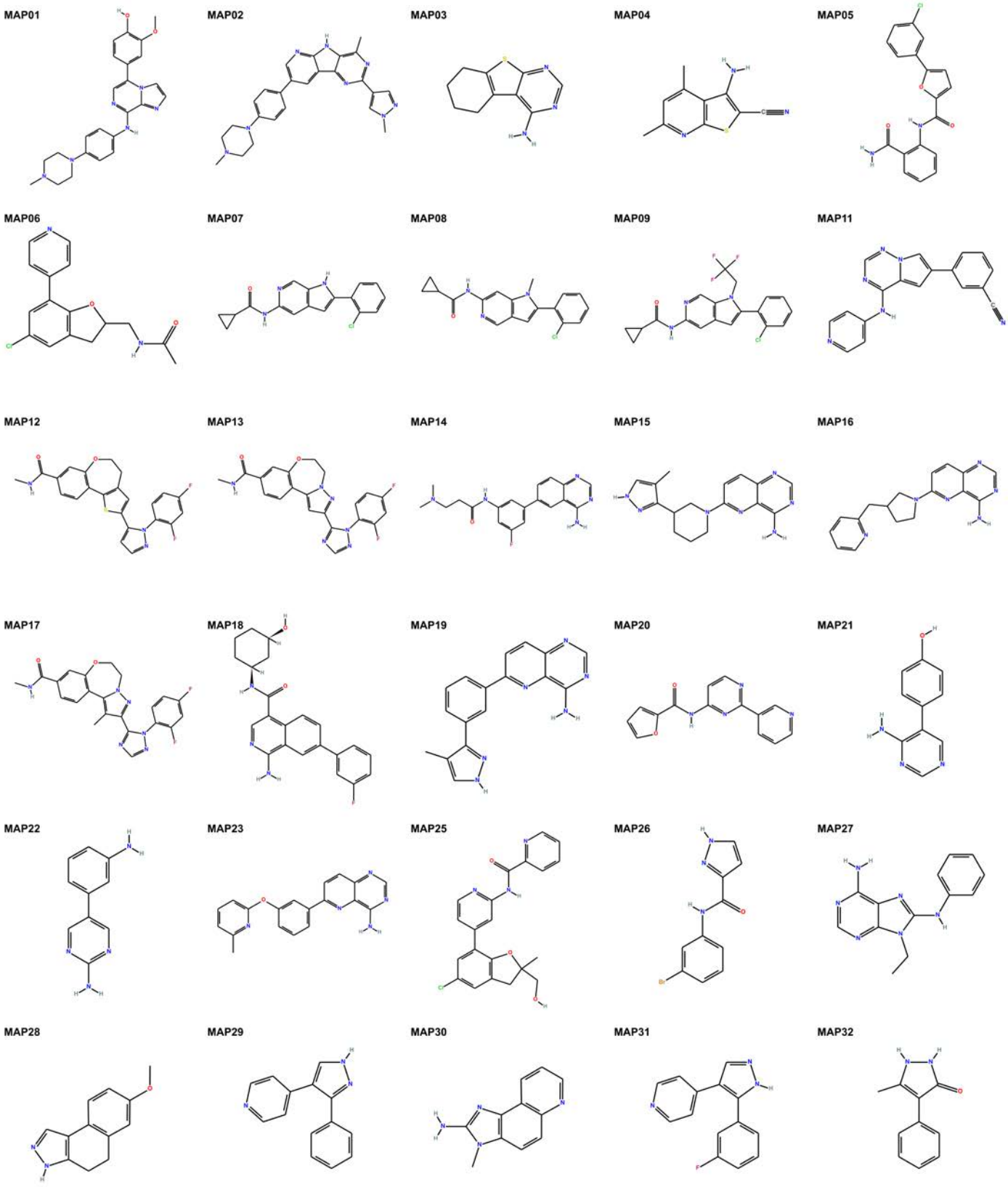
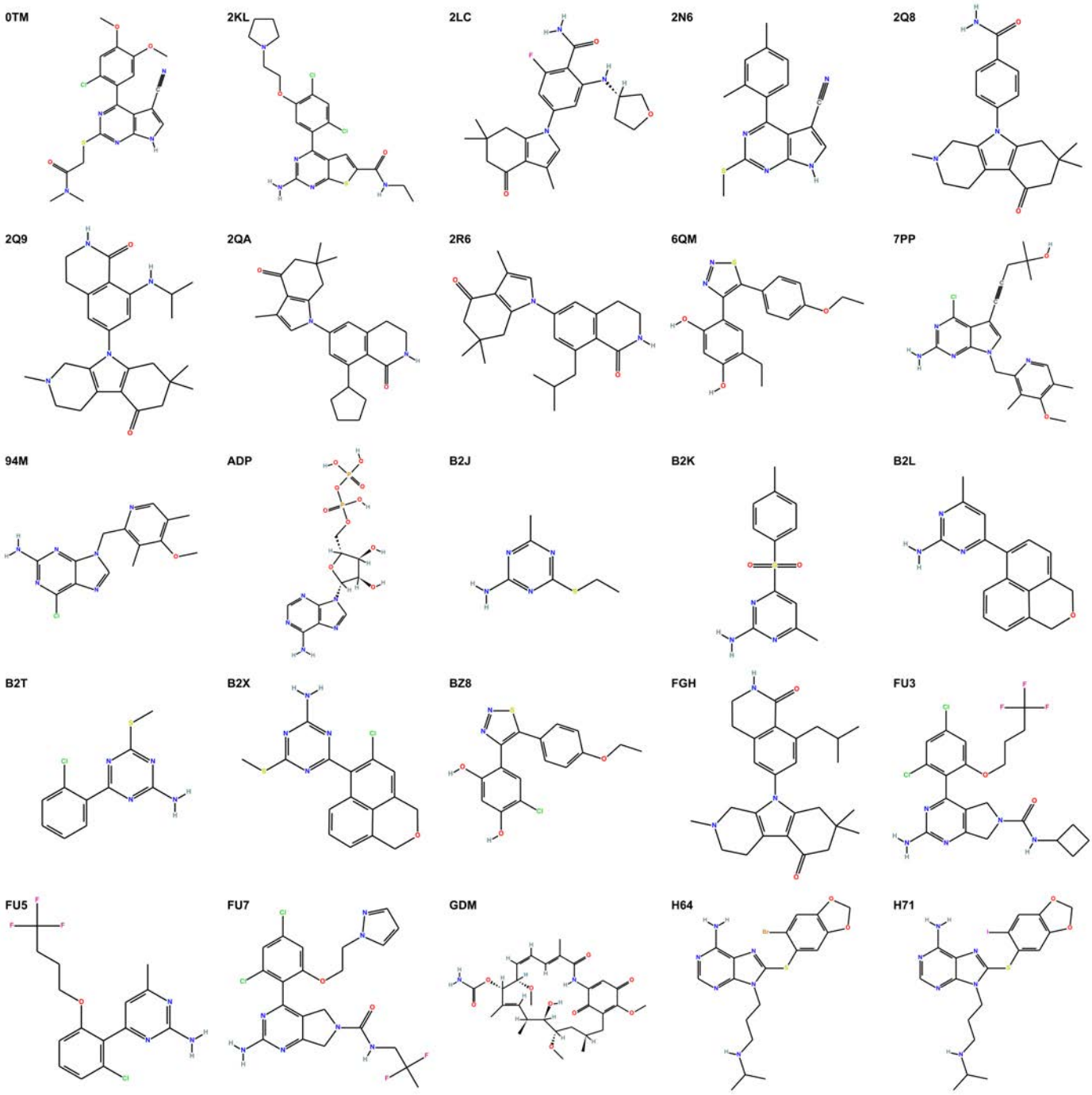
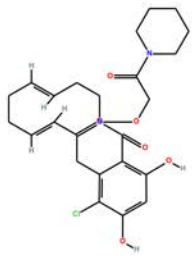


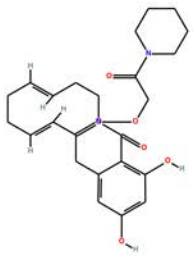
Figure S3. Chemical structures of HSP90 ligands found in the PDB that are also present in PubChem BioAssay with activity data. The compound names are the PDB ligand codes.



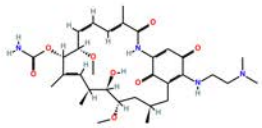
JZB



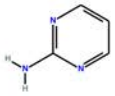
JZC



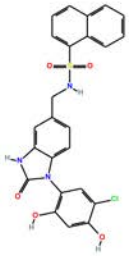
KOS



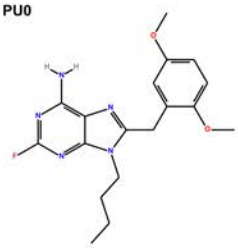
LGA



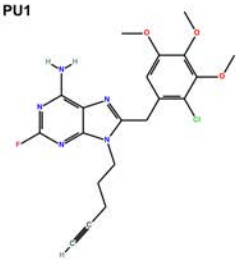
MEY



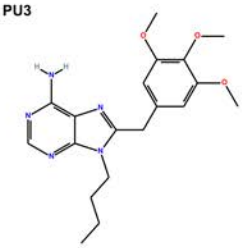
PU0



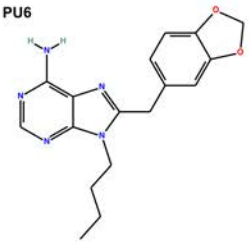
PU1



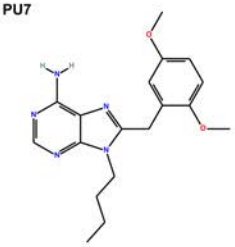
PU3



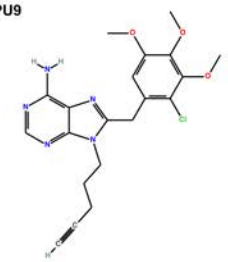
PU6



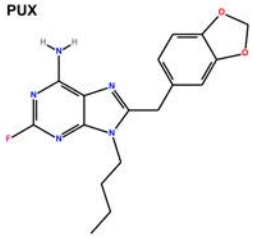
PU7



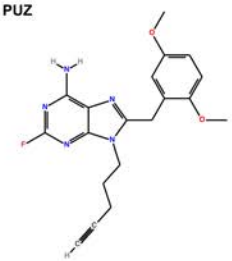
PU9



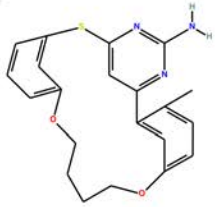
PUX



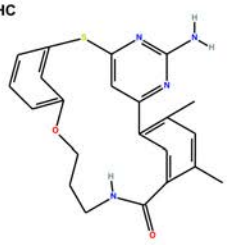
PUZ



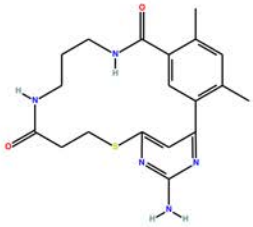
VHA



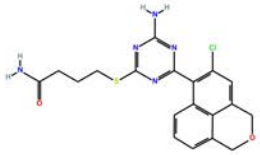
VHC



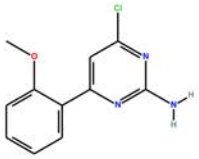
VHE



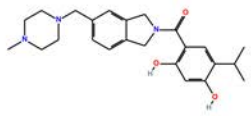
WHA



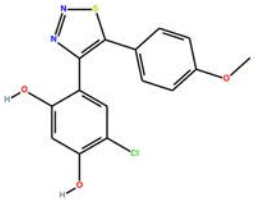
WOE



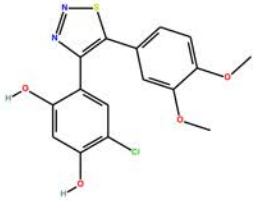
XJX



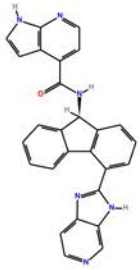
YI0



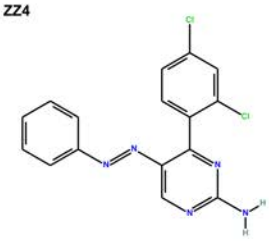
YI5



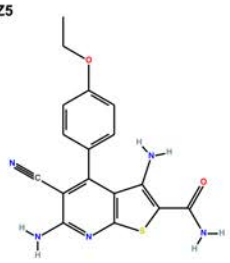
YKI



ZZ4



ZZ5



ZZ6

Figure S4. Chemical structures of MAP4K4 ligands found in the PDB. Compound names include the PDB ligand code and the PDB code of the structure containing the ligand.

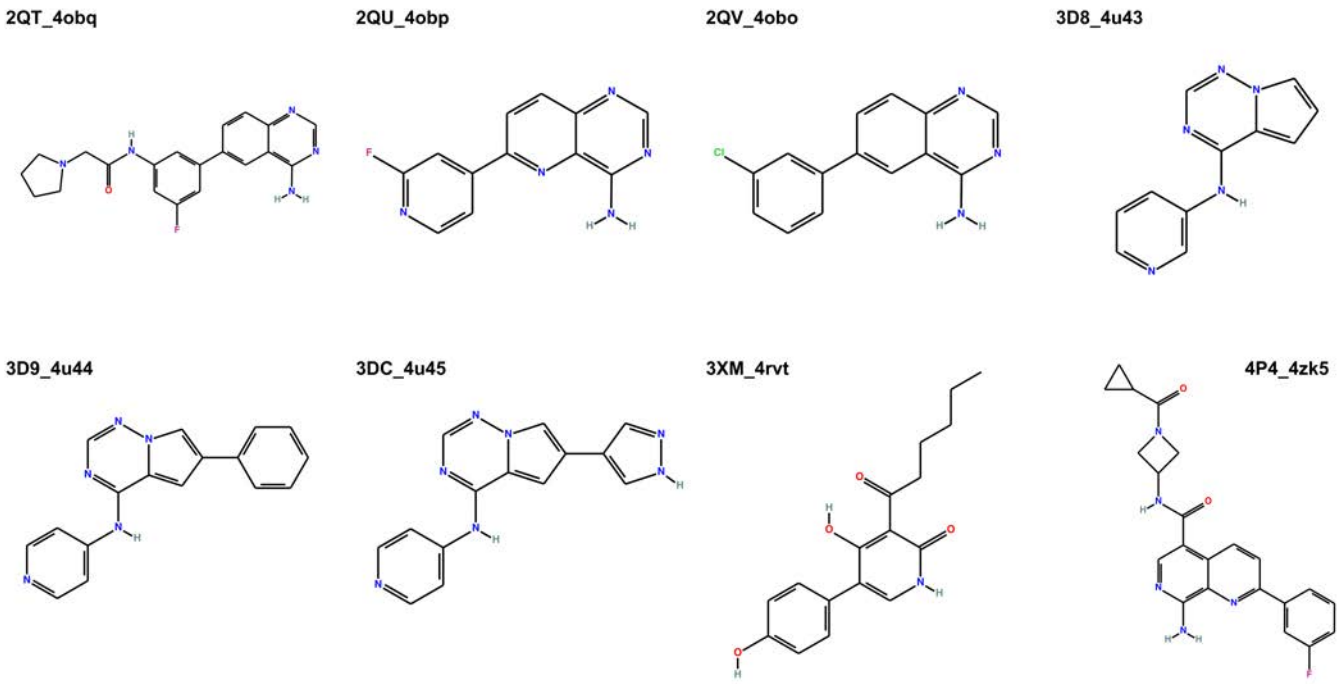


Figure S5. Performance of Phase 1 ranking submissions (Pearson R in blue and Kendall Tau in light blue) for the HSP90 (left) and MAP4K4 (right) datasets. The best predictions are those with values closest to 1. Our submissions are colored in red and light red, respectively.