



HAL
open science

Predicting binding poses and affinities for protein-ligand complexes in the 2015 D3R Grand Challenge using a physical model with a statistical parameter estimation

Sergei Grudin, Maria Kadukova, Andreas Eisenbarth, Simon Marillet,
Frédéric Cazals

► To cite this version:

Sergei Grudin, Maria Kadukova, Andreas Eisenbarth, Simon Marillet, Frédéric Cazals. Predicting binding poses and affinities for protein-ligand complexes in the 2015 D3R Grand Challenge using a physical model with a statistical parameter estimation. *Journal of Computer-Aided Molecular Design*, 2016, 30 (9), pp.791-804. 10.1007/s10822-016-9976-2 . hal-01377738

HAL Id: hal-01377738

<https://inria.hal.science/hal-01377738>

Submitted on 7 Oct 2016

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.

Predicting binding poses and affinities for protein - ligand complexes in the 2015 D3R Grand Challenge using a physical model with a statistical parameter estimation

Sergei Grudinin · Maria Kadukova · Andreas Eisenbarth ·
Simon Marillet · Frédéric Cazals

Received: date / Accepted: date

Abstract The 2015 D3R Grand Challenge provided an opportunity to test our new model for the binding free energy of small molecules, as well as to assess our protocol to predict binding poses for protein-ligand complexes. Our pose predictions were ranked 3-9 for the HSP90 dataset, depending on the assessment metric. For the MAP4K dataset the ranks are very dispersed and equal to 2-35, depending on the assessment metric, which does not provide any insight into the accuracy of the method. The main success of our pose prediction protocol was the re-scoring stage using the recently developed Convex-PL potential. We make a thorough analysis of our docking predictions made with AutoDock Vina and discuss the effect of the choice of rigid receptor templates, the number of flexible residues in the binding pocket, the binding pocket size, and the benefits of re-scoring.

However, the main challenge was to predict experimentally determined binding affinities for two blind test sets. Our affinity prediction model consisted of two terms, a pairwise-additive enthalpy, and a non pairwise-additive entropy. We trained the free parameters of the model with a regularized regression using affinity and structural data from the PDBBind database. Our model performed very well on the training set, however, failed on the two

test sets. We explain the drawback and pitfalls of our model, in particular in terms of relative coverage of the test set by the training set and missed dynamical properties from crystal structures, and discuss different routes to improve it.

keywords : protein-ligand docking; machine learning; scoring function; ridge regression; parameter estimation

1 Introduction

The Community Structure-Activity Resource (CSAR) was created in 2008 from the desire to provide high-quality experimental data to the community that is developing new computational tools. CSAR organized the first benchmark exercises in 2010 that highlighted the difficulties to predict the binding free energy of protein-ligand complexes, and discovered that some properties such as hydrogen bonding cause more trouble than the ligand size [1]. The goal of the 2011-2012 blind exercises was to compare improvements for pose prediction, enrichment, and relative ranking of congeneric series of compounds with a set of four protein targets. One of its conclusions was that correct ranking and correct pose prediction are not necessarily correlated [2]. This led to the CSAR Benchmark Exercise 2013 and 2014, which focused on the further improvement of the ability of computational methods to correctly predict binding poses and affinities. The successor of CSAR, the Drug Design Data Resource (D3R, www.drugdesigndata.org), is a new effort for docking and scoring data hosted at UCSD. It serves the same purpose as CSAR of providing high-quality experimental data for testing

S. Grudinin, M. Kadukova, A. Eisenbarth
Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France
CNRS, LJK, F-38000 Grenoble, France
Inria, F-38000 Grenoble, France
Tel.: +33 4 38 78 16 91
E-mail: Sergei.Grudinin@inria.fr

S. Marillet, F. Cazals
Inria, F-06902 Sophia Antipolis, France
Tel.: +33 4 92 38 71 88
E-mail: Frederic.Cazals@inria.fr

and improving ligand-protein docking algorithms and their scoring protocols.

The 2015 D3R Grand Challenge provided two unpublished datasets of protein-ligand co-crystal structures with the measured affinities to assess the current algorithms for pose and affinity predictions. The first dataset, referred to as HSP90, contained 8 crystal structures of the ATP site of the Hsp90 protein with resolution < 2.0 Å, as well as binding data for 180 compounds across five orders of magnitude and three chemical series. The challenge for this dataset was to predict the binding poses of 6 ligands spanning all three chemical series, and also to predict binding affinities for these 6 ligands as well as for the other 174 ligands. The second dataset, referred to as MAP4K4, comprised 18 compounds and crystal structures of a kinase protein, which is a member of the STE20 family, with resolution < 2.5 Å and inhibition/binding data over four orders of magnitude, in many cases confirmed by multiple assay methods. The challenge for this dataset was to predict the poses of 30 co-crystal structures containing diverse chemical series, and also to predict binding affinities of 18 of the ligand structures. Here, more co-crystal structures than affinity data were available.

We participated in the 2015 D3R Grand Challenge mainly to assess our new developments on the binding affinity predictions for small molecules, but also to validate our pose prediction protocol introduced earlier [3]. As we said, our main goal was to assess a new approximation to the binding free energy given by a combination of two terms, a pairwise-additive enthalpy, and a non pairwise-additive entropy, which in principle consists of conformational entropy, vibrational entropy, and solvent entropy. Below we provide a detailed explanation of our model with the results obtained on the training set and the two test sets provided by the 2015 D3R Grand Challenge.

2 Test Datasets and Organization of the Challenge

The D3R Grand Challenge 2015 consists of two datasets. The MAP4K4 (protein kinase kinase kinase) dataset was provided by Genentech [4]. It consists of 30 ligand compounds, for which co-crystal structures are available. For 18 of these complexes, there are also two experimental values provided in the results: IC_{50} *inactive* and K_i *active*.

The HSP90 (Heat Shock Protein 90) dataset was provided by AbbVie and Prof. Heather Carlson and CSAR colleagues at the University of Michigan and consists of 180 ligand compounds of three chemical classes (aminopyrimidines, benzimidazolones, benzophenone-like). Experimental IC_{50} are available for all of these. For 8 of them, co-crystal structures are provided.

The 2015 D3R Grand Challenge was organized in two stages. For the HSP90 dataset, the organizers provided as input 4 protein-ligand co-crystal structures drawn from the PDB and 2 protein-ligand co-crystal structures from the blinded dataset that were solved with compounds of each chemical class in three series. The organizers also provided SMILES strings of the 6 ligands to be docked, of the 2 ligands in the revealed co-crystal structures, and of the additional 172 compounds for affinity prediction or ranking. For the MAP4K4 dataset, the organizers provided as input 2 protein-ligand co-crystal structures drawn from the PDB with SMILES strings of the 30 ligands to be docked, among which 18 were selected for affinity prediction or ranking. In the first stage, the goal was to predict binding poses and binding affinities for the selected ligands. After the end of the first stage, the organizers made available the blinded co-crystal structures for both datasets. In the second stage, the goal was to predict binding affinities for the selected ligands using the released co-crystal structures of protein-ligand complexes. After the end of the second stage, the experimental binding affinities for both datasets were made available.

At the first stage of the Challenge, we submitted one docking pose prediction for each of the datasets. At this stage, we additionally submitted nine binding affinity predictions for the MAP4K4 dataset. All of these binding affinity predictions were computed using docking poses selected by nine different strategies.

At the second stage of the Challenge for each of the datasets, we submitted four binding affinity predictions using different prediction models. These were computed using co-crystal structures for the MAP4K4 dataset and top-one docking poses for the HSP90 dataset.

3 Binding Affinity Model

3.1 Free Energy Approximation

We approximated the logarithm of the binding affinity with the binding free energy, which is given as

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S, \quad (1)$$

where ΔH is the enthalpic difference between the bound and the unbound states of the protein-ligand complex, T is the temperature, and ΔS is the entropic difference upon binding. The binding free energy is a much more sophisticated function compared to the potential energy, and involves not only interaction energy between the partners, but also changes in the internal energy of the compounds, interactions with solvent, rearrangement of solvent molecules and changes of conformational degrees of freedom corresponding to the entropic loss upon binding. Direct computation of the binding free energy of a molecular complex, or of the relative binding affinities associated with two ligands (i.e., the binding energy difference) can be carried out with thermodynamic integration techniques. However, while this method can be applied to complexes consisting of few partners of small size, it reaches a prohibitively high computational cost, unless specific properties of the system(s) studied can be exploited [5,6]. For example an improved relative binding affinity calculation was recently proposed, using free energy perturbation / replica exchange with solute tempering (FEP/REST, [7]). In this method, only a small region of interest is *heated up*, so that a small number of replicas is sufficient to increase sampling and reduce the effects of quasi-nonergodicity. However, even if the selection of this region can be automated to some extent [6], this task is non trivial in all generality.

Therefore, in this work we adopted the following computational strategy. First, we approximate the enthalpic contribution ΔH using a pair-wise distance-dependent knowledge-base potential function that has been successfully applied to protein-protein and protein-ligand scoring [3,8,9], and whose coefficients are obtained from knowledge-base using a machine-learning technique. Second, we approximate the entropic contribution $T\Delta S$ with a set of geometric descriptors. These represent the non pairwise-additive part of the binding free energy [10] and have been recently demonstrated to perform very well on the protein-protein structure-affinity benchmark [11]. Below, we provide a more detailed description of the two contributions into the binding free energy.

3.2 Enthalpic Contribution

We represent the enthalpic contribution to the binding free energy as a linear functional F defined for all possible configurations of a protein-ligand complex [3,9]. To simplify the shape of the functional, we make several assumptions. First, we assume that F depends only on the interface between the protein and the ligand. We defined the interface as a set of atom pairs at a distance smaller than a certain cutoff distance r_{max} such that the first atom in each pair belongs to the protein and the second atom in each pair belongs to the ligand. We use the value of the cutoff distance of 10 Å, which was adapted from previous works [12,13,14,15,16,17,18,19]. It has been previously demonstrated that even smaller distances of 6 Å work somewhat well and that the quality of the scoring function does not improve after the cutoff distance of 12 Å. Second, we assume that both a protein and a ligand can be represented as a set of discrete interaction sites located at the centers of atomic nuclei. We divide all interaction sites into M types according to the properties of corresponding atomic nuclei, such as element symbol, aromaticity, hybridization, and polarity, resulting in a total of $M \times (M+1)/2$ pairs of interactions. An interaction between two sites can be also regarded as an interaction between two atoms of certain types. Third, we assume that F depends only on the distribution of the distances between the interaction sites, i.e. the number of site pairs at a certain distance,

$$F(P) \equiv F(n^{11}(r), \dots, n^{kl}(r), \dots, n^{MM}(r)) \equiv F(n(r)), \quad (2)$$

where $n^{kl}(r)$ is the *number density of site-site pairs* at a distance r between two sites k and l , with site k on the protein, and site l on the ligand, where M is the total number of interaction types. Finally, we assume that F is a linear functional, such that we can write the approximation to the enthalpic contribution as

$$\Delta H \approx \sum_{k=1}^M \sum_{l=k}^M \int_0^{r_{max}} n^{kl}(r) U^{kl}(r) dr, \quad (3)$$

where $U^{kl}(r)$ are unknown functions that are deduced from the training set of binding affinities for protein-ligand complexes. From now on, we will call these functions *scoring potentials*. We compute the site-site number densities for each protein-ligand complex $n^{kl}(r)$ as a sum of Gaussian-distributed distances between interaction sites of types k and

l as

$$n^{kl}(r) = \sum_{uv} e^{-\frac{(r-r_{ij})^2}{2\sigma^2}}, \quad (4)$$

where σ is the standard deviation, which is constant for all distributions. The sum is taken over all pairs of sites uv of types k and l separated by a distance r_{ij} smaller than a certain threshold distance r_{max} , with the first site located on the protein, and the second site located on the ligand.

In order to determine unknown scoring potentials $U^{kl}(r)$ (Eq. 3), we decompose them along with the number densities $n^{kl}(r)$ in a polynomial basis,

$$\begin{aligned} U^{kl}(r) &= \sum_q w_q^{kl} \psi_q(r), \quad r \in [0; r_{max}] \\ n^{kl}(r) &= \sum_q x_q^{kl} \psi_q(r), \quad r \in [0; r_{max}], \end{aligned} \quad (5)$$

where $\psi_q(r)$ are orthogonal basis functions on the interval $[0; r_{max}]$, and w_q^{kl} with x_q^{kl} are the expansion coefficients of $U^{kl}(r)$ and $n^{kl}(r)$, respectively. Given this, the enthalpic contribution ΔH can be expanded up to the order Q as

$$\begin{aligned} \Delta H \approx \sum_{k=1}^M \sum_{l=k}^M \sum_q^Q w_q^{kl} x_q^{kl} = (\mathbf{w} \cdot \mathbf{x}), \\ \mathbf{w}, \mathbf{x} \in \mathbb{R}^{Q \times M \times (M+1)/2} \end{aligned} \quad (6)$$

We will refer to the vector \mathbf{w} as to the *scoring vector*, whose value is to be determined from the knowledge-base, and to the vector \mathbf{x} as to the *structure vector* that is computed from the structural data. For the estimation of binding affinity, we used the maximum expansion order $Q = 10$ and the number of types $M = 48$, as it is explained below. Thus, the dimensionality of the *scoring vector* \mathbf{w} and the *structure vector* \mathbf{x} is $Q \times M \times (M + 1)/2 = 11,760$.

3.2.1 Atom Types of the Enthalpic Contribution

Our knowledge-based enthalpic contribution to the binding free energy requires definition of the atom types. We assigned the types to the atoms of both proteins and small molecules according to their physical and chemical properties as well as their functional groups. To do so, we started with 164 internal atom types inspired by the fconv library[20] and grouped them by measuring the statistical similarity of distance distribution functions between different atom types in the training data set. Overall, our parameterization consists of 48 atom types.

More precisely, we have 17 types for nitrogen, 9 types for oxygen, 8 types for carbon, 4 types for sulphur, 2 types for phosphorus, and 8 types for halogens. Table S1 from Electronic Supplementary Material lists all the used atom types along with the chemical properties of the corresponding atoms.

We should specifically add that our model has no explicit hydrogen atoms. Generally, it is not a trivial task to correctly predict positions of all missing hydrogens. For example, their positions depend on experimental parameters of the system, such as pH. Neither our model has directional bonds. Introducing explicit hydrogen atoms or angular dependency would add additional degrees of freedom (or features) to our model, thus increasing the size of the feature space.

3.3 Entropic Contribution

Of particular interest when modeling macro-molecular complexes are terms meant to code entropic properties, in particular conformational entropy, vibrational entropy, and solvent entropy. To model the latter two terms, we use geometric features recently introduced in [10], which proved instrumental to obtain state-of-the-art affinity predictions for the protein - protein complexes of the structure-affinity benchmark [11]. In a nutshell, these descriptors dissect the surface and volume properties of the isolated subunits and of the complex whose affinity is to be estimated. Their rationale is to replace the *buried surface area* (BSA), namely the portion of the exposed surface of both partners which gets buried upon complex formation [21], with a more informative descriptor. While the BSA exhibits a remarkable correlation with various biophysical quantities [22], it does not account for interface shape and packing properties, these latter attributes being of interest to model entropic changes.

Interface definition. We define interface atoms using the parameter free *Voronoi interface model*. It is based on the α -complex derived from the Voronoi (power) diagram of the atoms, and has been proposed in [23,24]. In a nutshell, define the *restriction* of an atom as the intersection between its ball in the solvent accessible model and its cell in the Voronoi diagram. The Voronoi interface identifies pairs of neighboring restrictions, such that each pair involves either two different partners or a partner and the interfacial solvent. The atoms

found in at least one such pair are denoted \mathcal{I} and their complement \mathcal{I}^C .

Inverse volume-weighted internal path length (IVW-IPL) and variants. This parameter encodes the size and morphology of the interface and takes atomic packing into account. Its definition involves three ingredients.

The first one consists in the interface atoms \mathcal{I} contributing to the Buried Surface Area, identified with the aforementioned Voronoi based model [24]. Once identified, the interface atoms of a partner define its *binding patch* (Fig. 1(A)). The second one is the notion of *shelling order* (SO), namely an integer value assigned to the interface atoms of a partner, and measuring the distance to the bounding patch boundary (Fig. 1(B)). The shelling order of atom a is denoted $SO(a)$. The third one is the packing of an atom. Atomic packing is defined for an atom in a structure which may be the isolated subunit containing this atom or the complex defined by these subunits. Following classical works [25], we define the atomic packing as the volume of the solid defined by the intersection between the ball of the atom (in the solvent accessible model) and the Voronoi region of this atom in the structure (the isolated subunit or the complex) ([26] and Fig. 1(CD)). This volume is denoted `volume_bound(a)` for atom a .

Using these notions, the IVW-IPL is defined as follows:

$$IVW-IPL = \sum_{a \in \mathcal{I}} \frac{SO(a)}{\text{volume_bound}(a)} \quad (7)$$

On the one hand, the shelling order refines so-called core-rim models [21]. Borrowing to the notion of cooperative effects involving non-bonded weak interactions, an isotropic or disk-like interface is indeed expected to be more stable than an elongated one—even if their surface areas match. On the other hand, the atomic packing encodes the local density of neighbors of a given atom, and thus provides a measure for local interactions (hydrogen bonds, van der Waals interactions). Note that packing is a subtle quantity related to the enthalpy - entropy compensation [27,28], as its properties strike a balance between enthalpy (a large number of neighbors favors interactions) and entropy (by restricting the moving freedom of atoms, tight packing is detrimental for dynamics yielding an entropic penalty assuming they were free to move in the first place).

We also use two variants of the previous quantity. First, the *inverse volume sum* which is the same as IVW-IPL except that $SO(a)$ is replaced by 1 in Eq. (7). Second, the number of atoms at interface which can be written as in Eq. (7), but replacing $SO(a)/\text{Vol}(a)$ with 1. These two quantities essentially isolate the contributions of the numerator and denominator in Eq. (7) so that the model parameters can be adjusted separately.

Remark 1 The previous quantities were computed using the following programs from the Structural Bioinformatic Library (<http://sbl.inria.fr> > Applications): `sbl-vorshell-bp-ABW-atomic.exe` [24] for binding patches and atomic shelling orders ¹, and `sbl-vorlume-pdb.exe` [26] for atomic volumes and surface areas ².

Solvent Interactions and Electrostatics. In order to account for the interaction between the partners and water molecules, we use the fractions of charged and polar residues on the non interacting surface (NIS) [29], NIS^{polar} and $NIS^{charged}$. The NIS consists of the exposed surface of the partners not involved in the interface (Fig. 1(A), \mathcal{I}^C and $SASA > 0$).

We also include the atomic solvation energy, σ , from Eisenberg et al [30], describing the free energies of transfer from 1-octanol to water per surface unit (\AA^2). This value is positive for non-polar atoms, and negative for polar atoms. The corresponding variable, `ATOM_SOLV`, is a weighted sum of atomic solvent accessible surface areas (Eq. (8)), and may be seen as the atomic-scale counterparts of $NIS^{charged}$ and NIS^{polar} :

$$ATOM_SOLV = \sum_{a \in \mathcal{I}^C} SASA(a) \cdot \sigma(a) \quad (8)$$

Finally, we include an intermediate-grained description of the non-interacting surface which consists in the total area of polar atoms of the complex (according to Eisenberg’s solvation parameters). The corresponding term, `POLAR_SASA` (Eq. (9)), is also a weighed sum of exposed areas.

$$POLAR_SASA = \sum_{a \in \mathcal{I}^C \text{ and } \sigma(a) < 0} SASA(a) \quad (9)$$

Using the introduced interactions terms, we represent the entropic contribution to the binding free

¹ User manual: http://sbl.inria.fr/doc/Space_filling_model_shelling_diagram_surface_encoding-user-manual.html

² User manual: http://sbl.inria.fr/doc/Space_filling_model_surface_volume-user-manual.html

energy as a linear combination of the following form,

$$T\Delta S \approx w_1\text{IVW-IPL} + w_2\text{NIS}^{\text{polar}} + w_3\text{NIS}^{\text{charged}} + w_4\text{ATOM.SOLV} + w_5\text{POLAR.SASA} \quad (10)$$

Similar to the vector of free parameters \mathbf{w} of the enthalpic term, we will deduce free parameters $w_1\dots w_5$ using a machine-learning procedure, as explained below.

Remark: Note that all the atomic descriptors except $\text{NIS}^{\text{polar}}$ and $\text{NIS}^{\text{charged}}$ are computed for both partners. Since the ligand is not made of amino-acids, these two residue-level descriptors are computed for the protein receptors only.

3.4 Training Set

To train our model of binding free energy given by Eqs. 6 and 10, we collected structural and binding affinity information from the PDBBind [32, 33] database, which provides experimentally determined protein-ligand complexes deposited in the Protein Data Bank supplied with the measured binding affinity data. To construct the PDBBind database, its authors manually examined the primary references for each protein-complex and collected experimentally determined binding affinity data (K_d , K_i and IC_{50}). These constitute the “general set” of the database consisting of 14,620 complexes. Then the authors of PDBBind additionally compiled the “refined set” to select protein-ligand complexes with better quality out of the general set. To do so, they applied a number of filters to the selection regarding binding data, crystal structures, as well as the nature of the complexes. For the training, we used 3,706 complexes from the latter “refined set” of PDBBind release 2015, which contains three-dimensional structures of resolution equal to or better than 2.5 Å with the corresponding binding data, which includes K_d , K_i , and IC_{50} values.

3.5 Training Models

Entropy only model. We trained a simple model consisting only of the five variables encoding the entropic contribution. Because of its small size, no

regularization was needed and we resorted to ordinary least-squares regression to compute the parameters. Namely, the following function is minimized:

$$C(\mathbf{w}) = \sum_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2, \quad (11)$$

where \mathbf{w}_i , $1 \leq i \leq 5$ are the parameters corresponding to variables \mathbf{x}_i , $1 \leq i \leq 5$, and y_i are experimentally measured $\log K_d/K_i$ values.

Enthalpy only and combined model. Our model for the binding affinity contains 11,765 free parameters (11,760 and 5, respectively introduced in sections 3.2 and 3.3) that were to be deduced from 3,706 experimentally measured data points. Because the dimensionality of our problem is several times larger compared to the size of the training set, to deduce the free parameters we chose a regression model with regularization. After several initial experiments with different regression models, we stopped at the linear ridge (LRR) and kernel ridge regressions (KRR), mainly because of their speed for our data [34]. The linear and kernel ridge regressions minimize the following cost function, which consists of the regression term and the regularization term,

$$C(\mathbf{w}) = \sum_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2 + \lambda \|\mathbf{w}\|_2^2, \quad (12)$$

where y_i are experimentally measured $\log K_d/K_i$ values, \mathbf{x}_i are features computed from 3D structures of protein-ligand complexes as described above, and λ is the adjustable regularization parameter that determines the importance of the regularization term with respect to the regression term. Its optimal value is determined using the cross validation procedure. For the kernel regression model, we used a Gaussian kernel to compute the inner product between feature vectors. This allowed us to introduce a non-linear relation between our features and the binding affinity data.

3.6 Computational Details

To train our prediction model with LRR and KRR, we used the SHOGUN C++ library [35]. We found two adjustable parameters, the regularization parameter λ , and the width of the Gaussian kernel, using a 5-fold cross validation. The confidence interval for the cross-validation was set to 95% and

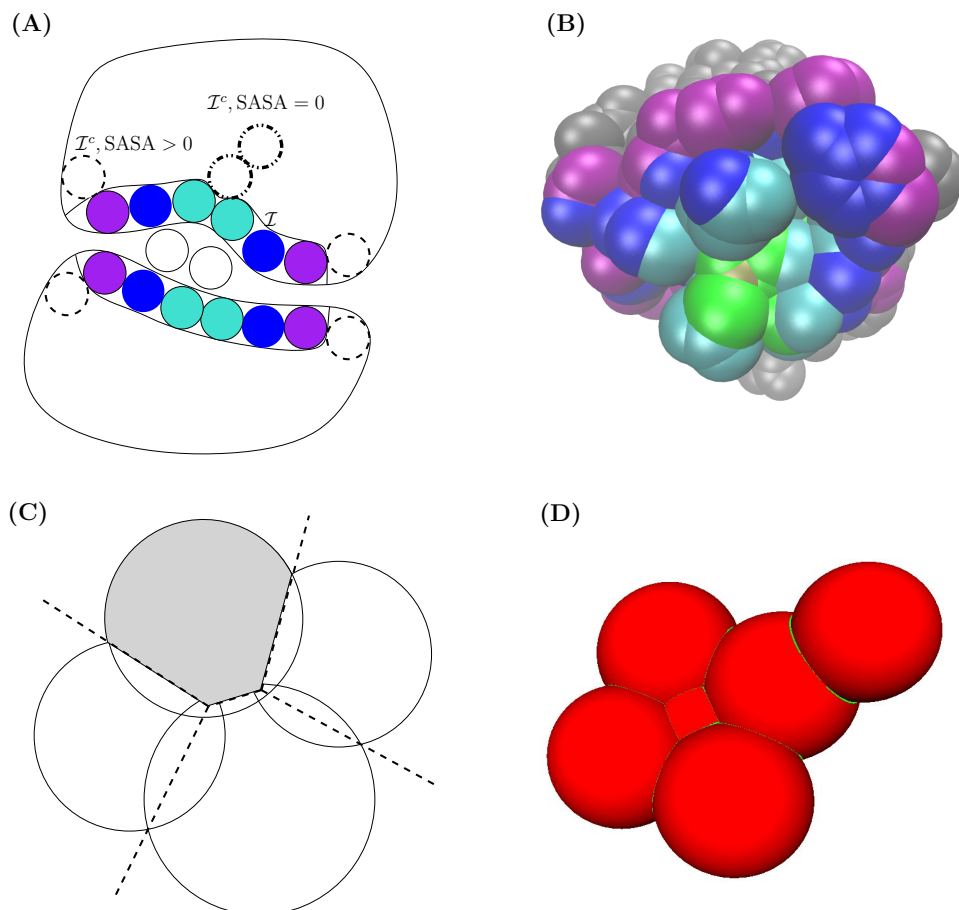


Fig. 1 Structural parameters used in this work. (A) A fictitious 2D complex and its interface atoms label highlighted in color. The binding patch on each partner consists of one layer of atoms (\mathcal{I} , colored solid balls), as identified by a Voronoi interface model [23,24]. The non interface atoms (\mathcal{I}^c) are ascribed to two categories: those which retain solvent accessibility ($SASA > 0$, dashed balls), and those which do not ($SASA = 0$, dash-dotted balls) (B) Each interface atom is assigned its *shelling order*, namely an integer equal to the smallest number of atoms traveled to reach an exposed non interface atom, i.e. an atom belonging to \mathcal{I}^c and with $SASA > 0$ (in grey) [31]. (C,D) Atomic packing measured by atomic volumes. The volume of an atom is defined as the volume of the intersection between its ball in the solvent accessible model, and its Voronoi cell [26]. Note that this quantity is defined even if the atom retains solvent accessibility. Practically, interfaces and binding patches are computed with `Vorshell`[24], while atomic surface areas and volumes are computed with `sbl-vorlume-pdb.exe` [26]. Both programs are available from the Structural Bioinformatics Library (SBL), see <http://sbl.inria.fr>.

the number of runs was set to 3. Atomic typization for the enthalpic features was computed using the Knodle library [36] available at <https://team.inria.fr/nano-d/software/Knodle>. Geometric descriptors for the entropic features were computed using the Structural Bioinformatics Library available at <http://sbl.inria.fr>.

4 Pose Prediction Protocol

To generate putative binding poses of a ligand with a protein, we used the AutoDock Vina software package with the default scoring function[37]. Using AutoDock Vina, we generated 1,000 docking solutions for the subsequent re-scoring. To start

the AutoDock Vina docking, we detected the binding pockets of each protein-ligand complex by analogy with the known homologous complexes. Using PyMOL and the AutoDock Vina plugin available at <http://wwwuser.gwdg.de/~dseelig/adplugin.html> [38], bounding boxes were fitted to the binding pockets. For the MAP4K4 dataset, we progressively tested three bounding boxes of increasing volumes, ranging from 6,125 to 8,685 \AA^3 for the 4OBO receptor, and ranging from 5,610 to 12,472 \AA^3 for the 4U44 receptor. In the obtained AutoDock Vina configuration files, the parameter `num_modes` was set to 1,000 and `exhaustiveness` to 100. We chose all the ligands to be flexible during the docking procedure, and we kept

all the protein residues inside the binding pockets rigid. To test the effect of flexible receptor residues, we ran two additional docking simulations for the MAP4K4 dataset. First, we allowed residues Tyr 36, Lys 54, Met 105, Phe 107, and Asp 171 to be flexible. Second, we allowed residues Tyr 36, Met 105, and Phe 107 to be flexible. However, we were not satisfied by the initial sampling results and did not re-score the predictions. We used the default assignment of ligand bond orders predicted by AutoDock Tools. However, one of the conjugated double bonds in the HSP90 set was falsely marked as rotatable. Also, we did not have flexibility in non-aromatic cycles, i.e. in cyclohexyl. To create the initial three-dimensional models of ligands given in the SDF format, we used the Open Babel conversion library[39]. We used the provided templates as the models of protein receptors, 4OBO and 4U44 for all the ligands in the MAP4K4 dataset and 2XDX (HSP90.73 and HSP90.179 ligands), 3OW6 (HSP90.40 and HSP90.44 ligands), 4YKR (HSP90.40 and HSP90.44 ligands), and 4YKY (HSP90.164 and HSP90.175 ligands) for the Hsp90 dataset. Then, we re-scored the docking solutions with the Convex-PL scoring function [3], without any further structure optimization, and ranked the poses according to the computed values. If multiple rigid receptor templates or multiple bounding boxes around the binding pockets were used, we re-scored the docking solutions together from all the predictions. For the first stage of the challenge, we submitted five top predictions for each protein-ligand complex for the two datasets. At the second stage of the challenge, we used top-one predictions for the HSP90 dataset for the subsequent re-scoring with multiple scoring functions to predict the binding affinities. At this stage, we did not predict binding poses for the MAP4K4 dataset, since the co-crystal structures were made available by the organizers.

For the assessment of docking poses we used the LigAlign analysis program [40] written for the PyMOL package [41]. We visually inspected the structural alignment of all the ligands in MAP4K and HSP90 datasets and corrected the alignment weight function of LigAlign when needed for the correct alignment.

5 Results and Discussion

5.1 Predicted Docking Poses

Table 1 lists results for our docking predictions for the two sets, MAP4K and HSP90. This table shows

the averaged assessment results for the top-1 binding poses, best binding poses, and all the five poses for each protein-ligand complex. As it can be seen from the table, we obtained good predictions for the HSP90 set with the mean RMSD for the best pose of 0.65 Å. These poses along with the crystallographic solutions are shown in Fig. 2. For the MAP4K set, however, AutoDock Vina experienced sampling problems for some of the large ligands. More precisely, in some regions of space no samples were generated while other were sampled densely. As a result, long ligands did not find their place into the corresponding binding pockets. For example, when using 4U44 template to model the receptor, the minimum RMSD in top 1,000 docking poses for MAP12 ligand was 11.2 Å, for MAP13 ligand was 6.5 Å, and for MAP17 ligand was 7.9 Å. Similarly, when using the 4OBO template to model the receptor, the minimum RMSD in top 1,000 docking poses for MAP12 ligand was 5.5 Å, for MAP13 ligand was 5.0 Å, and for MAP15 ligand was 4.5 Å. Conversely, for small ligands such as MAP03, MAP04, MAP21, MAP22, MAP30, and MAP32, we could obtain the minimum RMSD in top AutoDock Vina 1,000 docking solutions within 1 Å at least for one out of two rigid receptor templates. Also, since we did not relax the initial protein structures, some of our binding poses had steric clashes with bulky protein residues, such as phenylalanines. Overall, for the MAP4K set, we achieved the mean RMSD for the best pose of 3.52 Å. Fig. 3 shows some of the successful predictions for this set.

In the MAP4K set predictions there is a big difference of 3.58 Å between RMSD values of mean pose 1 and mean best pose. This is mostly due to the choice of different templates, namely, both 4OBO and 4U44 for all of the predicted ligands. If we discard 4U44-based predictions from the submission, the difference between RMSD values of mean pose 1 and mean best pose reduces to 1.08 Å. Similarly, if we discard 4OBO-based predictions, this difference reduces to 1.71 Å. The difference between RMSD values of mean pose 1 and mean best pose for the HSP90 set is explained by a flipped ring in the HSP90_164 ligand ranked first by the Convex-PL potential.

An interesting question is to which extent re-scoring with an additional potential is useful for successful pose predictions. To answer this question, we analyzed 1,000 top poses predicted by AutoDock Vina. Table 2 shows RMSD values of ten docking simulations for the MAP4K dataset

dataset	# submissions	Obtained results (overall rank)								
		RMSD, Å			RSR			RSCC		
		mean pose 1	mean best pose	mean all poses	mean pose 1	mean best pose	mean all poses	mean pose 1	mean best pose	mean all poses
MAP4K	35	7.1 (28)	3.52 (12)	6.54 (25)	3.39 (7)	2.72 (2)	3.43 (8)	0.37 (20-29)	0.15 (27-35)	0.35 (21-31)
HSP90	52	1.07 (8)	0.65 (5)	1.31 (5)	3.57 (5)	3.12 (3)	3.8 (3)	0.71 (8-9)	0.57 (7-9)	0.7 (5-6)

Table 1 Assessment of the docking pose predictions for the MAP4K and HSP90 datasets. Here, RMSD states for root mean square deviation, RSR states for real space R-factor, and RSCC states for real space correlation coefficient. RMSD is a similarity measure of an atomistic model with the one obtained experimentally. RSR and RSCC are measures of the similarity between an electron-density map calculated directly from a structural model and one calculated from experimental data.

computed with respect to co-crystal structures. We can clearly see that re-scoring with the Convex-PL potential significantly improves the quality of the best pose among top 5 poses. This is true for all AutoDock Vina simulations of this set. However, re-scoring did not improve the mean best pose, neither it improved the mean among top 5 poses. Table 3 shows the analysis of 100 poses predicted by AutoDock Vina for the HSP90 dataset. Here we see that re-scoring with the Convex-PL potential significantly improves all the RMSD values, namely, the mean top pose, the mean best pose, and the mean among all top 5 poses. To conclude, we strongly believe that the default scoring function of AutoDock Vina is too simplistic, since it is mainly designed to describe steric interactions, thus an additional re-scoring step with a more sophisticated scoring function is required. We also think that a subsequent gradient-based optimization step can further improve the quality of predictions.

Another interesting question is how much the quality of pose predictions depends on the choice of the rigid receptor template, the number of flexible residues in the binding pocket, and the binding pocket size. We cannot answer this question generally, since the conclusions should depend on the sampling method. However, we can draw certain conclusions from our docking experiments with AutoDock Vina, which is one of the most popular docking methods. First, if we analyze the effect of the size of the search space in the binding pocket, we can clearly see that the search smaller the volume, the higher the probability to find a better docking pose (see column 4 in Table 2). This emphasizes the insufficient sampling of the binding pocket by AutoDock Vina even with the sampling parameters chosen much higher than the default values (*num.modes* set to 1,000 and *exhaustiveness* set to 100). Second, flexibility of some of the receptor residues does not, generally, improve the quality of the predictions. This is probably connected to the previous problem, i.e. the insufficiently sampled search

space, which is much larger in this case. Finally, the success of the pose prediction critically depends on the chosen receptor template. This dependency cannot be seen clearly in results presented in Table 2 for the MAP4K dataset. However, we analyzed AutoDock Vina predictions for the HSP90 set, where several co-crystal structures are available for the same family of ligands. Table 3 shows RMSD values to co-crystal structures for the HSP90_73 and HSP90_179 ligand predictions if started with 2XDX and 2JJC templates, and also for HSP90_40 and HSP90_44 ligand predictions if started with 4YKR, 3OWD, and 3OW6 templates. We can see very large deviations in the best pose RMSD value depending on the chosen receptor template. Thus, it is very critical to pre-select the best receptor model for each of the ligand in a set prior to starting the docking simulations.

5.2 Affinity Model Training

As a first affinity prediction computational experiment, we trained our affinity models using the linear and the kernel ridge regressions. Table 4 lists the training results after a five-fold cross validation. We should mention that all the features were scaled to an interval [0; 1] prior to the training. As further computational experiments demonstrated, in order to have an affective regularization, it is sufficient to properly scale the five entropic features, since they have very different amplitudes compared to all the enthalpic features. We also noted that the regularization yielded a dense model, more specifically, the linear model had 5134 / 11765 i.e. 43% of non null terms. This is mainly because some of our 11760 enthalpic features are never met in protein-ligand complexes from the training set, and also because the l_2 norm used in our regression models produces a dense solution. This said, it appears that enthalpy is more important for the prediction of the binding affinities, compared to entropy. We can also see that when a linear model is used it is important to have both

dataset	template	# receptor flexible residues	search volume, Å ³	AutoDock Vina scoring function results							
				RMSD, Å, for 5 top poses			RMSD, Å, for 100 top poses		RMSD, Å, for 10 ³ top poses		
				mean pose 1	mean best pose	mean all poses	mean best pose	mean all poses	mean best poses	mean all poses	
MAP4K	4OBO	0 ^a	8,685	5.91	5.41	5.78	3.73	6.30	2.56	6.65	
MAP4K	4OBO	0 ^a	7,765	5.86	5.31	5.79	3.79	6.20	2.61	6.69	
MAP4K	4OBO	0 ^a	6,125	5.84	5.48	5.84	2.94	6.35	1.96	7.39	
MAP4K	4OBO	5 ^b	10,965	7.97	7.50	8.01	7.26	8.28	-	-	
MAP4K	4OBO	3 ^c	7,765	8.20	7.60	8.15	8.17	7.00	-	-	
MAP4K	4U44	0 ^a	12,472	5.50	5.13	5.43	3.60	5.96	2.84	6.21	
MAP4K	4U44	0 ^a	10,200	5.74	5.03	5.53	3.44	5.93	2.56	6.41	
MAP4K	4U44	0 ^a	5,610	5.42	5.03	5.42	3.15	5.86	2.40	6.39	
MAP4K	4U44	5 ^b	12,472	9.97	9.17	9.72	8.45	9.82	-	-	
MAP4K	4U44	3 ^c	8,479	5.71	4.72	5.59	3.45	5.84	-	-	

^a The receptor protein was kept rigid.

^b Five flexible residues on the receptor protein were chosen as Tyr 36, Lys 54, Met 105, Phe 107, and Asp 171.

^c Three flexible residues on the receptor protein were chosen as Tyr 36, Met 105, and Phe 107.

Table 2 Assessment of the docking pose predictions for the MAP4K set when only using the AutoDock Vina scoring function. For each of the two datasets five docking simulations were performed, three with the rigid receptor and bounding boxes around the binding site of a different size, and two with some residues of the receptor inside the binding pocket allowed to be flexible. Only the rigid simulations were used for the subsequent re-scoring. RMSD to the ligands of experimentally obtained structures are computed for the top 5 poses, top 100 poses, and top 1,000 poses.

dataset	template	AutoDock Vina scoring function results							
		RMSD, Å, for 5 top poses			RMSD, Å, for 10 top poses		RMSD, Å, for 100 top poses		
		mean pose 1	mean best pose	mean all poses	mean best pose	mean all poses	mean best poses	mean all poses	
HSP90	multiple	2.27	1.22	1.94	0.81	2.18	0.65	3.42	
HSP90_73	2XDX	1.09	1.00	1.09	1.00	2.51	1.00	4.03	
HSP90_73	2JJC	5.46	5.46	5.48	4.35	5.32	3.75	5.16	
HSP90_179	2XDX	2.86	2.71	2.80	0.50	2.45	0.41	3.39	
HSP90_179	2JJC	3.45	3.38	3.58	1.98	3.31	1.55	3.92	
HSP90_40	4YKR	0.59	0.46	0.58	0.43	0.95	0.43	4.23	
HSP90_40	3OWD	5.41	5.41	5.49	4.89	5.62	3.22	5.15	
HSP90_40	3OW6	0.70	0.70	2.53	0.70	3.98	0.70	5.39	
HSP90_44	4YKR	2.68	2.07	2.37	2.07	2.64	1.33	5.23	
HSP90_44	3OWD	7.88	5.75	6.92	5.75	6.92	5.60	7.80	
HSP90_44	3OW6	2.52	2.17	2.40	1.94	2.36	1.20	3.47	

Table 3 Assessment of the docking pose predictions for the HSP90 set when only using the AutoDock Vina scoring function. RMSD to the ligands of experimentally obtained structures are computed for the top 5 poses, top 10 poses, and top 100 poses. See main text for details.

terms in the approximation of the binding free energy. However, if we use a non-linear model, then the additional entropic term does not contribute to the overall accuracy of the prediction. This can be explained by the fact that the entropic term represents non-additive contributions to the binding free energy and it is very important to have it in the linear model. On the other hand, when we use the kernel ridge regression, the non-additivity of the entropic term is efficiently replaced by the non-linear kernel. Figure 4 shows predicted binding affinities versus experimentally measured values for both linear and non-linear models. We can see both from Table 4 and Fig. 4 that our affinity model predicts the training set very well, with the correlation coefficient for LRR of 0.82 and correlation coefficient for KRR of 0.92.

affinity model	# free parameters	RMSE, in log K_i/K_d units	
		LRR	KRR
$T\Delta S$	5	2.12	1.66
ΔH	11760	1.57	1.32
$\Delta H - T\Delta S$	11765	1.39	1.32

Table 4 Affinity model training accuracy measured as a root mean squared error.

5.3 Affinity Model Testing

The next computational experiment was to test the trained binding affinity model on the blind data of the MAP4K and HSP90 datasets. Particularly interesting for us was the MAP4K dataset, since the D3R challenge organizers provided 18 co-crystal structures for the prediction of binding affinities. Figure 5 shows the performance of our binding model on the MAP4K dataset. Notably, the accuracy of the model is much lower compared to the training set.

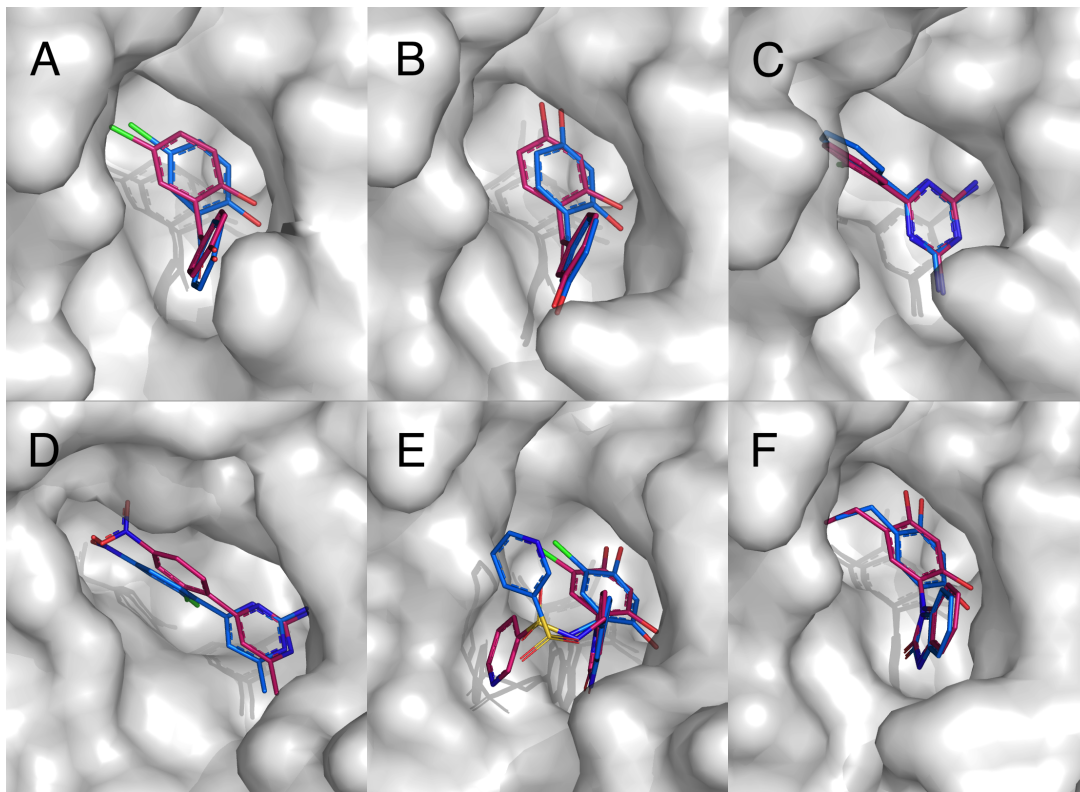


Fig. 2 Best binding poses for the HSP90 set. Crystallographic structures are shown in blue. Predictions are shown in red. Molecular surface is shown for the crystallographic structure. A) HSP90_40 ligand with the 4YKQ co-crystal structure. B) HSP90_44 ligand with the 4YKT co-crystal structure. C) HSP90_73 ligand with the 4YKW co-crystal structure. D) HSP90_164 ligand with the 4YKX co-crystal structure. E) HSP90_175 ligand with the 4YKZ co-crystal structure. F) HSP90_179 ligand with the 4YKU co-crystal structure.

To get insights on these disappointing performances, we checked to what extent cases of the test set are represented by cases from the training set, both for the enthalpic and entropic components – that is we work successively in the $d = 11760$ dimensional space coding enthalpic properties, and in the $d = 5$ dimensional space coding entropic properties. Following general results on the convergence of nearest neighbor based regressors [42], we compare two distributions of distances in the d -dimensional space just defined. We set the number nearest neighbors to be 8 ($\approx \log 3, 706$). The first one (D_{Train}) is the distribution of the average distance, for each point of the training set, to its 8 nearest neighbors in the training set itself (computed with the Euclidean distance in the d -dimensional space). The second (D_{Test}) is the distribution of the average distance, for each point of the test set, to its 8 nearest neighbors in the training set.

Consider first the 11760 space coding enthalpic properties. Computing the tuple (min, median, mean, max) yields (0.067, 0.481, 0.496, 1.535) for D_{Train} ,

and (0.505, 0.744, 0.739, 1.149) for D_{Test} (see also Fig. 6(A)). The mean and median distances almost double in moving from D_{Train} to D_{Test} . Consider now the five dimensional space. The same tuple (min, median, mean, max) yields (57.27, 397.20, 1113.00, 120200.00) for D_{Train} , and (271.9, 422.5, 418.1, 650.9) for D_{Test} (see also Fig. 6(B)). While the latter distribution occupies a more compact domain, the median value is significantly larger than that from the former distribution. Summarizing, both for the enthalpy and entropy related variables, the looser coverage of the test set by the training set intuitively explains why the prediction performance degrades when moving from the training set to the test set.

We should add that at the first stage of the Challenge for this dataset, we also submitted nine different binding affinity predictions. These were computed by re-scoring with the Convex-PL potential the top binding poses predicted using nine different docking strategies. More precisely, we predicted top-one binding poses with three bounding boxes of different size around the binding site.

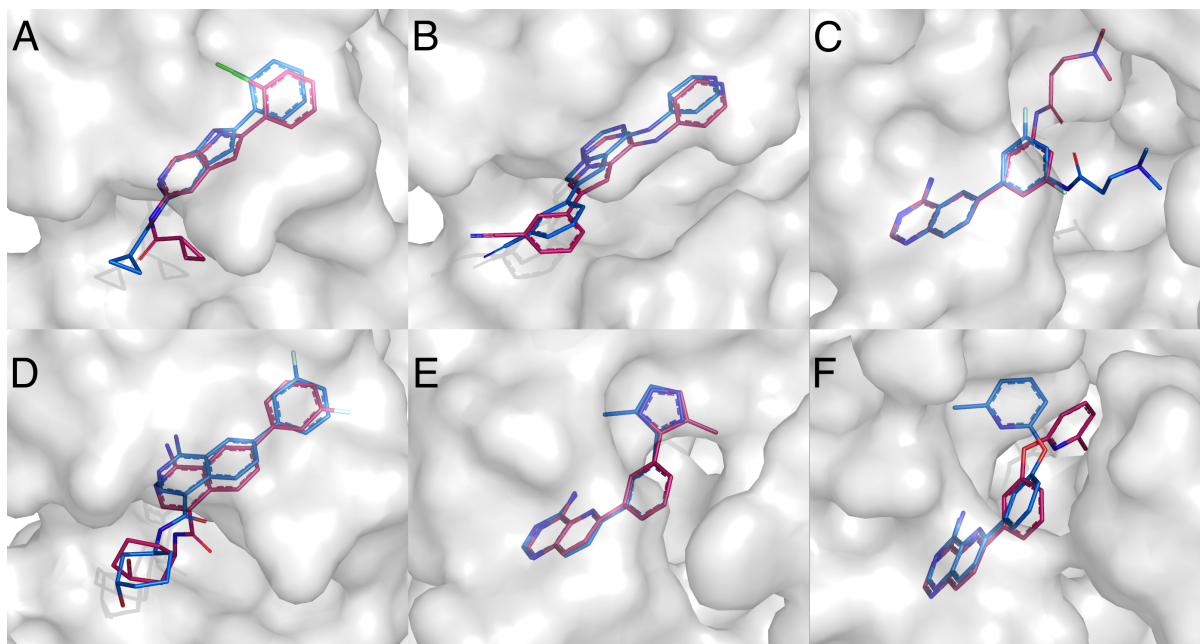


Fig. 3 Examples of some of the relatively successful predictions for the MAP4K set. Crystallographic structures are shown in blue. Predictions are shown in red. Molecular surface is shown for the crystallographic structure. A) MAP07 ligand prediction. B) MAP11 ligand prediction. C) MAP14 ligand prediction. D) MAP18 ligand prediction. E) MAP19 ligand prediction. F) MAP23 ligand prediction.

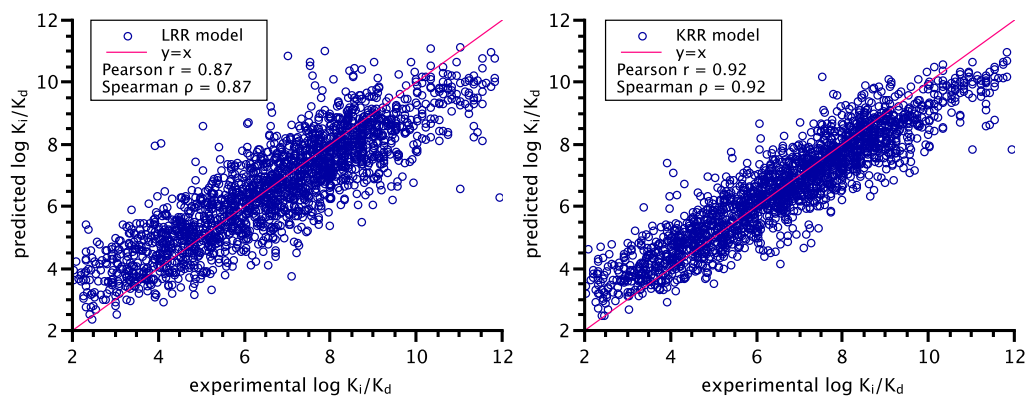


Fig. 4 Accuracy of the $\Delta H - T\Delta S$ model on the training set after a five-fold cross validation. Left : training of the linear model. Right : training of the non-linear model.

These were permuted with receptors chosen solely from the 4OBO template, solely from the 4U44 template, and a mix of two. The obtained Spearman ρ coefficients varied from 0.25 to 0.4. Respectively, the Kendal τ coefficients varied from 0.15 to 0.29.

At the second stage of the Challenge, we submitted four different binding affinity predictions. These were computed by re-scoring the co-crystal structures with the following scoring functions : a) the Convex-PL potential ($\rho = 0.25$, $\tau = 0.24$); b) the entropic term given by Eq. 10 trained with the LRR model without feature rescaling ($\rho = 0.44$,

$\tau = 0.33$, Eq. 11), which provided the best results; c) the free energy approximation given by Eqs. 1,6,10 trained with the LRR model without feature rescaling ($\rho = 0.3$, $\tau = 0.19$, Eq. 12); d) the free energy approximation given by Eqs. 1,6,10 trained with the KRR model without feature rescaling ($\rho = 0.12$, $\tau = 0.07$, Eq. 12).

In case of the HSP90 dataset, we obtained very low correlation coefficients. More precisely, we submitted four different binding affinity predictions following the same strategy as in the previous dataset. Re-scoring with the Convex-PL potential resulted in $\rho = 0.07$ and $\tau = 0.05$, re-scoring with the en-

tropic term trained with the LRR model (Eq. 11) without feature rescaling gave $\rho = 0.03$ and $\tau = 0.02$, re-scoring with the free energy approximation trained with the LRR model (Eq. 12) without feature rescaling gave $\rho = 0.07$ and $\tau = 0.05$, re-scoring with the free energy approximation trained with the KRR model (Eq. 12) without feature rescaling gave $\rho = 0.15$ and $\tau = 0.12$. The quality of the predictions is very low presumably because most of the co-crystal structures for this dataset were not available. Indeed, in our experiments we used the top-1 putative docking predictions for 180 protein-ligand complexes. Even though we obtained very high quality predictions for six complexes with experimentally solved structures (with a mean RMSD to crystallographic solution of about 1 Å), there is absolutely no guarantee that the same quality of predictions holds for the rest of 174 structures in this set.

To conclude the discussion, we would like to speculate about interpretability of the experimental affinity data. Indeed, IC_{50} is a measure of inhibition of a specific process and K_i can be used to relate it to binding affinity, but solely for competitive agonists and antagonists. On the other hand, K_d is related to the binding energy. Since binding does not necessarily result in inhibition and conversely, the conversion between IC_{50}/K_i and K_d is non-trivial. Therefore, since both the model selection and fitting were performed on $-\log(K_d)$, it is not unexpected that predictions of IC_{50}/K_i are not accurate including the very different scales.

5.4 Conclusion

In this challenge we assessed our docking pipeline and tested a new model to predict binding affinity. Our docking predictions appeared to be successful, especially for the HSP90 set. Nonetheless, we experienced certain difficulties in sampling large conformational space for long ligands of the MAP4K set, and this was the main source of inaccuracies in our pose predictions. We critically assessed the quality of pose predictions depending on the choice of rigid receptor templates, the number of flexible residues in the binding pocket, the binding pocket size, and the subsequent re-scoring with the Convex-PL potential. We demonstrated that re-scoring with the Convex-PL potential significantly improves at least the quality of the best pose among the top 5 poses. We also showed that it is very important to choose the right receptor template and to reduce the size of the binding pocket (consequently,

the size of the search space) as much as possible. Finally, we showed that flexibility of protein side chains inside the binding pocket does not improve the quality of docking poses.

However, our main goal in this challenge was to test the free energy model that consists of the enthalpic term and the entropic term. Free parameters of the model were deduced from the PDBBind database using regression procedures. Even though our model performed very well on the training set, its success rates on the MAP4K and, especially, HSP90 test sets were rather low. The mild performances of our model owe to several facts. First, the very large number of free parameters in the enthalpic term possibly causes overfitting of training data. Second, as in any regression problem, the coverage of the test set by the training set appears critical to ensure performance guarantees—as shown by our post hoc analysis. Given this observation, an interesting research direction to be explored is the application of other regression methods, possibly more tolerant to mild coverages of the parameter space. Finally, it should be kept in mind that binding affinity is a subtle thermodynamic property, calling for an exhaustive analysis of dynamics, including conformational entropy, vibrational entropy, and solvent entropy. While recent work has shown that reliable proxies existed for the latter two terms, accurately accounting for changes in conformational entropy remains an open problem. Filling this gap calls for efficient sampling procedures able to reveal low regions of the energy landscape of the system, together with their occupancy probabilities. As noticed above, such procedures would also be highly valuable to sample large conformational space.

To conclude, this work calls for two complementary extensions aiming at fostering our understanding of the emergence of function (i.e., binding) from a combination of structure and dynamics. The first one will address cases where compact models (such as our entropic model) can be used reliably to estimate affinity from static pauses. The second one will aim at going beyond such cases, so as to obtain equally reliable estimates exploiting conformational ensembles accurately coding more general entropic changes (in particular conformational entropy).

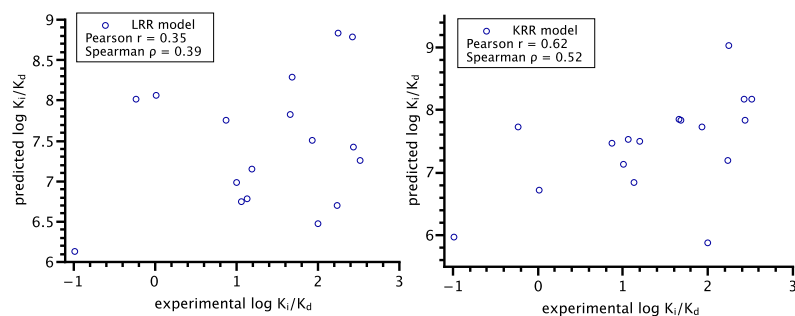


Fig. 5 Accuracy of the $\Delta H - T\Delta S$ model on the test set MAP4K. Left : training of the linear model. Right : training of the non-linear model.

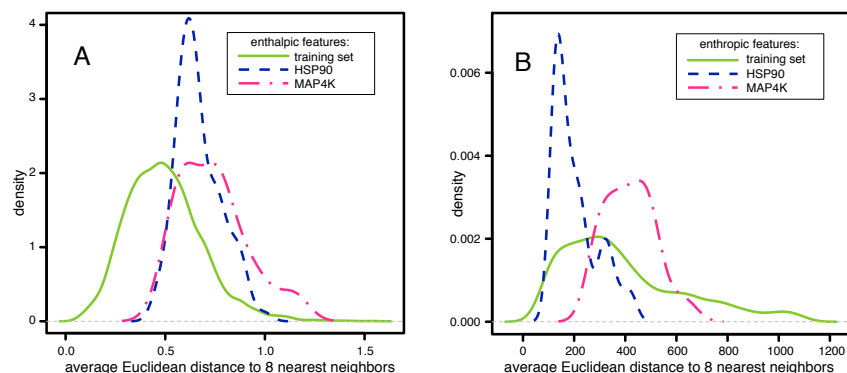


Fig. 6 Coverage of the test set by the training set assessed with pairwise distances. See section 5.3 for details. (A) Distribution of distances for the 11760 parameters introduced in section 5.3 (B) Distribution of distances for the five parameters introduced in section 3.3

Acknowledgments

The authors thank Dr. Petr Popov from MIPT Moscow for the initial analysis of the HSP90 targets.

References

- R.D. Smith, J. Dunbar, J. B., P.M. Ung, E.X. Esposito, C.Y. Yang, S. Wang, H.A. Carlson, *J. Chem. Inf. Model.* **51**, 2115 (2011)
- K.L. Damm-Ganamet, R.D. Smith, J.B. Dunbar Jr, J.A. Stuckey, H.A. Carlson, *J. Chem. Inf. Model.* **53**(8), 1853 (2013)
- S. Grudinin, P. Popov, E. Neveu, G. Cheremovskiy, *J Chem Inf Model* (2015). DOI 10.1021/acs.jcim.5b00339
- T.D. Crawford, C.O. Ndubaku, H. Chen, J.W. Boggs, B.J. Bravo, K. Delatorre, A.M. Giannetti, S.E. Gould, S.F. Harris, S.R. Magnuson, E. McNamara, L.J. Murray, J. Nonomiya, A. Sambrone, S. Schmidt, T. Smyczek, M. Stanley, P. Vitorino, L. Wang, K. West, P. Wu, W. Ye, *J Med Chem* **57**(8), 3484 (2014). DOI 10.1021/jm500155b
- N. Homeyer, H. Gohlke, *Journal of computational chemistry* **34**(11), 965 (2013)
- L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M.K. Dahlgren, J. Greenwood, et al., *Journal of the American Chemical Society* **137**(7), 2695 (2015)
- L. Wang, B. Berne, R.A. Friesner, *PNAS* **109**(6), 1937 (2012)
- M.F. Lensink, S. Velankar, A. Kryshtafovych, S.Y. Huang, D. Schneidman-Duhovny, A. Sali, J. Segura, N. Fernandez-Fuentes, S. Viswanath, R. Elber, S. Grudinin, P. Popov, E. Neveu, H. Lee, M. Baek, S. Park, L. Heo, G. Rie Lee, C. Seok, S. Qin, H.X. Zhou, D.W. Ritchie, B. Maigret, M.D. Devignes, A. Ghoorah, M. Torchala, R.A.G. Chaleil, P.A. Bates, E. Ben-Zeev, M. Eisenstein, S.S. Negi, Z. Weng, T. Vreven, B.G. Pierce, T.M. Borrmann, J. Yu, F. Ochsenbein, R. Guerois, A. Vangone, J.P.G.L.M. Rodrigues, G. van Zundert, M. Nellen, L. Xue, E. Karaca, A.S.J. Melquiond, K. Visscher, P.L. Kastiris, A.M.J.J. Bonvin, X. Xu, L. Qiu, C. Yan, J. Li, Z. Ma, J. Cheng, X. Zou, Y. Shen, L.X. Peterson, H.R. Kim, A. Roy, X. Han, J. Esquivel-Rodriguez, D. Kihara, X. Yu, N.J. Bruce, J.C. Fuller, R.C. Wade, I. Anishchenko, P.J. Kundrotas, I.A. Vakser, K. Imai, K. Yamada, T. Oda, T. Nakamura, K. Tomii, C. Pallara, M. Romero-Durana, B. Jiménez-García, I.H. Moal, J. Fernández-Recio, J.Y. Joung, J.Y. Kim, K. Joo, J. Lee, D. Kozakov, S. Vajda, S. Mottarella, D.R. Hall, D. Beglov, A. Mamonov, B. Xia, T. Bohnuud, C.A. Del Carpio, E. Ichiishi, N. Marze, D. Kuroda, S.S. Roy Burman, J.J. Gray, E. Chermak, L. Cavallo, R. Oliva, A. Tovchigrechko, S.J. Wodak, *Proteins* (2016). DOI 10.1002/prot.25007
- P. Popov, S. Grudinin, *J Chem Inf Model* **55**(10), 2242 (2015). DOI 10.1021/acs.jcim.5b00372

10. S. Marillet, P. Boudinot, F. Cazals, *Proteins: structure, function, and bioinformatics* **1**(84), 9 (2015). DOI 10.1002/prot.24946. URL <https://hal.inria.fr/hal-01159641>
11. P. Kastritis, I. Moal, H. Hwang, Z. Weng, P. Bates, A. Bonvin, J. Janin, *Protein Science* **20**, 482 (2011)
12. S.Y. Huang, X. Zou, *Proteins: Struct., Funct., Bioinf.* **72**(2), 557 (2008). URL <http://onlinelibrary.wiley.com/doi/10.1002/prot.21949/full>
13. G.Y. Chuang, D. Kozakov, R. Brenke, S.R. Comeau, S. Vajda, *Biophys. J.* **95**(9), 4217 (2008)
14. V.N. Maiorov, G.M. Grippen, *J. Mol. Biol.* **227**(3), 876 (1992)
15. J. Qiu, R. Elber, *Proteins: Struct., Funct., Bioinf.* **61**(1), 44 (2005)
16. R. Rajgaria, S. McAllister, C. Floudas, *Proteins: Struct., Funct., Bioinf.* **65**(3), 726 (2006)
17. D. Tobi, I. Bahar, *Proteins: Struct., Funct., Bioinf.* **62**(4), 970 (2006)
18. D. Ravikant, R. Elber, *Proteins: Struct., Funct., Bioinf.* **78**(2), 400 (2010)
19. M.H. Chae, F. Krull, S. Lorenzen, E.W. Knapp, *Proteins: Struct., Funct., Bioinf.* **78**(4), 1026 (2010)
20. G. Neudert, G. Klebe, *Bioinformatics* **27**(7), 1021 (2011)
21. L. Lo Conte, C. Chothia, J. Janin, *JMB* **285**(5), 2177 (1999)
22. J. Janin, R.P. Bahadur, P. Chakrabarti, *Quarterly reviews of biophysics* **41**(2), 133 (2008)
23. F. Cazals, F. Proust, R. Bahadur, J. Janin, *Protein Science* **15**(9), 2082 (2006). DOI doi:10.1110/ps.062245906
24. S. Lorient, F. Cazals, *Bioinformatics* **26**(7), 964 (2010). DOI 10.1093/bioinformatics/btq052. URL <http://hal.inria.fr/hal-00849822>
25. M. Gerstein, F. Richards, in *The international tables for crystallography (Vol F, Chap. 22)*, ed. by M.G. Rossmann, E. Arnold (Springer, 2001), pp. 531–539
26. F. Cazals, H. Kanhere, S. Lorient, *ACM Transactions on Mathematical Software* **38**(1), 1 (2011). DOI 10.1145/2049662.2049665. URL <http://hal.inria.fr/hal-00849809>
27. G. Meng, N. Arkus, M. Brenner, V. Manoharan, *Science* **327**(5965), 560 (2010)
28. J. Dunitz, *Chemistry & biology* **2**(11), 709 (1995)
29. P. Kastritis, J. Rodrigues, G. Folkers, R. Boelens, A. Bonvin, *J.M.B.* **426**, 2632 (2014)
30. D. Eisenberg, M. Wesson, M. Yamashita, *Chem. Scr. A* **29**, 217 (1989)
31. B. Bouvier, R. Grunberg, M. Nilgès, F. Cazals, *Proteins: structure, function, and bioinformatics* **76**(3), 677 (2009). DOI 10.1002/prot.22381. URL <http://hal.inria.fr/hal-00796032>
32. R. Wang, X. Fang, Y. Lu, C.Y. Yang, S. Wang, *J. Med. Chem.* **48**(12), 4111 (2005). DOI 10.1021/jm048957q. URL <http://www.ncbi.nlm.nih.gov/pubmed/15943484>
33. R. Wang, X. Fang, Y. Lu, S. Wang, *J. Med. Chem.* **47**(12), 2977 (2004). DOI 10.1021/jm030580l. URL <http://www.ncbi.nlm.nih.gov/pubmed/15163179>
34. D. Barber, *Bayesian reasoning and machine learning* (Cambridge University Press, Cambridge, 2012). URL <http://assets.cambridge.org/97805215/18147/cover/9780521518147.jpg>
35. S. Sonnenburg, G. Rätsch, S. Henschel, C. Widmer, J. Behr, A. Zien, F.d. Bona, A. Binder, C. Gehl, V. Franc, *The Journal of Machine Learning Research* **11**, 1799 (2010)
36. M. Kadukova, S. Grudin, *J Chem Inf Model* **56**(8), 1410 (2016). DOI 10.1021/acs.jcim.5b00512
37. O. Trott, A.J. Olson, *J. Comput. Chem.* **31**(2), 455 (2010)
38. D. Seeliger, B.L. de Groot, *J Comput Aided Mol Des* **24**(5), 417 (2010). DOI 10.1007/s10822-010-9352-6
39. N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, *J. Cheminform.* **3**, 33 (2011)
40. A. Heifets, R. Lilien, *Journal of Molecular Graphics and Modelling* **29**(1), 93 (2010). DOI 10.1016/j.jmgm.2010.05.005
41. Schrödinger, LLC, (2011)
42. L. Györfi, A. Krzyzak, *A distribution-free theory of nonparametric regression* (Springer, 2002)