



**HAL**  
open science

# Ranking docking poses by graph matching of protein–ligand interactions: lessons learned from the D3R Grand Challenge 2

Priscila da Silva Figueiredo Celestino Gomes, Franck Da-Silva, Guillaume Bret, Didier Rognan

► **To cite this version:**

Priscila da Silva Figueiredo Celestino Gomes, Franck Da-Silva, Guillaume Bret, Didier Rognan. Ranking docking poses by graph matching of protein–ligand interactions: lessons learned from the D3R Grand Challenge 2. *Journal of Computer-Aided Molecular Design*, 2018, 32 (1), pp.75-87. 10.1007/s10822-017-0046-1 . hal-02158149

**HAL Id: hal-02158149**

**<https://hal.science/hal-02158149>**

Submitted on 19 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

[Click here to view linked References](#)

## **Ranking docking poses by graph matching of protein-ligand interactions: Lessons learned from the D3R Grand Challenge 2**

Priscila da Silva Figueiredo Celestino Gomes<sup>1,2</sup>, Franck Da Silva<sup>1</sup>, Guillaume Bret,<sup>1</sup> Didier Rognan<sup>1\*</sup>

<sup>1</sup>Laboratoire d'Innovation Thérapeutique, UMR 7200 CNRS-Université de Strasbourg, 67400 Illkirch,  
France

<sup>2</sup>Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

\*To whom correspondence should be addressed (phone: +33 3 68 85 42 35, fax: +33 3 68 85 43 10,  
email: rognan@unistra.fr), ORCID ID: 0000-0002-0577-641X

## **Abstract**

A novel docking challenge has been set by the Drug Design Data Resource (D3R) in order to predict the pose and affinity ranking of a set of Farnesoid X receptor (FXR) agonists, prior to the public release of their bound X-ray structures and potencies. In a first phase, 36 agonists were docked to 26 Protein Data Bank (PDB) structures of the FXR receptor, and next rescored using the in-house developed GRIM method. GRIM aligns protein-ligand interaction patterns of docked poses to those of available PDB templates for the target protein, and rescore poses by a graph matching method. In agreement with results obtained during the previous 2015 docking challenge, we clearly show that GRIM rescoring improves the overall quality of top-ranked poses by prioritizing interaction patterns already visited in the PDB. Importantly, this challenge enables us to refine the applicability domain of the method by better defining the conditions of its success. We notably show that rescoring apolar ligands in hydrophobic pockets leads to frequent GRIM failures. In the second phase, 102 FXR agonists were ranked by decreasing affinity according to the Gibbs free energy of the corresponding GRIM-selected poses, computed by the HYDE scoring function. Interestingly, this fast and simple rescoring scheme provided the third most accurate ranking method among 57 contributions. Although the obtained ranking is still unsuitable for hit to lead optimization, the GRIM-HYDE scoring scheme is accurate and fast enough to post-process virtual screening data.

**Keywords:** docking, D3R, Drug Discovery Data resource, Grand challenge

## Introduction

Molecular docking is still the most straightforward computational technique to predict three-dimensional (3-D) atomic coordinates of a protein-ligand complex [1,2]. Any docking attempt is aimed at simultaneously solving two questions: (i) What is the conformation of the ligand when bound to its target?, (ii) What is the relative orientation of the ligand with respect to its host protein? Almost 35 years of practice and the development of over 60 different docking methods [3] have considerably helped the community to enhance the quality of the answers to the above two questions, and to better depict the precise applicability domain of the method. Very soon, computational chemists have realized that posing a drug-like low molecular weight ligand to a structurally druggable and reasonably rigid cavity is reachable in the large majority of cases [4]. However, rank these docking solutions to prioritize the most relevant one in first place is still a major issue [5]. Pushed by massive interest of the pharmaceutical industry in structure-based drug design approaches, the docking community has then organized various resources to challenge computational chemists to refine both their practices and methods. Among the most important resources are the Community Structure Activity Resource (CSAR)[6] and the Drug Design Data Resource (D3R) [7] which periodically organize challenges aimed at predicting protein-ligand coordinates prior to the release of their crystal structures, discriminating X-ray poses from decoys and scoring/ranking a set of ligands for well-defined targets [8,5,9-14].

The first D3R Grand Challenge [13], launched in 2015, was a good opportunity to test in real life conditions, our recently-developed graph matching algorithm (GRIM)[15] to rank docking poses. GRIM converts protein-ligand atomic coordinates into a graph where each node is matched to an interaction pseudoatom (IPA) describing a single protein-ligand interaction (apolar contact, aromatic interaction, hydrogen bond, ionic bond, metal chelation). For each interaction, three IPAs are derived, one on the ligand-interacting atom, one on the protein-interacting atom and one at the geometric barycenter of the later 2 IPAs. Since the graph precisely depicts the corresponding protein-ligand interaction pattern, two different protein-ligand complexes can thus be easily compared by using standard graph matching

techniques to determine a clique, in other words, the largest maximal common subgraph which describes the **protein-ligand interaction pattern** shared by the two investigated complexes. This simple graph alignment-based method produces a quantitative estimate of the similarity of two protein-ligand interaction patterns. It can therefore be used to rank docking poses by similarity to any reference interaction pattern (e.g. stored in the Protein Data Bank) or to post-process virtual screening docking data. In the latter two scenarios and starting from the same set of docking poses, GRIM significantly outperformed standard scoring functions in either posing a set of protein ligands or enriching virtual screening hits in true ligands [15].

In the first D3R challenge, GRIM rescoring of Surflex-poses was ranked 3rd out of 42 submissions for the pose-prediction of HSP90 $\alpha$  and MAP4k4 inhibitors, respectively [16]. To check the target and ligand set dependency of our method, we applied the same strategy to the D3R second challenge consisting in (i) pose prediction of 36 agonists of the Farnesoid X nuclear receptor (FXR), (ii) rank/score a set of 102 FXR agonists by decreasing binding affinity. This novel competition represents a real challenge for several reasons: (i) the FXR receptor structure exhibits a flexible binding site whose conformation depends on the bound-ligand chemotype [17], (ii) some but not all ligands require a bound water [18], (iii) about one third of the new ligands whose pose need to be predicted exhibit a chemotype never co-crystallized with the target, (iv) FXR ligands may adopt almost non overlapping poses due to the large size and hydrophobic nature of the ligand-binding cavity.

## Computational Methods

### FXR dataset

The structures of the FXR apo-receptor (PDB format) and 102 FXR ligands (SD files and SMILES strings, **Supplementary Table 1**) were downloaded from the D3R Grand Challenge 2 website [7] as a zipped archive file (417\_data\_517096.zip). In addition, 26 agonist-bound FXR X-ray structures were defined as templates (**Supplementary Table 2**) by searching the RCSB Protein Data Bank [19] for the text keyword 'FXR' and a known bound ligand, confirmed as a true FXR agonist. Existing hydrogen atoms were removed and added again while optimizing both the protonation and tautomeric states of all atoms using Protoss [20]. Water molecules were explicitly conserved at the condition that the water oxygen atom was closer than 6.5 Å from any ligand heavy atom and that at least two hydrogen bonds with either the protein and/or the ligand could be identified assuming a donor-acceptor distance lower than 3.5 Å and a donor-hydrogen-acceptor angle higher than 120 deg. Protein (including water atoms) and ligand atoms were then separately saved in MOL2 file format in SYBYL-X 2.1.1 [21]. Last, atomic coordinates (main chain atoms only) of the 26 protein templates were structurally aligned to that of the apostructure provided by the organizers of the D3R Grand Challenge, using the 'Align\_Structures' module of SYBYL X 2.1.1.

The SD files of the 102 ligands to dock were converted in three-dimensional (3-D) atomic coordinates with Corina v3.40 [22]. The protonation state of every ligand was assigned at pH 7.4 with Filter v.2.5.1.4 [23] and verified manually. 3-D coordinates were saved in MOL2 file format.

### Docking poses generation

Ligands were docked **to the above mentioned 26** protein structures using Surflex v.3066 [24]. For each protein input structure, a *protomol* was first generated using a list of binding site residues (including bound waters) for which at least one heavy atom was closer than 6.5 Å from one **co-crystallized** ligand

heavy atom. The *protomol* is a pseudo-ligand used as a target to generate putative alignments of fragments of an input ligand. It utilizes three probes (C=O, NH, CH<sub>4</sub>) whose positions correspond to energetically favored locations for hydrogen bond acceptors, hydrogen bond donors and apolar atoms, respectively.

In order to better define the docking zone, an additional 3 Å voxels of space around the volume specified by the *protomol* (*proto\_bloat* parameter set to 3) was chosen. The degree of buriedness (*proto\_thresh* parameter) for the primary volume used to generate the *protomol* was set to a value of 0.3. The protein residues side-chains were kept rigid. Other parameters were assigned as default. The docking accuracy parameter set *-pgeom* was used. The '*pgeom*' option starts each docking from four initial and different poses to ensure good search coverage, turns on ligand minimization prior to docking and after docking (in-pocket minimization), ensures that the returned poses are different from one another by at least 0.5 Å rmsd, and saves a total of 20 poses (ranked by Surflex energy score, from 000 to 019). Each of the 102 FXR ligands was docked to each of the 26 protein structure templates, thereby generating a set of 53,040 (102 x 20 x 26) docking poses. Only poses with a predicted Surflex score (in pK<sub>d</sub> unit) higher than 2.0 were kept for further analysis.

### GRIM rescoring

Protein-ligand interaction patterns were generated with IChem [25] for both the 26 X-ray structures and the docking poses, and further compared to that of 26 FXR-agonist X-ray templates with GRIM [15]. Each alignment was quantified using the empirical GRIMscore [15] (Eq1).

$$\text{GRIMscore} = 0.5006 + 0.0151.\text{NLig} + 0.0039.\text{NCenter} + 0.0143.\text{NProt} + 0.2098.\text{SumCl} - 0.0720.\text{RMSD} - 0.0003.\text{DiffI} \quad (1)$$

NLig: number of matched ligand-based IPAs

Ncent: number of matched centered IPAs

NProt: number of matched protein-based IPAs

$$\text{SumCl: } \frac{\sum \text{pair weights in clique}}{\sum \text{all possible pair weights}}$$

RMSD: root mean square deviation (in Å) of the matched clique

DiffI: absolute value of the difference in the number of IPAs between reference and query.

For every ligand, the five poses with the highest GRIMscore (GRIM-1 to GRIM-5) were retained.

In a previous study [15], the GRIMscore has been determined by fitting the above described six parameters to the IShape similarity score [15] parametrized on a training set of 1800 protein-ligand X-ray structures. In contrast to IShape, the GRIM score shows an excellent and sharp discrimination between known similar and known dissimilar complexes [15]. Please note that GRIM assigns a weight to each IPA (used in the SumCl parameters), whose value is inversely proportional to the frequency of the corresponding interaction among 9877 sc-PDB protein-ligand complexes [13].

### **Root-mean square deviations to atomic coordinates**

Root mean square deviations (rmsd) of predicted poses to the X-ray structures, **released at the end of the first stage of the challenge**, were computed using the *rms* routine of the Surflex package. Rmsd were computed only for heavy atoms and included symmetry operations to explicitly account for equivalent atoms (e.g. carboxylate oxygen atoms). One compound (FXR\_33) was excluded from the pose prediction phase of the challenge due to an oxidation during the co-crystallization process, generating a structure in the crystal that did not correspond to that initially provided.

### **Ligand ranking**



For each of the 102 FXR ligands (FXR\_1 to FXR\_102) to rank, the corresponding protein-ligand complex with the absolute best GRIMscore was selected and its binding free energy was estimated using the HYDE scoring function [26] (see Eq. 2) implemented in seeSAR [27].

$$\Delta G_{HYDE}^i = \sum_{atom\ i} \Delta G_{dehydration}^i + \Delta G_{H-bond}^i \quad (2)$$

Briefly, the HYDE scoring function relies on atom type-specific hydration and desolvation terms that have been carefully calibrated using octanol/water partition coefficients of small molecules. For more details on the HYDE scoring energy terms, the reader is referred to the original article [26].

## Results and Discussion

### Suitability of known PDB interaction patterns for FXR agonist docking

We first examined the currently available protein-ligand interaction patterns in the PDB by selecting 25 entries (one having 2 chains considered) corresponding to FXR structures co-crystallized with 26 agonists of three different chemotypes (**Supplementary Table 2**). At this step, we only considered FXR structures bound to a known agonist for two main reasons: (i) the FXR receptor, like most nuclear hormone receptors, is known to change its overall conformation with respect to the functional effect of the ligand [28], (ii) the purpose of the current D3R challenge was focused on FXR agonists only. The FXR structure in complex with ivermectin (PDBID 4wvd) was discarded from the list because of the very unique binding mode of the latter compound [29]. **Ivermectin exhibits both a peculiar structure and protein-ligand interaction pattern with respect to the 26 X-ray templates as well as the 102 FXR agonist to dock. We therefore decided to remove this structure, as we believe it would add more noise than true signal to the training set. The remaining 26 X-ray templates, despite evident structural similarities, were all kept since the computing effort for docking and GRIM rescoring remains quite negligible.**

For each of the 26 PDB templates, a protein-ligand interaction pattern graph was obtained with IChem [25] and a full pairwise similarity matrix was **generated**, considering GRIMscore as a similarity descriptor. The corresponding heat map (**Fig. 1**) clearly shows three groups according to the chemotypes of the bound ligands. Benzimidazoles appear to bind to FXR with a very homogeneous interaction pattern evidenced by very high pairwise similarities (GRIMscore > 1.0). For **isoxazoles**, the observed interaction patterns are clearly more diverse with a few highly similar complexes (e.g. 3rut, 3ruu, and 3p89; pairwise GRIMscore > 1.0) and much more complexes with still statistically significant interaction patterns but lower GRIMscores (between 0.7 and 1.0). The binding mode observed in the 3fxv entry seems to be unique to this group (**Fig. 1**). Last, the miscellaneous class is characterized by two different interaction patterns corresponding to FXR structures crystallized with either cholic acids (1osv, 1ot7\_A, **1ot7\_B**, 4qe6) or tetrahydroazepinoindoles (3fli, 111b; **Fig. 1**).

To ascertain that the 26 X-ray structures are indeed suitable templates to select high-quality poses for the 102 FXR agonists to dock within this challenge, we next cross-docked each of the 26 template ligands into the remaining 25 available protein template structures (self-docking was not permitted here). For each docking pose, the corresponding interaction pattern graph was compared and matched, with the GRIM alignment method [15], to the 25 interaction pattern graphs used as references. The benefit of rescoring Surflex poses by GRIM is nicely visible by comparing the root mean square deviations (rmsd) to the X-ray structure of the best predicted poses with or without GRIM rescoring (**Fig. 2**). Very high quality poses are picked by the GRIM scoring function for all 26 ligands (**Fig. 2**). With a single exception (PDB entry 1ot7, chain B), all corresponding ligands could be docked within 1.0 Å rmsd (median rmsd = 0.607 Å), thereby validating both the docking engine and the GRIM rescoring method as suitable for this particular dataset. Conversely, the best poses selected by the native Surflex score exhibit a higher and wider distribution of rmsd across the 26 reference ligands (**Fig. 2**). The GRIMscores associated to the GRIM-selected poses are also very high (median GRIMscore = 1.040) and demonstrate that the set of interaction patterns available in the PDB is sufficient to precisely pick high-quality docking solutions for previously uncharted ligands [30].

### **GRIM rescues Surflex for selecting accurate poses for 36 novel FXR agonists**

The first stage of the D3R Grand Challenge 2 aimed at predicting the binding mode of 36 novel FXR agonists (**Supplementary Table 1**) prior to the release of their FXR-bound X-ray coordinates. Out of the 36 agonists, 25 ligands exhibit chemotypes already crystallized in complex with FXR and thus in the list of 26 PDB templates (21 benzimidazoles, two isoxazoles, one tetrahydroazepinoindole, one cholic acid) whereas 9 ligands (3 spiro[indoline-3,4'-piperidine]-2-ones, 3 sulfonamides, 3 miscellaneous) exhibit chemotypes not yet crystallized with FXR. One compound (FXR\_33) was excluded from the dataset following organizers' request due to an oxidation of the compound during the co-crystallization process.

Following our strategy successfully applied in the previous challenge [16], all ligands were first docked with Surflex [24] to all PDB templates. We eliminated the poses with a Surflex score inferior to 2.0. The remaining poses were then converted into interaction pattern graphs and further compared to the 26 reference interaction patterns. For each ligand, the 5 poses with the highest GRIMscore values, whatever their initial Surflex rank, are last retained. The overall quality of the docking/rescoring protocol is judged by the rmsd of the ligand heavy atoms coordinates to the later released X-ray coordinates. GRIM rescoring clearly aids the Surflex docking engine to find high quality poses as evidenced by the mean rmsd which is lower for highest ranked GRIM poses, GRIM-1 (rmsd = 3.25 Å), with respect to highest ranked Surflex poses, Surflex-1 (rmsd = 4.12 Å; **Fig. 3**). As already observed in the previous challenge [16], the GRIM-1 pose is usually the most accurate. If one considers the absolute lowest rmsd pose out of the five proposed by GRIM, there is no significant decrease of the mean rmsd (3.08 vs. 3.25 Å; **Table 1, Fig. 3**), thereby confirming that GRIM rescoring does not necessitate to output more than a single solution. However, we have to admit that the obtained results are not entirely satisfactory since the mean rmsd is clearly above the threshold value (rmsd = 2.0 Å) generally considered as acceptable by the docking community. Considering the absolute lowest rmsd pose out of the 520 (20 x 26) generated for each ligand (**Table 1, Fig. 3**), we clearly evidence that the present challenge was not difficult from a pure docking viewpoint since the mean rmsd of the best possible pose is really low (rmsd = 1.34 Å; **Fig. 3**). The overall quality of the poses is certainly due to the use of multiple protein structures (26 in the present case) accounting for the ligand-dependent flexibility of the binding site. Scoring these excellent poses appears to be more challenging since neither GRIM nor the Surflex scoring function were able to rank them high enough to be selected. GRIM rescoring is however definitely better than relying on the original docking score provided by Surflex.

Analyzing the obtained rmsd values by chemotype, we clearly see that only agonists from the benzimidazole series were accurately docked with 18 out of 23 compounds posed with a rmsd below 2.0 Å (**Fig. 4**). None of the two isoxazoles (FXR\_4, FXR\_23) was properly docked although this series had already been co-crystallized with the target. Agonists bearing a new chemotype (sulfonamides,

spiro compounds, miscellaneous) were systematically misdocked, considering either GRIM or the Surflex scoring function. Only the novel tetrahydroazepinoindole (FXR\_5) was accurately posed (**Fig. 4**).

Since GRIM rescoring is a knowledge-based approach, we next looked at which preexisting protein-ligand interaction patterns have been used to select the final poses. As to be expected, GRIM has chosen available interaction patterns from the benzimidazole series to rescore docking poses of compounds sharing this chemotype (**Table 2**). Since this series adopt a very homogenous binding mode (**Fig.1**), it is therefore no surprise that most of the new compounds from this series were docked with low rmsd values. Similarly, the two isoxazoles were also rescored according to interaction patterns from the same chemical series. Nevertheless, our knowledge-based scoring approach failed in selecting accurate poses (**Table 2**). We previously demonstrated that known interaction patterns for this chemotype were more diverse (**Fig.1**), thereby explaining the difficulty to dock such compounds (**Table 2**). Interestingly, the low two-dimensional (2-D) similarity of the new isoxazoles to the existing templates (**Table 2**) may explain the hard attempt of GRIM to find out the existing good poses generated by Surflex (**Table 1**) but not scored high enough. For agonists exhibiting a novel chemotype (sulfonamides and spiro compounds), none of the existing interaction patterns was adequate for rescoring (**Table 2**). In most cases, these compounds were rescored considering only shape similarity to known templates, as evidenced by the Npol value, which accounts for the number of polar nodes in the selected alignment subgraph (**Table 2**). In GRIM, nodes have seven possible pharmacophoric properties corresponding to the encoded interaction type (hydrophobic, aromatic, hydrogen bond donor, hydrogen bond acceptor, negative ionisable, positive ionisable, and metal complexation). Whereas apolar nodes (hydrophobic, aromatic) describe mostly the shape of the bound ligand, polar nodes (remaining five properties) are placed at the location of polar protein-ligand interactions (hydrogen and ionic bonds), therefore increasing the confidence for the selected pose.

In the present case, the Npol descriptor exhibited quite often a null value, demonstrating that the GRIM alignment method has not found a single common protein-ligand polar interaction with the template. Among the miscellaneous series, only the novel tetrahydroazepinoindole FXR\_5 was accurately docked, by similarity to the interaction pattern observed for the 3fli template of the same series (**Table 2**). Last, the cholic acid analog FXR\_34 was misdocked, even if four interaction pattern templates from the same series were available (1osv, 1ot7\_A, 1ot7\_B, 4qe6) but not chosen by GRIM (**Table 2**).

### **Defining good practices for GRIM rescoring: Are docking failures predictable?**

To better delineate the applicability domain of GRIM rescoring, we plotted the observed rmsd of the top-1 GRIM pose with respect to three simple descriptors: GRIMscore, 2-D similarity to the X-ray ligand chosen for interaction template matching, and number of conserved polar interactions between the selected pose and the template pose (Npol parameter). The corresponding scatter plots (**Fig. 5**) clearly show some conditions that favor GRIM selection. They are here recapitulated by the following three rules:

*Rule 1: a very high GRIMscore ( $> 1.0$ ) generally leads to a low rmsd ( $< 2.0$ ). There are only four exceptions to this rule, one for which the GRIMscore is below the 1.0 threshold (FXR\_9) although the rmsd is less than 2.0, and three (FXR\_13, FXR\_18 and FXR\_35) for which the rmsd is higher than 2.0 although the GRIMscore is above 1.0. (**Fig. 5A**).*

*Rule 2: A high 2-D similarity to the template ligand ( $Tc > 0.6$ ) generally leads to a low rmsd ( $< 2.0$  Å).* This rule is verified for 31 out of 35 ligands analyzed. There are no cases for which GRIM could prioritize a good pose from a template in which the X-rayed ligand had a low 2-D similarity to the compound to dock (**Fig. 5B**). In four cases (FXR\_8, FXR\_13, FXR\_18, FXR\_35), a sufficiently high 2-D similarity to the template ligand could not guide GRIM to find out a good pose. When this 2-D similarity is low ( $< 0.60$ ), GRIM fails with no exception. Interestingly, such a strong 2-D chemical similarity to PDB templates was

not required to properly dock HSP90 $\alpha$  and MAP4K4 inhibitors in the previous D3R challenge [16]. This rule might thus be dataset-dependent.

*Rule 3: Good poses correspond to interaction patterns with at least two conserved polar interactions ( $N_{pol} \geq 2$ ).* In this challenge, we prioritized pose selection by the overall GRIMscore. This strategy clearly fails in case the common interaction pattern subgraph chosen for pose selection is lacking enough polar nodes ( $N_{Pol}$  parameter). Assuming only shape conservation ( $N_{pol} = 0$ ), there is only one ligand (FXR\_7) out of nine that is well docked by GRIM (**Fig. 5C**). In case  $N_{pol}$  is equal to 1 (one common FXR-agonist polar interaction), no docking success ( $rmsd < 2.0$ ) could be reported as well. It is only when  $N_{pol}$  is greater or equal to 2 that good rmsd values begin to be observed (**Fig. 5C**). There are only three ligands (FXR\_13, FXR\_34, FXR\_35) for which the rule is not verified despite a  $N_{pol}$  value of 3.0. These three rules clearly define a strict applicability domain for GRIM rescoring, at least for the present dataset of 36 agonists. New challenges will enable to precise whether these rules are transferable to new datasets or not.

### Reasons for failure

After the submission of the first set of predictions (stage1: prediction of up to five poses for each ligand), X-ray structure of the 36 new agonists in complex with FXR were released. The comparison of the corresponding protein-ligand interaction patterns with that of our 26 PDB templates illustrates the difficulty of the present challenge (**Fig. 6A**). With the exception of all benzimidazoles and one tetrahydroazepinoindole (FXR\_5), all novel FXR agonists adopt an unprecedented binding mode (**Fig. 6A**). A molecular explanation to this observation lies in the large (750-1025 Å<sup>3</sup>) and hydrophobic cavity in FXR that is able to accommodate very different and almost non-overlapping agonists. Despite these significant hurdles, observed failures to predict binding modes of these ligands are not due to real docking issues. Among the set of proposed poses by Surflex, true docking failures (Surflex-best rmsd < 3.0 Å) were only reported for three ligands (FXR\_1, FXR\_18 and FXR\_23). The case of the isoxazole

FXR\_23 is perfectly illustrative. The top-1 GRIM pose (**Fig. 6B**) was generated by analogy of the corresponding interaction pattern found on the 3hc5 template of the same chemical series, nicely overlapping the aryl-isoxazole moieties of both compounds. Surprisingly, the disclosed X-ray pose of FXR\_23 departs considerably from that already seen for FXR isoxazole agonists (**Fig. 6B**). None of the solutions generated by Surflex are similar to this unexpected pose. It is interesting to notice that two out of these three ligands are among the less potent agonists of the dataset, with *in vitro* binding affinities in the micromolar range.

GRIM failed to propose reliable poses for three benzimidazoles (FXR\_8, FXR\_13, FXR\_35) although they exhibit the canonical benzimidazole-binding mode (**Fig. 6A**) and despite the existence of suitable poses generated by Surflex (Surflex-best rmsd < 2.0 Å, **Table 1**). The overall cross-like shape of the ligands as well as the exact binding amino acids are found, but the benzimidazole ring that switched to the area normally occupied by a cyclohexyl ring (**Fig. 6C**). In these examples, the GRIM score is dominated by the conservation of apolar contacts (in other words, the overall ligand shape) at the cost of an important hydrogen bond between the benzimidazole and Tyr373 side chain. The same observation applies to most of the scoring failures for which the GRIM-selected pose has been dominated by shape conservation (**Fig. 5C**). The herein described pose selection protocol, biased towards the highest GRIMscore is therefore not suited for hydrophobic ligands binding to predominantly apolar cavities. Removing poses with no common polar interactions to the chosen template (Npol = 0) before final GRIM ranking should avoid some of the above reported failures in the future.

Altogether, combining Surflex for pose generation and GRIM for ranking remains a robust docking strategy. When compared to the 51 other contributions to the first phase of the D3R Grand Challenge 2, the Surflex/GRIM protocol lies among the top contributions sharing a posing accuracy in the 2-3 Å rmsd range (**Fig. 7**).



## Ranking GRIM poses by the HYDE scoring function: Estimating the binding affinity of 102 FXR agonists.

In a second step, the binding affinity of 102 FXR agonists (**Supplementary Table 2**), including the 36 ligands (FXR\_1-FXR\_36) considered for the pose prediction phase and 66 additional ligands (FXR\_37-FXR\_102), was predicted and compared to the experimental affinity data released once the challenge was closed. The GRIM-1 pose of each ligand was thus rescored with the HYDE scoring function [26]. This scoring function was chosen because of the very careful estimation of the desolvation contributions to the overall binding free energy of a wide array of diverse protein-ligand complexes. Importantly, HYDE is not **parametrized against any particular training set** and provides a balanced assessment of the energetics of desolvation considering three atomic physicochemical properties: local hydrophobicity, solvent accessible surface, and contact surface area. When applied to the FXR dataset of 102 ligands, HYDE scoring of the GRIM-1 pose yielded the third most accurate ranking with a Kendall's  $\tau$  ranking coefficient of 0.442 (**Fig. 8A**) and a Pearson's  $\rho$  correlation coefficient of 0.593 (**Fig. 8B**). Predicted HYDE binding free energies were systematically higher than that derived from in vitro experimental binding affinity data but an overall trend could be confirmed. The current performance in affinity ranking is slightly better than that obtained for the previous D3R challenge for two different datasets (180 HSP90 $\alpha$  inhibitors, 18 MAP4K4 inhibitors) [13], but far from being usable for hit to lead optimization, thereby confirming most previous attempts to accurately predict binding free energies for a set of heterogeneous compounds [31,11]. Since the observed ranking and correlation coefficients were not improved for the subset of benzimidazoles for which docking poses are of **good** quality, consistent failures in predicting binding free energies cannot be attributed to error in atomic coordinates but merely to inconsistent treatment of energetics.

## Conclusions

We have herein applied a knowledge-based graph matching method (GRIM) to rank docking poses of 36 novel FXR agonists. Although the current poses have been generated by Surflex, the rescoring protocol is fully independent on any docking tool, at the condition that correct MOL2 file formats can be generated. This second D3R docking challenge was significantly more demanding than the previous one for the major reason that the large and apolar ligand-binding site of the chosen target (FXR receptor) is able to accommodate chemically heterogeneous ligands in many different ways. Docking predictions were reasonably accurate (mean rmsd = 1.97 Å) for the class of 21 benzimidazoles already co-crystallized with FXR, which exhibited a conserved binding mode. For other chemical classes, the existence (isoxazoles) or not (spiroindolines and sulfonamides) of known similar PDB templates, did not impact GRIM posing, which was very unsatisfactory (rmsd above 4 Å). Several reasons could be invoked to explain these systematic rescoring failures: (i) the inability of the docking engine to provide a single accurate pose for further rescoring, (ii) the impossibility to predict completely novel binding modes, (iii) the bias in GRIM rescoring towards poses with the highest protein-ligand interaction pattern similarity to that of existing PDB templates, thereby favoring quite often shape matching for ligands with a predominant hydrophobic character.

However, the current challenge has taught us interesting rules to improve future predictions. First, the obtained data confirm the well-known bias of fast energy-based scoring functions to rank improper poses among the top-ranked solutions. Any knowledge-based rescoring scheme [32-35] similar in spirit to GRIM is therefore preferable to relying only on docking scores. In our current rescoring protocol, poses with very low predicted binding energies (predicted pkd < 2) were filtered out and not subjected to GRIM rescoring. Later, the comparison of the experimental solutions with our predictions surprisingly identified several cases where reasonably good poses (rmsd to X-ray pose < 2 Å) were rejected for the simple reason that the predicted pkd was below 2. We will therefore remove this filter in our next rescoring protocol.

Second, GRIM rescoring relies on the existence of a similar protein-ligand interaction **pattern** in the PDB. For unbalanced interaction patterns dominated by apolar contacts, we identified a bias in prioritizing shape overlap over the conservation of a few but very important hydrogen bonds or salt bridges. As a workaround, we thus propose to eliminate from the set of GRIM selections, any pose for which the number of aligned graph polar nodes to the chosen template is below a value of 2.

Third, rescoring the GRIM highest ranked pose using the HYDE scoring function provided one of the most accurate affinity ranking strategy submitted to this challenge. Due to the overall speed of the procedure, we therefore propose to apply this rescoring protocol to virtual screening where GRIM is used to prioritize a few thousand ligands to be further rescored with HYDE within a reasonable amount of cpu time (ca. 10 s/compound).

GRIM presents several advantages over alternative knowledge-based rescoring strategies: (i) it can be coupled to any docking algorithm, (ii) it does not constrain ligand docking but rewards interaction patterns already present among PDB templates, (iii) it takes advantage of ligands with similar binding modes and not necessarily similar chemical structures [16], (iv) it can be applied in a target family-biased pose selection process in which PDB templates from the same protein but also from similar targets can be used to store reference interaction patterns, (v) it permits to directly quantify binding mode similarity between a predicted protein-ligand complex and any PDB template at a very high throughput.

GRIM is embedded in the IChem toolkit and available for non-profit research at <http://bioinformatics.u-strasbf.fr/labwebsite/download.html>.

### **Supporting information.**

List of 102 FXR agonists, List of FXR-agonist PDB templates.

**Acknowledgements.**

We thank the CAPES foundation and the Ministry of Education of Brazil for a postdoctoral fellowship to P.S.F.C.G (Computational Biology program grant, issue n°51/2013). The Computing Center of the IN2P3 (CNRS, Villeurbanne, France) is acknowledged for allocation of computing time and excellent support.

**Table 1.** Accuracy of pose selection (rmsd in Å to X-ray solution) for 36 FXR agonists

Compound	rmsd to X-ray, Å						
	GRIM-1 <sup>a</sup>	GRIM-2 <sup>b</sup>	GRIM-3 <sup>c</sup>	GRIM-4 <sup>d</sup>	GRIM-5 <sup>e</sup>	Surflex-1 <sup>f</sup>	Surflex-best <sup>g</sup>
FXR_1	4.00	4.14	5.84	4.77	3.90	3.53	3.08
FXR_2	7.31	7.63	7.77	7.67	8.04	7.10	1.03
FXR_3	7.76	4.93	8.13	5.94	4.75	6.59	2.16
FXR_4	3.89	6.50	3.92	6.23	7.35	7.24	1.18
FXR_5	0.48	0.69	0.51	0.68	0.63	7.50	0.48
FXR_6	0.75	0.73	0.70	0.76	0.69	1.86	0.70
FXR_7	1.11	1.10	1.30	1.21	n.a. <sup>h</sup>	1.52	1.10
FXR_8	3.88	3.88	6.45	3.87	5.06	1.50	0.98
FXR_9	1.19	4.78	4.77	6.93	0.89	1.47	0.89
FXR_10	4.81	4.52	4.80	4.44	n.a.	4.79	1.34
FXR_11	4.50	2.69	4.48	3.38	4.75	5.05	2.16
FXR_12	8.66	8.60	8.68	8.85	8.44	6.90	1.29
FXR_13	4.61	4.60	7.78	4.61	4.66	1.66	1.66
FXR_14	0.61	0.65	1.57	0.73	1.46	1.28	0.61
FXR_15	7.81	4.66	5.08	7.80	7.61	7.91	2.02
FXR_16	5.78	5.78	4.32	5.61	7.95	7.28	1.11
FXR_17	2.69	5.82	6.14	8.63	6.13	2.80	1.15
FXR_18	7.57	9.09	7.74	7.56	7.60	7.26	4.18
FXR_19	1.17	0.79	1.20	4.94	1.20	1.56	0.79
FXR_20	0.73	1.17	0.93	n.a.	n.a.	2.20	0.73
FXR_21	0.97	0.77	0.76	0.77	0.77	1.53	0.77
FXR_22	0.79	0.90	2.00	0.68	2.04	1.70	0.68
FXR_23	7.82	7.81	6.82	6.49	n.a.	7.02	2.92
FXR_24	0.74	1.39	1.41	1.44	n.a.	2.19	0.74
FXR_25	1.73	1.73	1.73	5.04	n.a.	5.21	1.42
FXR_26	1.32	1.35	1.36	1.34	1.34	1.68	1.32
FXR_27	0.50	0.96	0.76	0.73	0.67	1.80	0.50
FXR_28	0.62	0.56	0.94	1.15	0.97	2.25	0.56
FXR_29	0.79	0.84	0.85	0.84	0.82	1.64	0.79
FXR_30	1.94	1.98	1.97	1.84	1.83	6.07	1.42
FXR_31	1.82	1.15	1.98	1.17	1.93	2.16	1.15
FXR_32	2.07	2.06	2.08	2.07	2.06	2.52	1.43
FXR_34	5.41	4.42	4.48	3.91	n.a.	12.82	2.38
FXR_35	6.93	6.74	1.10	6.81	4.51	7.04	1.10
FXR_36	1.25	1.16	1.10	1.20	6.79	1.79	1.10
Mean rmsd	3.25	3.33	3.47	3.82	3.98	4.12	1.34
median rmsd	1.94	2.06	2.00	3.89	2.57	2.51	1.11

<sup>a</sup> 1<sup>st</sup> ranked pose according to GRIMscore

<sup>b</sup> 2<sup>nd</sup> ranked pose according to GRIMscore

<sup>c</sup> 3<sup>rd</sup> ranked pose according to GRIMscore

<sup>d</sup> 4<sup>th</sup> ranked pose according to GRIMscore

<sup>e</sup> 5<sup>th</sup> ranked pose according to GRIMscore

<sup>f</sup> 1<sup>st</sup> ranked pose according to Surflex score

<sup>g</sup> lowest rmsd pose out of 20 Surflex solutions.

<sup>h</sup> not available

**Table 2.** Characteristics of GRIM top-ranked docking poses for 35 FXR agonists

Ligand	Protein <sup>a</sup>	Pose <sup>b</sup>	Template <sup>c</sup>	GRIMscore <sup>d</sup>	Npol <sup>e</sup>	Tc <sup>f</sup>	rmsd <sup>g</sup>
FXR_1	3olf	6	3olf	0.90	1	0.32	4.00
FXR_2	3olf	7	3olf	0.78	1	0.55	7.31
FXR_3	3oki	15	3oki	0.93	0	0.55	7.76
FXR_4	3rut	8	3rut	0.84	0	0.42	3.89
FXR_5	3fli	3	3fli	1.11	3	0.77	0.48
FXR_6	3oki	11	3oki	1.15	3	0.67	0.75
FXR_7	3ook	0	3ook	1.06	0	0.66	1.11
FXR_8	3oki	10	3oki	0.94	0	0.76	3.88
FXR_9	3oof	2	3oof	0.93	3	0.59	1.19
FXR_10	1osh	3	1osh	0.80	0	0.34	4.81
FXR_11	3oof	4	3oof	0.81	1	0.51	4.50
FXR_12	3olf	3	3olf	0.84	0	0.50	8.66
FXR_13	3oof	5	3oof	1.03	3	0.75	4.61
FXR_14	3oki	13	3oki	1.25	3	0.63	0.61
FXR_15	3ook	15	3ook	0.84	0	0.53	7.81
FXR_16	3omk	1	3omk	0.80	0	0.43	5.78
FXR_17	3oof	7	3oof	0.80	3	0.46	2.69
FXR_18	3omk	6	3omk	1.02	0	0.60	7.57
FXR_19	3oki	8	3oki	1.10	3	0.76	1.17
FXR_20	3olf	0	3omk	1.06	6	0.69	0.73
FXR_21	3ook	18	3ook	1.19	3	0.85	0.97
FXR_22	3oki	3	3oki	1.12	3	0.78	0.79
FXR_23	3ruu	7	3hc5	0.87	1	0.58	7.82
FXR_24	3oof	1	3oof	1.17	6	0.85	0.74
FXR_25	3omk	1	3omk	1.08	3	0.82	1.73
FXR_26	3oof	6	3oof	1.10	3	0.72	1.32
FXR_27	3ook	3	3ook	1.39	5	0.98	0.50
FXR_28	3ook	18	3ook	1.39	9	0.67	0.62
FXR_29	3ook	10	3ook	1.42	9	0.67	0.79
FXR_30	3olf	17	3olf	1.07	9	0.75	1.94
FXR_31	3oof	2	3oof	1.08	3	0.68	1.82
FXR_32	3ook	14	3ook	1.08	2	0.74	2.07
FXR_34	3olf	0	3olf	0.89	3	0.47	5.41
FXR_35	3oof	0	3oof	1.13	3	0.63	6.93
FXR_36	3olf	11	3olf	1.30	12	0.78	1.25

<sup>a</sup> set of protein coordinates used for docking<sup>b</sup> Surflex pose number<sup>c</sup> set of protein-ligand coordinates used as template for graph matching<sup>d</sup> GRIMscore<sup>e</sup> number of polar nodes in the clique<sup>f</sup> 2-D chemical similarity (Tanimoto coefficient) between query and template ligands, calculated from 166-bit MDL public keys.<sup>g</sup> root mean square deviations (in Å) from X-ray pose

## Figure legends

**Fig. 1.** Heat map of protein-ligand interaction pattern similarity (GRIMscore) among 26 reference FXR-agonist X-ray structures. Entries are grouped according to the three chemotypes (benzimidazoles, isoxazoles, miscellaneous) of the bound FXR-agonists. Pairwise similarity GRIMscores are calculated with IChem v5.2.6, and colored as follows: blue, no similarity (GRIMscore <0.7); yellow, medium similarity (0.7 <GRIMscore < 1.0); red, high similarity (GRIMscore > 1.0)

**Fig.2** Distribution of root mean square deviations (in Å) from the true X-ray pose, for 26 reference FXR agonists docked to 25 FXR X-ray structures. Poses were scored according to GRIMscore or the native Surflex scoring function. The boxes delimit the 25th and 75th percentiles, the whiskers delimit the 5th and 95th percentiles. The median and mean values are indicated by a horizontal line and a square, respectively. Crosses delimit the 1% and 99th percentiles, respectively. Minimum and maximum values are indicated by a dash.

**Fig.3** Mean root mean square deviations (in Å) from the true X-ray pose, for 35 new FXR agonists docked to 26 FXR X-ray structures. Poses were scored according to GRIMscore or the native Surflex scoring function. Top ranked poses by GRIM and Surflex are labelled GRIM-1 and Surflex-1, respectively. The lowest rmsd pose identified by GRIM and Surflex are labelled GRIM-best and Surflex-best, respectively. The median values are indicated by a horizontal line, the standard deviation to the mean by a whisker.

**Fig. 4.** Root mean square deviations (in Å) of the GRIM-1 pose from the true X-ray pose for 35 novel FXR agonists.

**Fig. 5.** Variation of the root mean square deviations (in Å) to the X-ray poses for 35 FXR agonists as a function of three properties: **(A)** GRIMscore of the pose, **(B)** 2-D similarity of the FXR agonist to the FXR-bound template ligand used by GRIM, **(C)** number of polar nodes (polar interactions, Npol) in the clique common to the protein-FXR ligand docked pose and the protein-ligand PDB template chosen by GRIM for pose selection. Docking poses complying to a precise rule are enclosed by a red circle.

**Fig. 6.** Predicted vs. X-ray protein-ligand interaction patterns for 35 FXR agonists. **A)** *A posteriori* observed binding mode similarity of 35 novel FXR agonists to 26 known PDB ligands, as a function of the root mean square deviations (in Å) of the GRIM-1 pose to the true X-ray structure. For each novel FXR-agonist X-ray structure, the highest GRIMscore considering all 26 PDB templates is used to quantify binding mode similarity. **B)** Predicted (cyan sticks) versus X-ray pose (tan sticks) of the agonist FXR\_23 bound to the FXR receptor (white ribbons). GRIM-1 pose was obtained by docking FXR\_23 to the 3ruu atomic coordinates and selected according to the similarity of its interaction pattern to that of the 3hc5 ligand (light green sticks). Heteroatoms are colored in blue (nitrogen), red (oxygen), yellow (sulfur) and dark green (chlorine). **C)** Predicted (cyan sticks) versus X-ray pose (tan sticks) of the agonist FXR\_8 bound to the FXR receptor (white ribbons). GRIM-1 pose was obtained by docking FXR\_8 to the 3oki atomic coordinates and selected according the similarity of its interaction pattern to that of the 3oki ligand (light green sticks). Heteroatoms are colored in blue (nitrogen) and red (oxygen)

**Fig. 7.** Accuracy of 52 contributions to the stage 1 (pose prediction) of the D3R Grand Challenge 2. The accuracy of each contribution is described by the mean rmsd of the highest-ranked pose to the X-ray



coordinates of 35 test FXR agonists, released after closing the challenge. Incomplete submissions are indicated by white bars.

**Fig. 8.** Assessment of the GRIM-HYDE rescoring protocol for ranking and affinity prediction of 102 FXR agonists. **A)** Comparison of the GRIM-HYDE protocol with respect to 57 competing contributions in this challenge, estimated by the Kendall's  $\tau$  ranking coefficient. Error bars were obtained by re-computing all statistics in 10,000 rounds of resampling with replacement, where, in each sample, the experimental  $IC_{50}$  data were randomly modified based on the experimental uncertainties (data provided by the organizers). **B)** Predicted ( $\Delta G_{pred}$ ) vs. experimental ( $\Delta G_{exp}$ ) absolute binding free energies in kJ/mol. The red line indicates a linear regression fit of predicted to experimental values.

## References

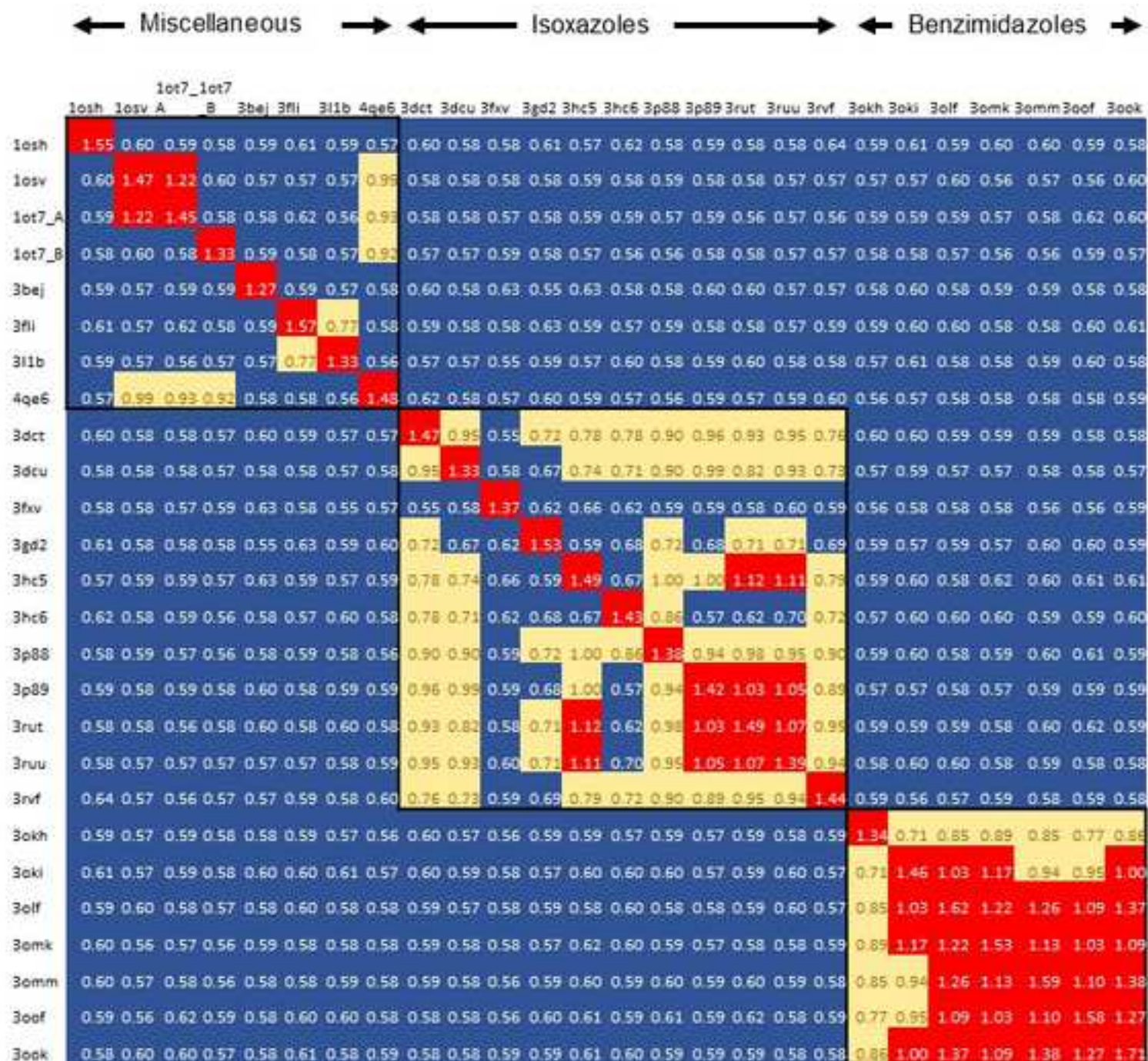
1. Brooijmans N, Kuntz ID (2003) Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* 32:335-373.
2. Rognan D (2017) The impact of in silico screening in the discovery of novel and safer drug candidates. *Pharmacol Ther* 175:47-66.
3. Chen YC (2015) Beware of docking! *Trends Pharmacol Sci* 36:78-95
4. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* 49:5912-5931
5. Smith RD, Dunbar JB, Jr., Ung PM, Esposito EX, Yang CY, Wang S, Carlson HA (2011) CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. *J Chem Inf Model* 51:2115-2131
6. Community Structure Activity resource, <http://www.csardock.org>
7. Drug Design Data Resource, <https://drugdesigndata.org/about/grand-challenge-2>
8. Dunbar JB, Jr., Smith RD, Yang CY, Ung PM, Lexa KW, Khazanov NA, Stuckey JA, Wang S, Carlson HA (2011) CSAR benchmark exercise of 2010: selection of the protein-ligand complexes. *J Chem Inf Model* 51:2036-2046
9. Damm-Ganamet KL, Smith RD, Dunbar JB, Jr., Stuckey JA, Carlson HA (2013) CSAR benchmark exercise 2011-2012: evaluation of results from docking and relative ranking of blinded congeneric series. *J Chem Inf Model* 53:1853-1870
10. Dunbar JB, Jr., Smith RD, Damm-Ganamet KL, Ahmed A, Esposito EX, Delproposito J, Chinnaswamy K, Kang YN, Kubish G, Gestwicki JE, Stuckey JA, Carlson HA (2013) CSAR data set release 2012: ligands, affinities, complexes, and docking decoys. *J Chem Inf Model* 53:1842-1852
11. Carlson HA (2016) Lessons Learned over Four Benchmark Exercises from the Community Structure-Activity Resource. *J Chem Inf Model* 56:951-954

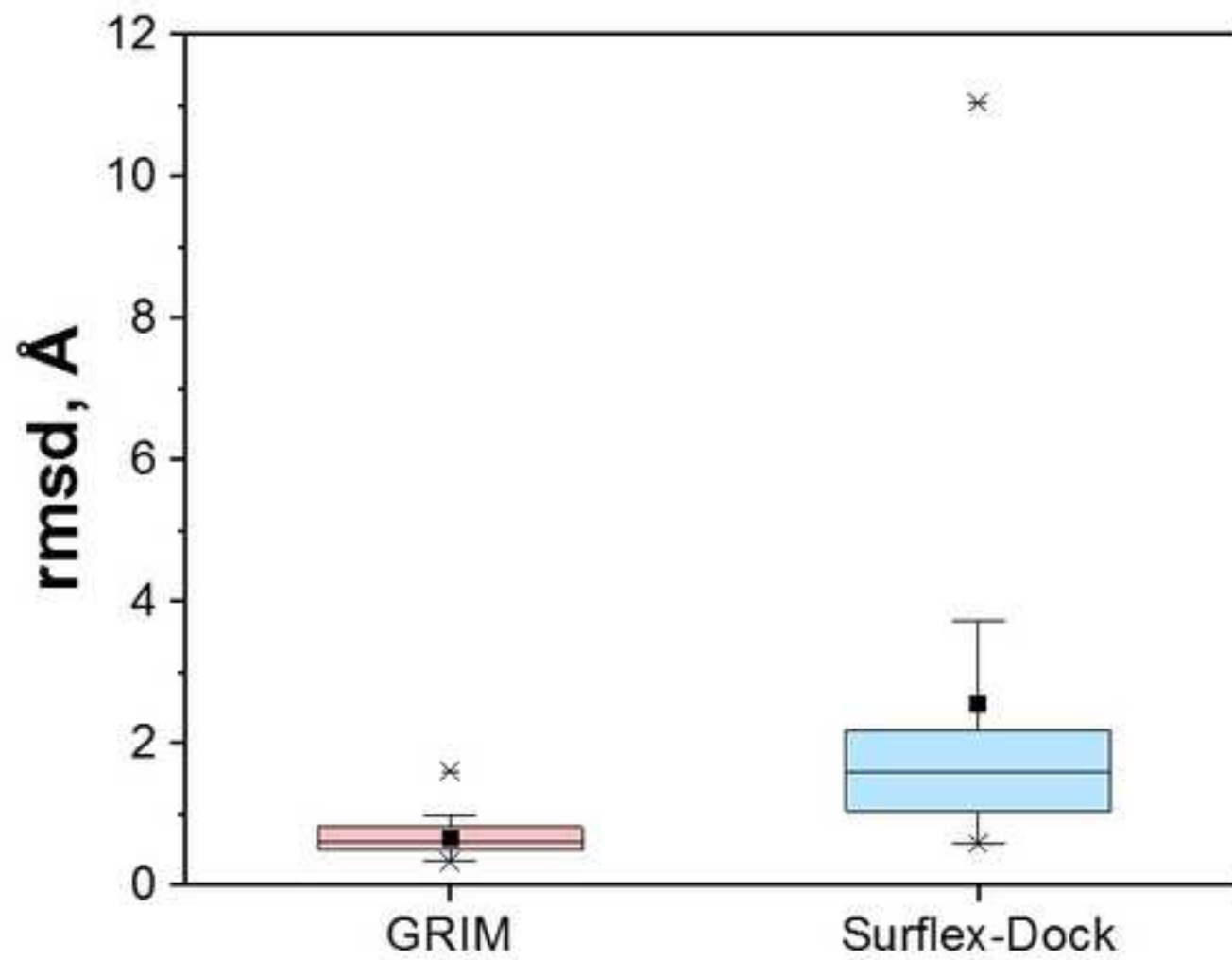
12. Carlson HA, Smith RD, Damm-Ganamet KL, Stuckey JA, Ahmed A, Convery MA, Somers DO, Kranz M, Elkins PA, Cui G, Peishoff CE, Lambert MH, Dunbar JB, Jr. (2016) CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. *J Chem Inf Model* 56:1063-1077
13. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposto J, Kubish G, Dunbar JB, Jr., Carlson HA, Burley SK, Walters WP, Amaro RE, Feher VA, Gilson MK (2016) D3R grand challenge 2015: Evaluation of protein-ligand pose and affinity predictions. *J Comput Aided Mol Des* 30:651-668
14. Smith RD, Damm-Ganamet KL, Dunbar JB, Jr., Ahmed A, Chinnaswamy K, Delproposto JE, Kubish GM, Tinberg CE, Khare SD, Dou J, Doyle L, Stuckey JA, Baker D, Carlson HA (2016) CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge. *J Chem Inf Model* 56:1022-1031
15. Desaphy J, Raimbaud E, Ducrot P, Rognan D (2013) Encoding protein-ligand interaction patterns in fingerprints and graphs. *J Chem Inf Model* 53:623-637
16. Slynko I, Da Silva F, Bret G, Rognan D (2016) Docking pose selection by interaction pattern graph similarity: application to the D3R grand challenge 2015. *J Comput Aided Mol Des* 30:669-683
17. Richter HG, Benson GM, Bleicher KH, Blum D, Chaput E, Clemann N, Feng S, Gardes C, Grether U, Hartman P, Kuhn B, Martin RE, Plancher JM, Rudolph MG, Schuler F, Taylor S (2011) Optimization of a novel class of benzimidazole-based farnesoid X receptor (FXR) agonists to improve physicochemical and ADME properties. *Bioorg Med Chem Lett* 21:1134-1140
18. Bass JY, Caravella JA, Chen L, Creech KL, Deaton DN, Madauss KP, Marr HB, McFadyen RB, Miller AB, Mills WY, Navas F, 3rd, Parks DJ, Smalley TL, Jr., Spearing PK, Todd D, Williams SP, Wisely GB (2011) Conformationally constrained farnesoid X receptor (FXR) agonists: heteroaryl replacements of the naphthalene. *Bioorg Med Chem Lett* 21:1206-1213
19. 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

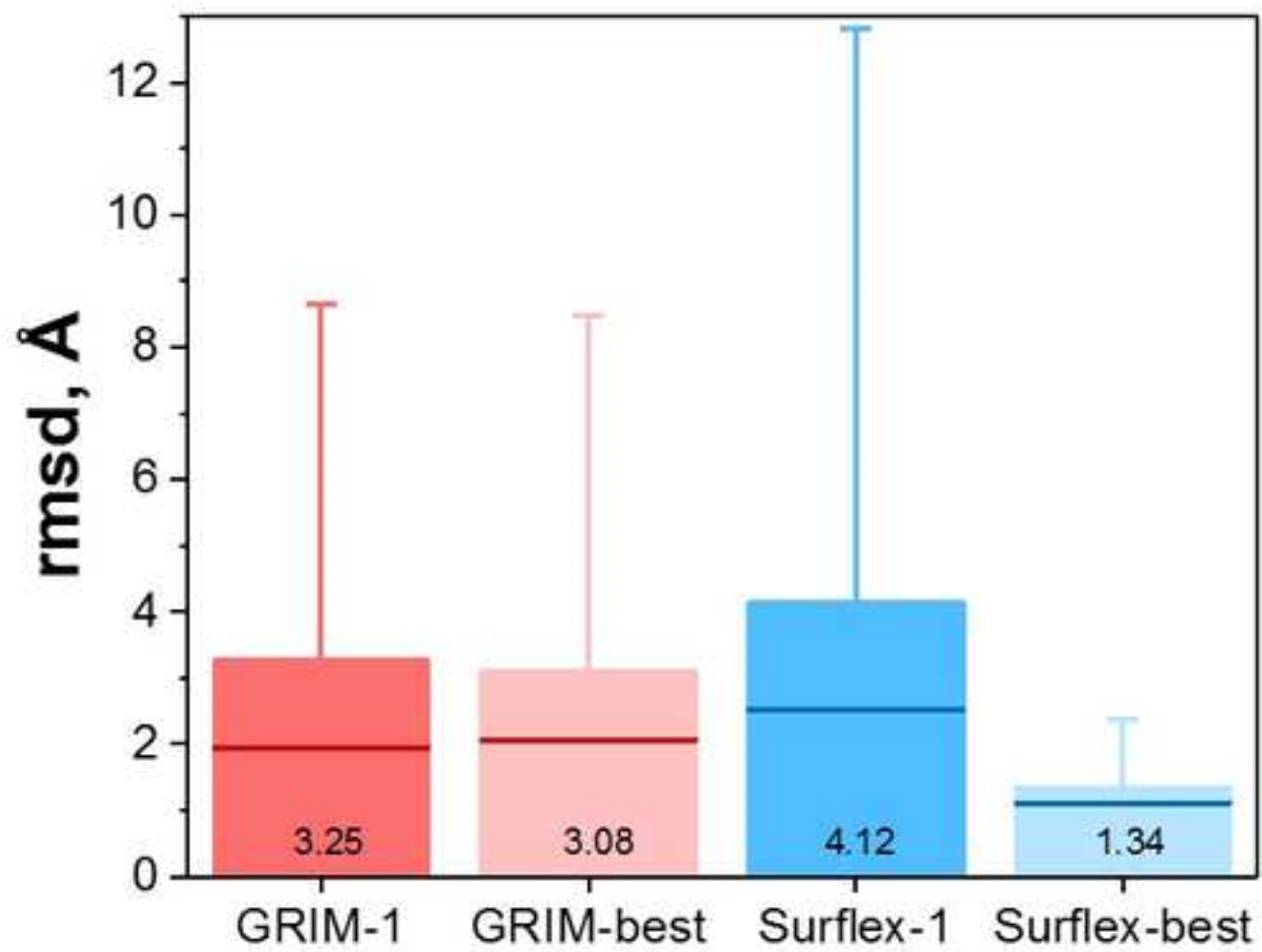
20. Bietz S, Urbaczek S, Schulz B, Rarey M (2014) Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. *J Cheminform* 6:12
21. Certera, L.P., Princeton, NJ 08540, U.S.A.
22. Molecular Networks GmbH, Erlangen, Germany
23. OpenEye Scientific Software, Santa Fe, NM, U.S.A.
24. Jain AN (2007) Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J Comput Aided Mol Des* 21:281-306
25. Desaphy J, Da Silva F, Rognan D IChem v.5.2.6: <http://bioinfo-pharma.u-strasbg.fr/labwebsite/downloads>.
26. Schneider N, Lange G, Hindle S, Klein R, Rarey M (2013) A consistent description of HYdrogen bond and DEhydration energies in protein-ligand complexes: methods behind the HYDE scoring function. *J Comput Aided Mol Des* 27:15-29
27. BioSolveIT GmbH, Sankt Augustin, Germany.
28. Xu X, Xu X, Liu P, Zhu ZY, Chen J, Fu HA, Chen LL, Hu LH, Shen X (2015) Structural Basis for Small Molecule NDB (N-Benzyl-N-(3-(tert-butyl)-4-hydroxyphenyl)-2,6-dichloro-4-(dimethylamino) Benzamide) as a Selective Antagonist of Farnesoid X Receptor alpha (FXRalpha) in Stabilizing the Homodimerization of the Receptor. *J Biol Chem* 290:19888-19899
29. Jin L, Feng X, Rong H, Pan Z, Inaba Yu, Qiu L, Zheng W, Lin S, Wang R, Wang Z, Wang S, Liu H, Li S, Xie W, Li Y (2013) The antiparasitic drug ivermectin is a novel FXR ligand that regulates metabolism. *Nature Commun* 4:1937
30. Previous experiences suggest that GRIMscore values above 0.70 correspond to a statistically significant similarity of the two interaction patterns under comparison.
31. Kim R, Skolnick J (2008) Assessment of programs for ligand binding affinity prediction. *J Comput Chem* 29:1316-1331
32. Gao C, Thorsteinson N, Watson I, Wang J, Vieth M (2015) Knowledge-Based Strategy to Improve Ligand Pose Prediction Accuracy for Lead Optimization. *J Chem Inf Model* 55:1460-1468

33. Kelley BP, Brown SP, Warren GL, Muchmore SW (2015) POSIT: Flexible Shape-Guided Docking For Pose Prediction. *J Chem Inf Model* 55:1771-1780
34. Anighoro A, Bajorath J (2016) Three-Dimensional Similarity in Molecular Docking: Prioritizing Ligand Poses on the Basis of Experimental Binding Modes. *J Chem Inf Model* 56:580-587
35. Kumar A, Zhang KY (2016) Application of Shape Similarity in Pose Selection and Virtual Screening in CSARdock2014 Exercise. *J Chem Inf Model* 56:965-973

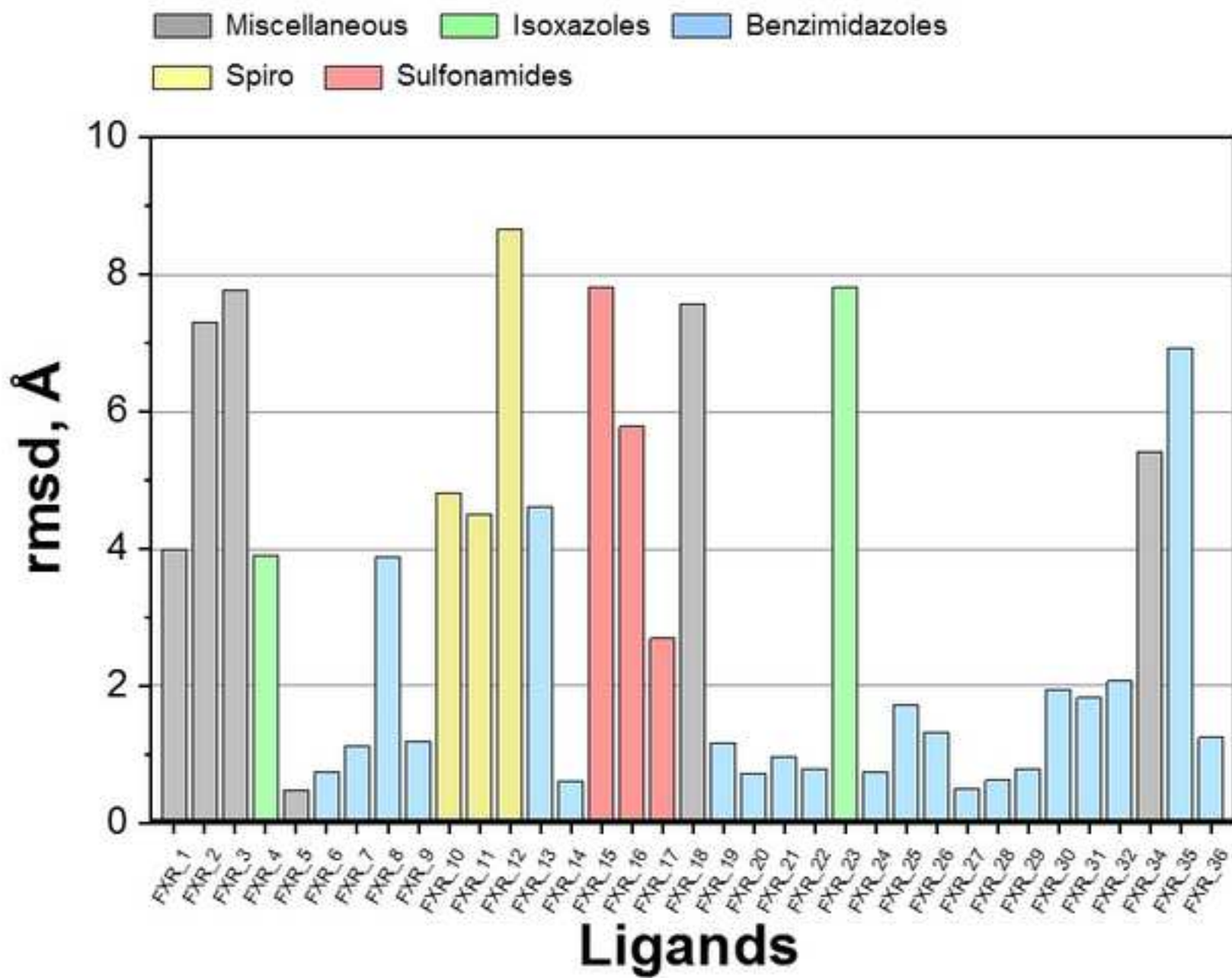
Figure 1

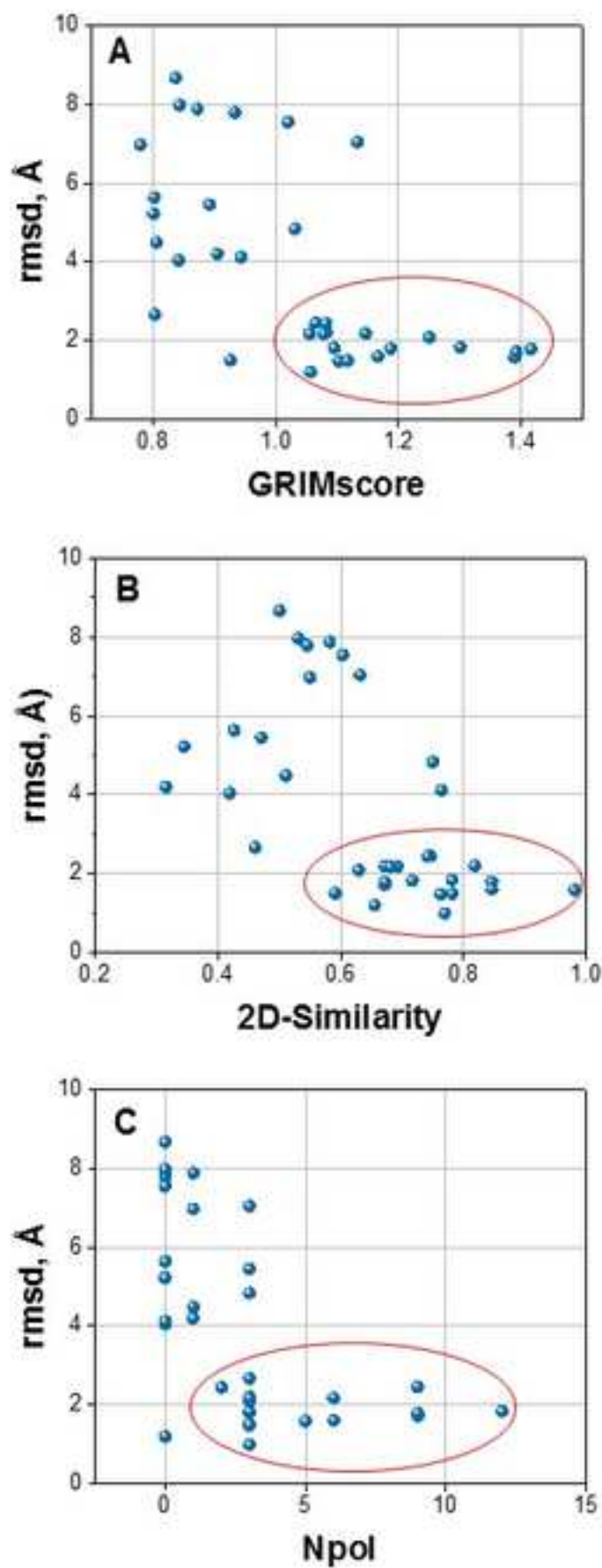
[Click here to download Figure Fig1.jpg](#)

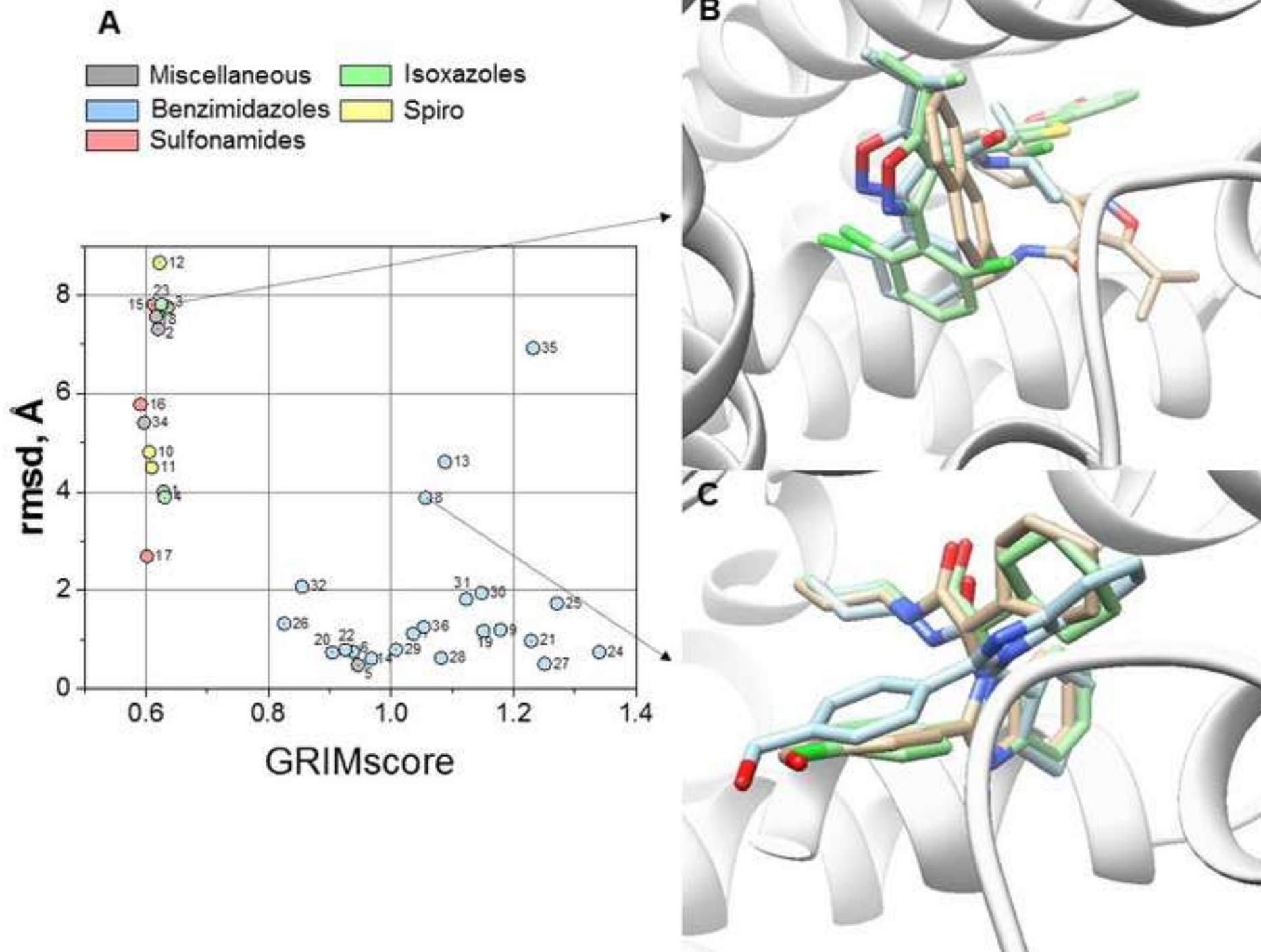


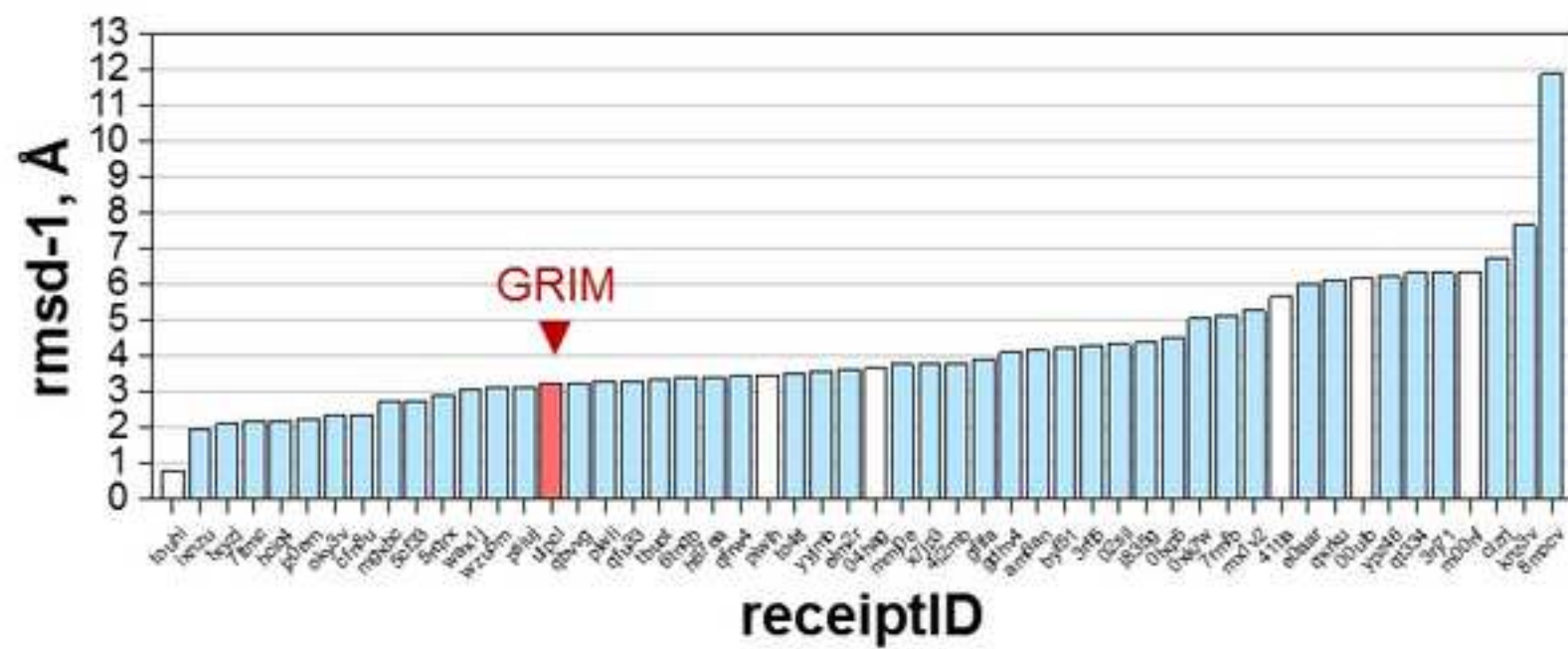


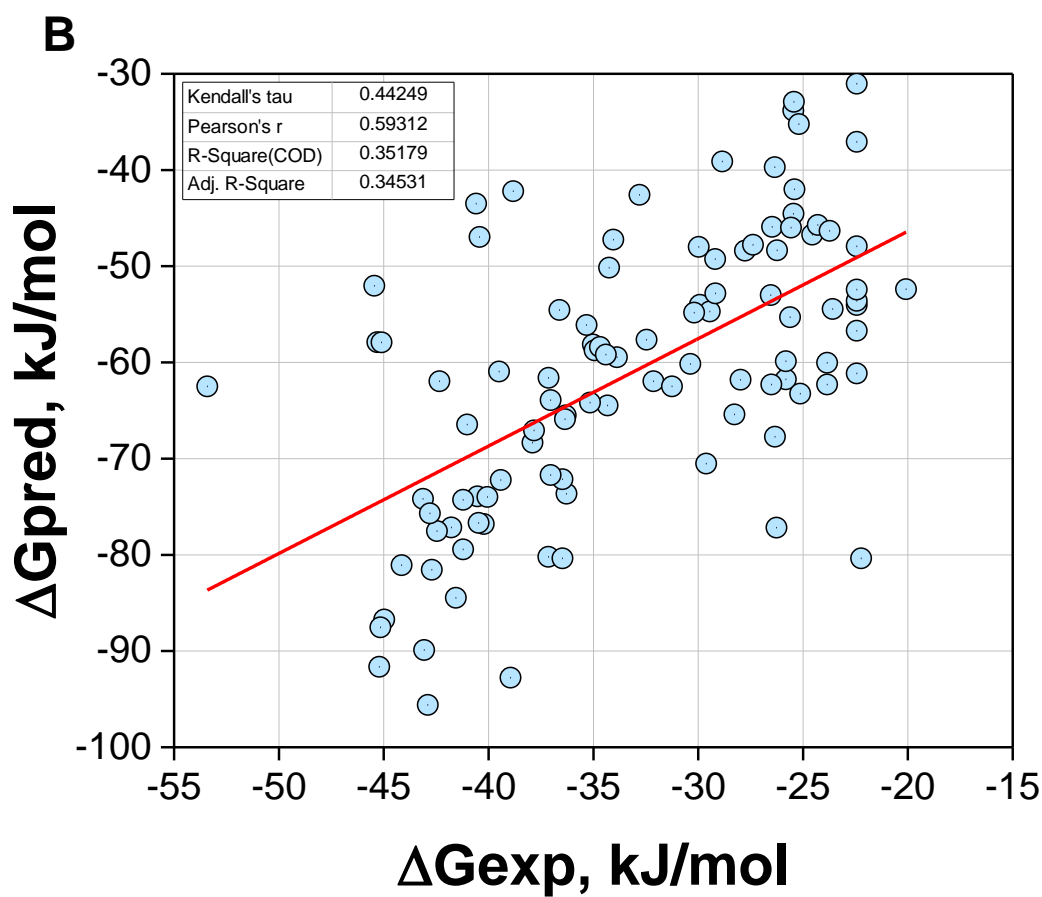
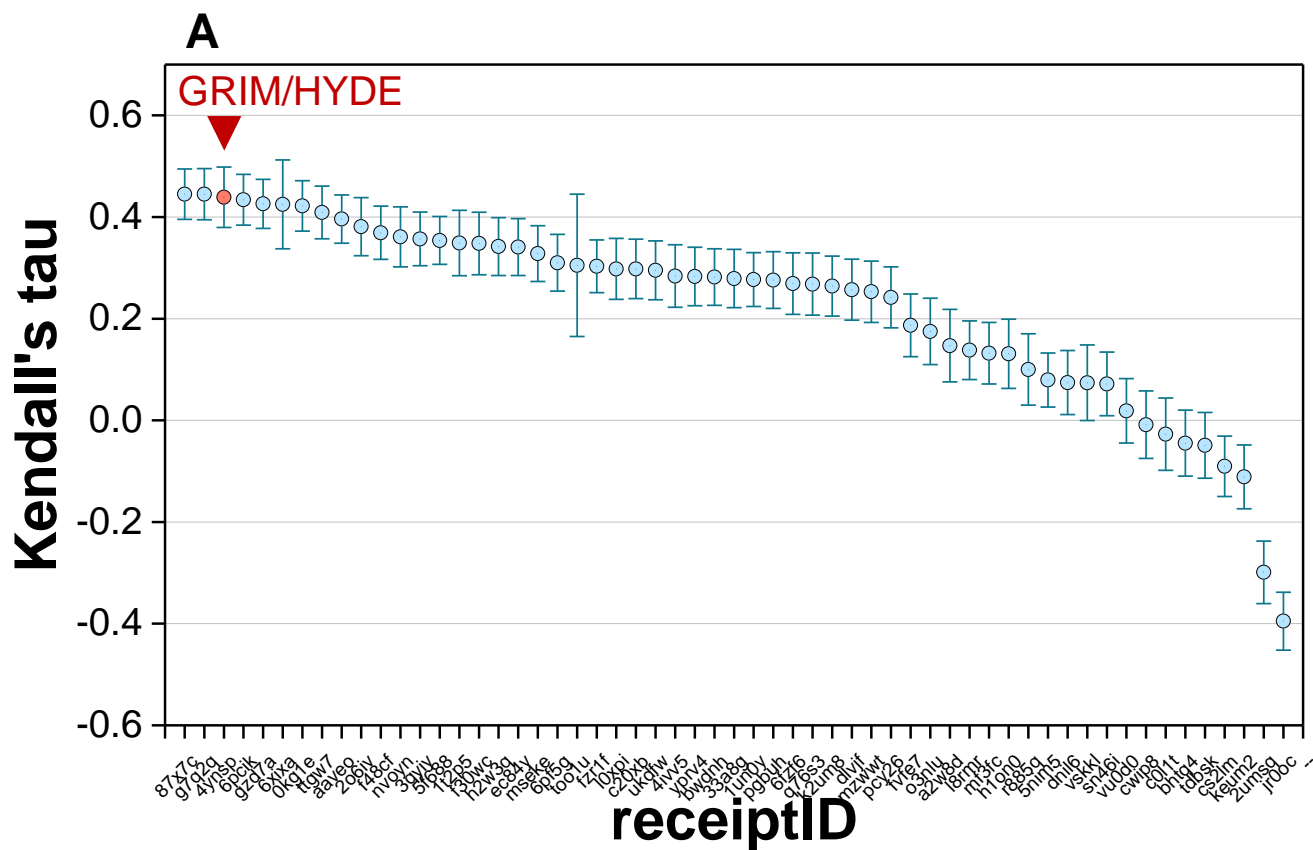














[Click here to access/download](#)

**Electronic supplementary material**  
Supplementary Table 1.xlsx





[Click here to access/download](#)

**Electronic supplementary material**  
Supplementary Table 2.xlsx

