UNIVERSITAT DE BARCELONA

Virtual screening for novel mechanisms of action: applications and methodological developments

Sergio Ruiz Carmona

Universitat de Barcelona
Facultat de Farmàcia
Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica

Virtual screening for novel mechanisms of action: applications and methodological developments

Sergio Ruiz Carmona

2016

Programa de Doctorat en Biomedicina, Universitat de Barcelona
Director de Tesi: Dr. Xavier Barril Alonso

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

DEPARTAMENT DE FARMÀCIA I TECNOLOGIA FARMACÈUTICA I FISICOQUÍMICA

PROGRAMA DE DOCTORAT EN BIOMEDICINA

VIRTUAL SCREENING FOR NOVEL MECHANISMS OF ACTION: APPLICATIONS AND METHODOLOGICAL DEVELOPMENTS

Aquesta tesi ha estat realitzada per Sergio Ruiz Carmona sota la direcció del Dr. Xavier Barril Alonso, Professor d'Investigació ICREA en el Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica de la Facultat de Farmàcia de la Universitat de Barcelona. Es presenta aquesta memòria per optar al títol de doctor per la Universitat de Barcelona en el Programa de Doctorat en Biomedicina.

Xavier Barril Alonso
Director de tesi

Sergio Ruiz Carmona
Doctorand

Sergio Ruiz Carmona
Setembre 2016

A vosotros papa y mama.
Aunque no entendais nada,
sin vosotros no habría nada que entender.

Acknowledgements

Primer de tot m'agradaria donar-li les gràcies al Xavi. Des de que vaig començar el màster ja fa més de 5 anys fins l'últim dia de la tesi has estat no només un director, sinó també una referència per mi en tot aquest projecte. En la meva humil opinió, crec que és molt difícil dirigir un grup de persones i més encara ser un bon líder, però tu has sabut fer-ho i, a més a més, fer-me gaudir (i molt) de tota l'experiència.

Cuando entré en el grupo me encontré con la primera generación: Jesús, Peter y Mousumi ya estaban a punto de acabar. Gracias por hacerme fácil algo que en aquel momento era un cambio total para mi tranquila vida de clases y exámenes. Després, i durant la major part d'aquesta tesi, vaig compartir les hores amb la segona generació: la Montse i el Dani. Moltes gràcies als dos per tot el temps que hem passat junts davant de l'ordinador, el Biacore i les cerveses del bar. I al Pep i la Natàlia, per ensenyar-me a "matar el temps" de manera productiva a Biologia. Más tarde el grupo se fue ampliando con Aida, Laura, Yvonne y Kevin (y luego las demás Minoryxtas). Gracias por vuestro granito de arena con el que me habeis ayudado a realizar esta tesis, sobretodo con los crucigramas a la hora de comer!

Con el cambio a Farmacia, he podido compartir más tiempo con Ornella, Marina, Sonia y Silvana. Muchas gracias a vosotras por acogernos en el ático y hacernos sentir mejor incluso que en nuestra antigua casa, es una lástima no haber hecho la mudanza un año antes! Y aunque hayamos estado más lejos gracias también a todos los Javis: Toni, William, Constantí, Javi, Jordi, Salo, Carol, Flavio y Ana. Tampoco me olvido de los visitantes argentinos Juan y Elias.

Gràcies també als seniors del grup: a Campa, Curu, Ramon (i les seves calçotades) i a l'Axel per aguantar-me a mi i les nostres baralles amb els ordinadors. Y sobretodo a Javi, por ayudarme cuando lo he necesitado.

I ja finalment, la tercera generació: Míriam, Serena i Guillermo primer i Maciej i Moira després. Ha estat molt fàcil crèixer amb vosaltres. Fins i tot amb el Carles! Bromes a part, des de que vas arribar que tot ha evolucionat positivament, moltes gràcies a tu també.

Por otro lado, no me quiero olvidar de los *muggles*. A mi familia, gracias papa y mama por esas comidas regulares en el japo y el Andreu. Ahora papa espero que dejes ya la clásica broma de "estudiar" (as it can be seen in Figure 1)! Gracias Carlos, Guillermo, Bruno y Adriana por aguantarme todos estos años. Sobretodo a los Cool Bros, sois los mejores. Os quiero mucho a todos.

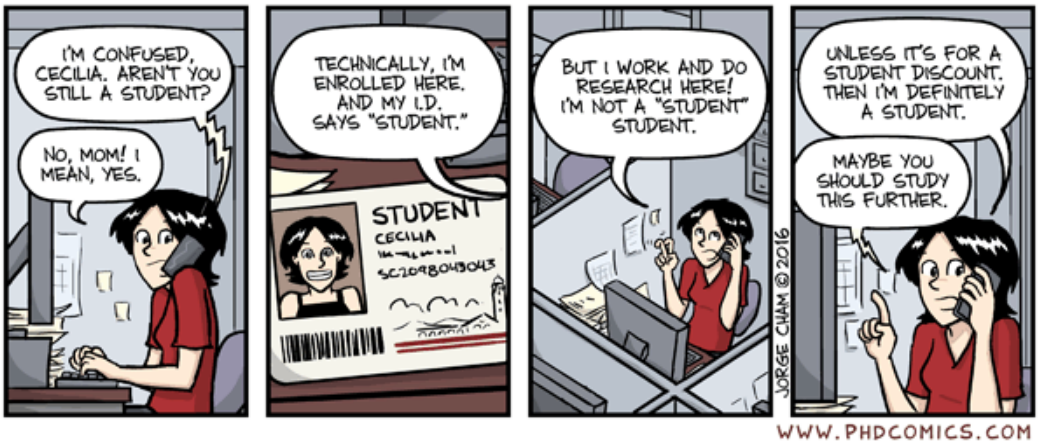


Figure 1: Thanks also to PhD Comics for the word "procrastination".

També als meus amics, amb qui he compartit gran part del temps fora de la pantalla

negra. A l'Arnau (i la Glòria) i tot el temps a Neopàtria. Al Pau (com sempre, "gràcies per convidar-me") i l'Aitor, el Josep, el Marc i el Sevi. Als companys de la uni, Berta, Dr. Santi, Marçal, Claudia, Laures, Maria i Cris. I al futbol. Moltes gràcies a tots per fer-me la vida menys avorrida.

I per acabar, a tu, Rosa. Perquè no hagués pogut fer això sense tu, perquè m'has ajudat tots els dies, perquè m'has escoltat quan ho he necessitat encara que no entenguessis res del que et deia, perquè vam començar esta aventura separats i l'acabarem junts i perquè sí.

Después de los agradecimientos personales me gustaría acabar con una cita que escuché decir a alguien que estoy seguro que era muy sabio, aunque ahora no recuerde el nombre completo, como agradecimiento a todo el mundo anónimo que queriendo o sin querer me ha ayudado con esta tesis.

I know you believe you understand what you
think I said, but I am not sure you realize that
what you heard is not what I meant

Gaussian 09 Rev. E01

Contents

Chapter 1

Introduction

Isaac Asimov said a long time ago that "The saddest aspect of life right now is that science gathers knowledge faster than society gathers wisdom", which I think it is still true nowadays. Science moves forward at a very high rate and, to a certain degree thanks to the boom of computational science during the last decades, it is something that will be continued in the next few years.

During this thesis, whose work is presented over the next chapters, I feel that I have contributed to this progress and I strongly hope the reader gets convinced too.

1.1 Overview of Drug Discovery and Development

It is estimated that the drug discovery and development process, on average, can take up to 10 or 15 years to get a drug from early stages of research to being commercialized to the public, with an associated cost of more than 500 million dollars.

Several stages have to be undertaken from starting point until getting a drug to the market (summarized in Figure 1.1).

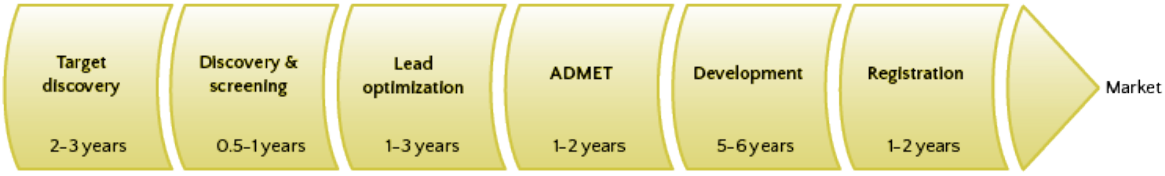


Figure 1.1: Scheme of the different stages and corresponding times that have to be dealt with in the development of a drug.

In Target Discovery, the process implies finding out the target that causes a particular disease of interest. Next, putative drugs (drug-like molecules [1, 2], biological compounds, natural products, etc.) are screened experimentally against these identified targets in order to find hits, which will be optimized to find lead compounds and develop them to drugs. The lead compounds undergo a more exhaustive optimization during Lead Optimization and ADMET (abbreviation for absorption, distribution, metabolism, excretion and toxicity) stages. Many iterations in these stages involving *in vitro* and *in vivo* assays, syntethic chemistry, genomic technology or bioinformatics can increase the length of this period for up to 5 years.

After that, the lead compound has been optimized and becomes a more likely drug, which will then be carried to the development stage of clinical trials. These are experiments carried on humans and are divided in different stages, as depicted in Figure 1.2. In phase I, a safety screen is performed within tens of people in order to establish safe dosage ranges and to identify early visible side effects. In phase II, a comparison with a placebo determines the efficacy of the drug. Moreover, in a larger group of people, uncommon side effects will be easier to spot. Phase III serves as the final confirmation of safety and efficacy whereas side effects are still monitored in thousands of participant people.

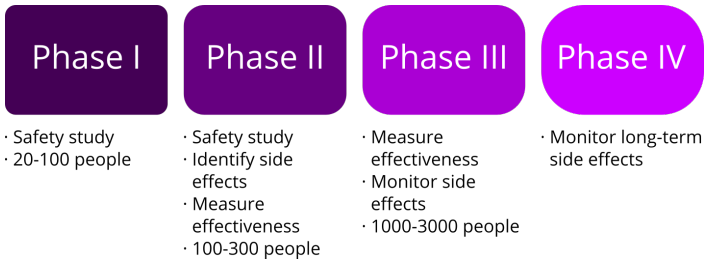


Figure 1.2: Summary of all phases of the clinical trials that a drug has to overcome.

Finally, the last step is the registration and approval. A new drug has to be reviewed and approved by the different regulatory authorities, which check that the safety, efficacy and potency is well demonstrated. After the approval of the drug and after it has been marketed, the phase IV of the clinical trials monitors the long-term side effects and if non-investigated adverse events appear. These studies are carried on during sales, and are used to optimize, when necessary, the risks, benefits and indications.

1.2 Main approaches in Drug Discovery

There are different procedures to approach the Drug Discovery and Development, which could be classified into different categories:

- Classical approach
- Rational approach
- Gene therapy
- Biologics
- ...

The classical approach is the most used one in history of Drug Discovery and most drugs available today have been discovered by applying so. It involves experimental observations of the effects from testing chemical compounds. Thousands of compounds are tested in a similar way thousands of keys should be tried in order to open a given lock. The positive hits are isolated and characterized in detail and further optimization is carried on.

The rational approach, in contrast with the classical one, needs previous knowledge of the target structure. Drugs are then designed to make an interaction with the target structure in order to cause a beneficial outcome. This is the approach taken during this thesis and more details about it will be shown in the next section.

Gene therapy and biologics are relatively hot topics with much less cases with respect to the previous ones. Gene therapy aims at treating a disease by inserting a missing gene or correcting a malfunctioning one. The final goal is to alter the disease pathway or to restore the missing proteins or enzymes. Biologics are mainly antibodies, vaccines or proteins that are designed to act as drugs. The biologic drugs are manufactured externally and inserted to the body to perform their function.

1.3 Rational Approach

The main feature of rational approach is the necessity to know the 3D structure of the receptor as well as the structure of the chemical structures being tested. X-ray crystallography or NMR spectroscopy can help in solving the three-dimensional structure of the proteins, but, in the cases where that is not possible, molecular modelling can predict the structure.

The role of computers

Computers have been used in biology and chemistry since a long time ago. Year after year, computers evolve and methods are improved and widely applied. In recent years, computers have helped in scanning DNA sequences to determine location of genes, determine possible functions and structures of proteins, predict binding sites for drug interactions and provide information for drugs to be designed to fit the binding sites.

1.4 *In Silico* Drug Discovery

More specifically, I will focus on how computers aid in the Drug Discovery field. The main stages where they are applied are the early parts of Drug Discovery and Development pipeline, as shown in Figure 1.3.

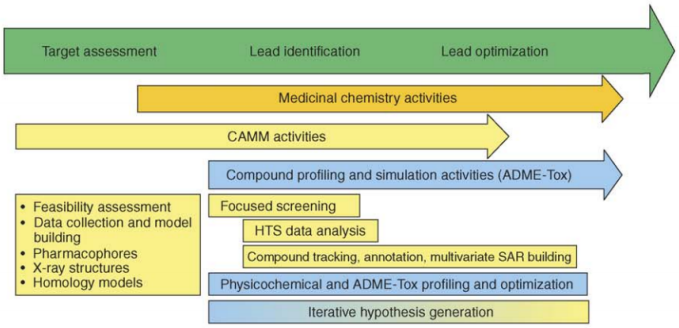


Figure 1.3: Role of computers in Drug Discovery and Development.

It encloses a different number of techniques that deal with different aspects such as studying the molecular basis of ligand-protein interactions, develop target-specific compound libraries, model target proteins, identify hits by ligand and structure-based virtual screening, estimate binding free energy, and optimize lead compounds, all of which can be used to rationalize and increase the efficiency, speed, and cost-effectiveness of the drug discovery process. The increment of such methods, the algorithmic and software development, the large number of web servers and the decreasing cost of computational power have contributed to the success of computational drug lead discovery.

However, it is reasonable to think that more accurate and reliable methods would surely help to overcome the stagnation in the number of approved drugs in recent years [3], especially if *in silico* drug discovery is coupled with druggability assessments early in the drug discovery process.

1.5 Structure Based Drug Design

With current developments in bioinformatics, gene-sequencing and molecular biology techniques, drug targets are identified at an increased pace and the limiting step is finding the appropriate drug. Within target-based discovery, owing to the availability of the target atomic structure, two approaches are identified: **structure based drug design** (SBDD) and **ligand based drug design** (LBDD). Structure-Based Drug Design encompasses all the tools and techniques exploiting the target's structural information to rationally guide the design process [4]. On the other hand, when the target structure is not available, Ligand-Based Drug Design, or ligand based approaches, exploit the information contained in the active hits found during *in vitro* experimental assays. Therefore, second approach is usually employed when some active ligands are already known and SBDD is the preferred tool in initial hit finding.

The holy grail of SBDD is that, knowing the target structure, we have all the information needed to design effective, selective and safe drugs. But we are not still there: a gap between structure and successful drug is present. Are we lacking target structures or is the process of understanding the information failing?

Thanks to ultimate technology advances (e.g. more precise and clear X-ray light, more potent electromagnetic fields, etc.), crystallography and Nuclear Magnetic Resonance (NMR)

techniques can now easily elucidate many macromolecule's structure with a very low resolution. The evolution of these techniques is directly linked to the number of structures released and, as shown in Figure **??**, the exponential growth is notorious. Currently, the Protein Data Bank or PDB (main database for macromolecular structures) contains more than 100.000 structures. All this wealth of structural information and the expected near future advances make the target structure not the problem in SBDD.

It is our understanding on the fundamental process of molecular recognition in biological systems which is still limited, and thus are the predictions, which generate the drug design guiding hypothesis. It is still a hot topic how to correctly use the information contained within the target structure and its interactions with the ligands in the more appropriate manner.

1.5.1 Target identification, validation and structure elucidation

First crucial step in drug discovery processes is the correct selection of the appropriate target entity responsible for the disease or effect we are willing to modulate. The target should be responsible for the activity and its modulation should produce the desired effect without altering normal functioning pathways in the organism. In this stage, bioinformatic genetic studies are increasingly contributing to the discovery of potential drug targets [5]. After a target is selected and approved for its correct behavior (i.e. knock-out models should reproduce the drug effect), obtaining a good atomic resolution structure is probably the most important task.

Usual techniques (crystallography and NMR) cannot grant the success on difficult to produce and purify systems yet (e.g. membrane receptors, ion channels). In some situations where experimental techniques fail, it is still possible to predict the structure of the target in study if the target has a close homolog protein with already known structure.*Homology modeling*, the computational discipline in charge of such predictions, finds its roots in the lower diversity of protein folds than sequences: current SCOP classification (one of the main authorities in protein tertiary structure classification) identifies 1390 different folds independently on the origin organism; whereas only human body is estimated to have about 50000 different proteins. It is considered that above a 30% sequence identity it is possible to obtain a reliable structure estimate [6].

If it is not possible to be obtain the target structure, ligand based approaches (e.g. QSAR models, pharmacophore identification) have been successfully and widely applied before the technology was able to yield as many protein structures as we have now. Ligand approaches are still a powerful tool in combination with structure-based approach, specially during the optimization stages where potency and safety should be improved.

1.5.2 Target structural analysis

Druggability and cavity detection

One important task to reduce attrition rates in drug discovery, that was somehow omitted in early projects, is the assessment of the target's druggability. The term druggability is understood as the suitability of a target for binding drug-like molecules.

A correct classification of a macromolecule as either druggable or undruggable helps to choose the appropriate target in early stages of the drug design process, discarding those which are presumably more difficult to be modulated by drug-like molecules. Good predictions have an evident economical impact on the project and help to direct all efforts to most promising targets.

The assessment of druggability can be done experimentally or computationally. Experimental methods include the retrospective analysis of the hit rates in high-throughput screening or fragment screening campaigns[7]. But obviously, these approaches are not applicable in a prospective manner. The most promising approaches are all computational. Based on a correct cavity identification and characterization, usually by the definition of some descriptors, many methods have been developed to help in this assessment. A dataset of known proteins already classified as druggable or not is used to train and validate the predictive models. Schrodinger introduced a druggability score in their proprietary *SiteMap* cavity prediction software[8] and *fpocket*, open-source software, based on an extended dataset also proposed a druggable score[9].

Druggability, however, is not understandable without a previous *cavity* or *binding site* definition and, actually, both tasks are usually linked. Cavity identification might be obvious when targeting enzymes or receptors by classical mechanisms of action (e.g. competitive agonism or antagonism) but, when targets are not naturally evolved to bind a small molecule substrate or one seeks non-competitive mechanisms of action, this task becomes complex (e.g. protein-protein interfaces). Several computational tools, mostly based on geometrical and shape parameters, have been developed to aid in the binding site identification and selection.[10] For instance, **fpocket** or **SiteMap** previously mentioned are two representative of these approaches.

Binding site characterization

Besides cavity identification, the correct characterization of the pocket will likely increase the probabilities of success. Binding sites specify structural and physicochemical constraints that must be met by any putative ligand. Hence, it is imperative to analyze the constitution of the binding site by mapping the characteristics that are essential for ligand recognition. This is particularly relevant during the project first stages when still no ligand has been identified.

1.5.3 Molecular Docking: Virtual screening and compound ranking

After the binding site has been identified, the normal process will follow with the docking and scoring of a virtual compound library [11, 12, 13]. This process, known as *virtual screening*, aims to identify potential binders from a pool of virtual molecules [14]. If the scoring process is correct, molecules with the lower energy values ranked at the top, should be the most active. The advantage of this computational tool over experimental high throughput screening is obvious: being virtual, there is no need to set up an assay, purchase nor synthesize all the compounds to be assayed, with implied time saving and reduced costs [15, 16].

Space sampling and receptor definition

Taking into account that millions of poses for each molecule and thousands to millions of different molecules should be evaluated, the limiting factor of virtual screening is speed. Therefore, it is not possible to exhaustively explore all possible receptor or ligand geometries. For this reason, the receptor is usually considered as a rigid body and, in the best cases, some of the binding site side chains are allowed to move or hydrogen atoms to rotate. Even then, the virtual screening process is still far from identifying induced binding sites or complex ligand-protein conformational changes.

Moreover, it is known that water molecules play an important role in the binding process of many drug-target complexes. However a common task in virtual screening is to remove all

receptor solvating waters if there is not enough information pointing otherwise. The outcome will largely depend on their presence, and will likely fail if important waters were removed. More detailed discussion on this topic is presented in the next section.

Scoring functions

On the other side, accurate energy estimation is still a major problem to be addressed in the field of computational chemistry. Usually, costly and complex computations (e.g. free energy methods) are needed to roughly estimate binding affinities with an acceptable error. The process involves molecular simulations and hours of computation for a single molecule. It is therefore impossible to apply such accurate methods in a high-throughput manner, as it would be required in virtual screening. The solution is to use empirically or statistically derived scoring functions which try to offer the best balance between accuracy and speed of calculation. These functions can evaluate a single pose in milliseconds.

Current scoring functions include several terms describing the ligand-receptor interactions from a mechanical point of view. Many terms are used to build current scoring functions; each one introduced with as many variations as scoring functions exists: Van der Waals for steric effects, Coulomb for electrostatic interactions, geometrical terms for hydrogen bonds and a desolvation term. All these terms are parametrized according to a training set. That is, taking into account previous knowledge, a weighing factor is applied to each term in order to reproduce the experimental set results. Then, if the function is not over-fitted, it will be able to predict binding modes in novel systems not part of the training set.

But, what if we try to target binding sites very different from any training set used up to date? Will it be possible to find allosteric or protein-protein inhibitors using classically parametrized scoring functions (i.e. functions working well on natural binding sites)?

Professor Gisbert Schneider defines Virtual Screening as an endless staircase[17]: despite all continued developments, still its impact on the success of drug discovery projects is controversial. The lack of novel approaches is a major drawback to advance in the field of computer aided drug design.

Including experimental information boosts results

Once first active ligands are identified, the determination of most and less important interactions provides valuable information for guiding the virtual screening process. For instance, if some mutational study highlights the importance of certain aminoacid for the ligand binding, it is possible to introduce some bias in the virtual process to give higher score to ligands fulfilling that interaction. If we have a resolved crystal structure and some water is mediating the interaction, its consideration in the docking process will likely yield better results. Also if there are several active ligands with known binding mode, a *pharmacophore* can be derived, to guide future virtual screening processes and improve the outcome.

1.5.4 Experimental assay and iterations

As mentioned before, scoring functions are somehow limited and being at the top of the ranked ligand list, still does not guarantee a molecule will be active in experimental assays. For this reasons, an expert will have to carefully examine the top ranking compounds and select those which are more likely to be active based on his own experience. To somehow

simplify this process, a common task is to re-score the list using several scoring functions and finally sort the list with a consensus score[18].

Once a list is proposed, the compounds will be purchased or synthesized and tested in the most suitable biologic assay. Active and inactive compounds information will be useful to establish an initial structure-activity relationship (SAR) to guide future design. This SAR is constructed to determine which are the optimal ligand structure for gaining the best potency possible. Several informatic tools (e.g. QSAR, 3D-QSAR) are used to build predictive quantitative models and estimate possible affinities before any compound is synthesized and tested.

The active ligand list can be further optimized by looking for similar scaffolds or evolving the structure in a synthesis campaign guided by the binding mode hypothesis and the primitive SAR. Knowing what interactions are more or less important will determine the improvement success. The process iterates and the more molecules are found active, the more precise the models and the SAR become and more potent ligands are likely to be found.

Not only the potency is important. Safety is also a crucial aspect of a good drug candidate. Thus ligand structure is in parallel also optimized for having good ADME properties and be less likely to fail in future drug discovery stages.

MDmix

In 1997, Leipinsh and Otting described the presence of small organic molecules in the binding site of Hen Egg White Lysozime (HEWL) in their NMR studies [19]. Not only they described the unspecific ability of DMSO, methanol, acetonitrile and other small co-solvents to bind in the protein active site but they could also determine their binding affinity. Multiple solvent crystal structures (MSCS) methodology, published in 2006 by Mattos *et al.*, was also presented as a method to determine binding sites and hots pots using different organic solvent mixtures [20, 21]. This unspecificity is not surprising if we understand binding sites as regions naturally designed to be desolvated for binding substrates, and the co-solvent molecules as a minimum representation of a drug-like molecule.

As discussed in previous sections, the correct identification and characterization of binding sites is crucial for conducting a successful drug design campaign, and these techniques can clearly help in such determination. However, several limitations hamper their applicability: it must be feasible to produce protein and determine its structure in the presence of the co-solvents. Unfortunately, this is not possible yet for many systems.

Consequently, in last years, our group developed a computational method (MDmix)[22, 23] which uses molecular simulations to mimic the experimental techniques. As an *in silico* version of previous experiment with aqueous-organic mixtures by Otting, Mattos and other, the method simulates the behaviour of the target protein in a solvent mixture (an isopropanol-water mixture in this case). The whole system is modeled and simulated using classical Molecular Dynamics which allows the system to move and evolve in time. By calculating the regions in the space with higher occupancy of isopropanol molecules, the study aims at identifying binding sites and to estimate the maximum binding affinity a perfect ligand would attain upon binding. This latter measure was given as an indicator of druggability (i.e. more druggable binding sites will have lower estimated energies).

The advantages of computational techniques in this area are obvious and, as proven elsewhere[24], the simulation results are often as reliable and robust as experimental methods, giving higher confidence on their predictive power.

The workflow

Two main actors are needed to setup a calculation: the solvent mixtures (selected according to the probes we are interested in) and the target macromolecule (usually a protein but it could also be a nucleic acid chain). After choosing the simulation parameters (e.g. temperature, number of replicas, time of simulation, etc.), the calculations can be submitted. The process of Molecular Dynamics simulations, for those readers not familiar with it, can be seen as a movie making process: the atoms move during time and their movement is captured in a trajectory.

All results are contained within these films or trajectories and the analysis process in the lower box in First step is the alignment of all the replica's trajectories to a single reference frame, usually the starting configuration of the protein. Then, superposing a rigid grid that partitions the space, it is counted how many times the probe atoms in the co-solvent fall in each of the small space partitions (the grid points or voxels). This way a density map is obtained. For instance, in a simulation with ethanol and water mixture, a density map for the oxygen in the ethanol will represent hydroxyl interactions (or hydrogen bond donor and acceptor interactions in general) and the tail carbon will show hydrophobic interactions [25, 26, 27].

Steered Molecular Dynamics

Steered molecular dynamics can be used for predicting affinity of small molecules upon binding to proteins [28]. However, this process can get really hampered by the difficulty in identifying a valid reaction coordinate, which can be very complex due to the size if the protein and the diversity of potential ligand molecules. In this thesis, we propose an alternative approach that consists in a reduction of the system size focused around a key interaction point [29], which facilitates the choice of a reaction coordinate, decreases the time of the simulations and allows us to differentiate between active and inactive compounds in a relatively high throughput scenario.

This and other methods are applied in real Drug Discovery campaigns. The development and real implementation of all of them is presented and discussed in this thesis. Moreover, challenging targets demand better virtual screening methods: In this direction, the development of such methods has been an important part of this thesis, whereas the application of these methods in real challenging systems is one of the main motivations. A final discussion and conclusions covers the main concepts introduced here and in the rest of the chapters.

Chapter 2

Objectives

As discussed in the introduction, better and novel methods are needed to improve drug discovery and, in particular, computer-aided drug discovery. In our group, rDock has always been the main docking tool for all our projects. However, it lacks a comparison with other programs that are commonly used as well as a public release after the evolution from Ribotarget's RiboDock. Moreover, docking programs are usually far from perfect and a lot of noise is present in their results: false negatives and, more importantly, false positives increase the time and economic cost of drug discovery campaigns. Complementary methods that can help in this aspect would be extremely useful.

Hence, the global objective of this thesis is to develop, apply and validate novel tools for drug discovery in order to improve the actual landscape of available methods in Structure Based Drug Design.

Specific Objectives

In particular, the specific objectives are the following:

1. Validation of rDock by comparing it to other reference docking programs.
2. Improve docking performance by introducing knowledge-based scoring biases.
3. Docking-based Virtual Screening and post-filtering of hits with complementary methods.
4. Develop a novel approach based on Steered Molecular Dynamics and establish a proof-of-concept in prospective and retrospective applications.

Chapter 3

Papers

Three papers have been published as a result of this thesis. Each of them, preceded by an overview and a brief summary of the results and conclusions, are included in this chapter.

The first paper, entitled "rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids", introduces rDock to the scientific community. rDock is a molecular docking software released as open source that improves results obtained by other commonly used docking programs (more details on the results in section 4.1).

In the second paper, entitled "Dynamic Undocking and the Quasi-Bound State as Tools for Drug Design", we present the "Dynamic Undocking", a new tool for drug discovery that can help improve virtual screening results by several fold. All the details on the methodology development and experimental validation using Hsp90 are found in section 4.2.

The third paper, entitled "Docking-Undocking Combination Applied to the D3R Grand Challenge 2015", summarizes our participation in the public challenge organized by the Drug Design Data Resource (D3R) (in section 4.3). Our approach was based on a combination of docking with rDock and Dynamic Undocking, which placed us amongst the best participants.

rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids

Sergio Ruiz-Carmona[1,2], Daniel Alvarez-Garcia[1,2], Nicolas Foloppe[3], A. Beatriz Garmendia-Doval[4], Szilveszter Duhos[5], Peter Schmidtke[6], Xavier Barril[1,2,7], Roderick E. Hubbard[3,8] and S. David Morley[9,10]

PLoS Computational Biology, vol. 10, p. e1003571, Apr. 2014.

1 Departament de Fisicoquimica, Facultat de Farmacia, Universitat de Barcelona, Barcelona, Spain, **2** Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain, **3** Vernalis (R&D) Ltd, Granta Park, Cambridge, United Kingdom, **4** Amper Programas, Madrid, Spain, **5** Omixon Biocomputing, Budapest, Hungary, **6** Discngine, Paris, France, **7** Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain, **8** YSBL, University of York, Heslington, York, United Kingdom, **9** Enspiral Discovery Limited, Cambridge, United Kingdom, **10** Ariana Pharma, Paris, France

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids

Sergio Ruiz-Carmona[1,2], Daniel Alvarez-Garcia[1,2], Nicolas Foloppe[3], A. Beatriz Garmendia-Doval[4], Szilveszter Juhos[5], Peter Schmidtke[6], Xavier Barril[1,2,7]*, Roderick E. Hubbard[3,8]*, S. David Morley[9,10]*

1 Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain, **2** Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain, **3** Vernalis (R&D) Ltd, Granta Park, Cambridge, United Kingdom, **4** Amper Programas, Madrid, Spain, **5** Omixon Biocomputing, Budapest, Hungary, **6** Discngine, Paris, France, **7** Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain, **8** YSBL, University of York, Heslington, York, United Kingdom, **9** Enspiral Discovery Limited, Cambridge, United Kingdom, **10** Ariana Pharma, Paris, France

Abstract

Identification of chemical compounds with specific biological activities is an important step in both chemical biology and drug discovery. When the structure of the intended target is available, one approach is to use molecular docking programs to assess the chemical complementarity of small molecules with the target; such calculations provide a qualitative measure of affinity that can be used in virtual screening (VS) to rank order a list of compounds according to their potential to be active. rDock is a molecular docking program developed at Vernalis for high-throughput VS (HTVS) applications. Evolved from RiboDock, the program can be used against proteins and nucleic acids, is designed to be computationally very efficient and allows the user to incorporate additional constraints and information as a bias to guide docking. This article provides an overview of the program structure and features and compares rDock to two reference programs, AutoDock Vina (open source) and Schrödinger's Glide (commercial). In terms of computational speed for VS, rDock is faster than Vina and comparable to Glide. For binding mode prediction, rDock and Vina are superior to Glide. The VS performance of rDock is significantly better than Vina, but inferior to Glide for most systems unless pharmacophore constraints are used; in that case rDock and Glide are of equal performance. The program is released under the Lesser General Public License and is freely available for download, together with the manuals, example files and the complete test sets, at http://rdock.sourceforge.net/

Citation: Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, et al. (2014) rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Comput Biol 10(4): e1003571. doi:10.1371/journal.pcbi.1003571

Editor: Andreas Prlic, UCSD, United States of America

Received November 29, 2013; **Accepted** February 19, 2014; **Published** April 10, 2014

Funding: The rDock program was developed from 1998–2006 by the software team at RiboTargets (subsequently Vernalis (R&D) Ltd). Maintenance, benchmarkings and subsequent developments have been funded by the Wellcome Trust (C2D2 initiative; York U.) and the Spanish Ministerio de Economía y Competitividad (SAF2009-08811& SAF2012-33481; Barcelona U.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: I have read the journal's policy and have the following conflicts: A. Beatriz Garmendia-Doval is a paid employee of Amper Programas. Szilveszter Juhos is a paid employee of Oximon Biocomputing. Peter Schmidtke is a paid employee of Discngine. Roderick E. Hubbard and Nicolas Foloppe are paid employees of Vernalis Ltd. S. David Morley is a paid employee of Enspiral Discovery Limited and Ariana Pharma. This does not alter our adherence to the PLOS policies on sharing data and materials. All other authors have declared that no competing interests exist.

* E-mail: xbarril@ub.edu (XB); roderick.hubbard@york.ac.uk (REB); d.morley@enspiral-discovery.com (SDM)

This is a *PLOS Computational Biology* Software Article.

Introduction

The discovery of small molecules with biological activities is important to probe biological mechanism in chemical biology and to provide drug candidates as potential therapeutic agents. The first step in this process is to identify compounds that bind to a specific target (hits); experimentally this is usually achieved with high-throughput (HTS) or fragment screening (FS). The resulting hits are then optimised to higher affinity compounds, usually guided by a model of how the compounds bind to the target, increasingly with crystal structures of the target used to guide the optimisation.

Computational methods are often used as a central part of this process. Molecular docking can play an important role in the optimisation, where a proposed position and conformation (so-called pose) of the compound can be generated and provide useful models for how the compounds are binding, in advance of any experimental structure determination. However, if the structure of the target is known and a druggable cavity has been identified [1], molecular docking can also be used to screen virtual chemical collections to identify those molecules that offer good shape and chemical complementarity [2]. Such virtual screening (VS) offers opportunities for small research groups without access to HTS or FS to identify new hit compounds, as setting up a low-throughput assay to test a few tens of compounds is relatively fast and inexpensive. Such VS has been successful, but it requires a docking program that is computationally efficient and can be finely tuned to achieve optimal performance [3–5]. rDock is a molecular docking platform which has been optimised for such tasks.

rDock has its origins in the program RiboDock [6], designed initially for VS of RNA targets. Developed at the company now known as Vernalis (http://www.vernalis.com), the software,

scoring functions, and search protocols have been refined continuously over a number of years to meet the demands of in-house discovery projects on heat-shock proteins [7–9], kinases [10–13] and other targets. The major components of the platform now include fast intermolecular scoring functions (vdW, polar, desolvation) validated against protein and RNA targets, a Genetic Algorithm (GA)-based stochastic search engine, a wide variety of external restraint terms (tethered template, pharmacophoric restraints), and novel Genetic Programming-based post-docking filtering [14]. In this paper we describe the platform, benchmark it against two other state of the art docking programs for both binding mode prediction and VS and discuss its use in high-throughput VS (HTVS).

Design and implementation

The rDock platform is a collection of command-line programs and scripts (Table 1 and Figure S1). The main tasks are carried out by the programs *rbcavity* (cavity generation) and *rbdock*(docking). rDock is written in ANSI C++ and compiles under the Linux operating system using the GNU g++ compiler. Apart from the C++ Standard Template Library (STL) there are minimal external dependencies (*e.g.* OpenBabel bindings for running *sdtether* and *sdrmsd* [15]). The core functionality is compiled into a single shared library, which is linked with each of the (light-weight) command-line applications. Scoring functions and docking protocols are assembled at run-time from well-defined C++ object class hierarchies, allowing for customisation at source code level by extending the base classes. Ancillary scripts are provided for file management and output processing and are described in the manuals.

Preparation

The receptor is provided in Tripos MOL2 format with standard atom typing. Amino acid ionisation states in the vicinity of the cavity must be defined, as the rDock scoring functions depend on formal charge assignments. Metal ions, cofactors and structural water molecules can be included as part of the receptor. The user should also resolve other structural issues such as alternate locations or missing atoms. The docking volume is defined by the *rbcavity* program which provides two mapping algorithms; the accessible volume within a specific distance of a reference ligand, and a two probe sphere method [6]. In the examples presented in this paper, the reference ligand method is used with a distance of 6 Å.

Ligands to be docked are read in the MDL SDFile format (SDF) and should have the correct topology and bond orders. The program can protonate and deprotonate certain ionisable groups, but pre-processing the ligands with a dedicated program is preferable. Since the program only samples exocyclic dihedral angles, a correct input geometry is required for bonds, angles and rings. In the case of flexible rings, a variety of low-energy conformers should be pregenerated by a suitable program. We have used LigPrep [16] for all ligand preparation steps. The execution of the programs is controlled by a series of parameter (.prm) files; this allows user controlled tuning of the docking protocol and scoring functions (described in more detail in the Manual). The following sections describe the main characteristics of the program and the available options.

Scoring

The rDock master scoring function (S^{total}) is a weighted sum of intermolecular (S^{inter}), ligand intramolecular (S^{intra}), site intramolecular (S^{site}), and external restraint terms ($S^{restraint}$). S^{inter} is the main term of interest as it represents the protein-ligand (or RNA-ligand) interaction score. S^{intra} reports the change in energy of the ligand relative to the input ligand conformation. Similarly, S^{site} represents the relative energy of the flexible regions of the active site. In the current implementation, the only flexible bonds in the active site are terminal OH and NH_3^+ bonds. $S^{restraint}$ is a collection of non-physical restraint functions that can be used to bias the docking calculation in several useful ways (*vide infra*). S^{inter}, S^{intra}, and S^{site} are built from a common set of constituent potentials, which are described in the Manual. Briefly, they mainly consist of a van der Waals potential (vdW), an empirical term for attractive and repulsive polar interactions, and an optional desolvation potential that combines a weighted solvent accessible surface area approach [17] with a rapid probabilistic approximation to the calculation of solvent accessible surface areas [18] for computational efficiency. The vdW term can be calculated during docking, or precalculated and stored on grid files by the ancillary

Table 1. List of main programs and utilities included in the rDock package.

Name	Language	Use	Description
rbdock	C++	Docking	The main rDock docking engine
rbcavity	C++	Cavity definition	Cavity mapping and preparation of docking site (.as file).
rbcalcgrid	C++	Preparation	Calculation of vdW grid files (usually called by make_grid.csh wrapper script)
sdtether	python	Preparation	Prepares a ligand SD file for tethered scaffold docking, annotating the atom indices of the tethered substructure. Requires OpenBabel python bindings [15]
sdrmsd	python	Analysis	Calculation of ligand Root Mean Squared Displacement (RMSD) between reference and docked poses, taking into account ligand topological symmetry. Requires OpenBabel python bindings [15]
sdfilter	perl	Analysis	Utility for filtering SD files by arbitrary data field expressions. Useful for simple post-docking filtering by score components.
sdsort	perl	Analysis	Utility for sorting SD files by arbitrary data field. Useful for simple post-docking filtering by score components.
sdreport	perl	Analysis	Utility for reporting SD file data field values in tab-delimited or CSV format.

doi:10.1371/journal.pcbi.1003571.t001

program *rbcalcgrid*; this increases computational performance. Two distinct scoring functions have been optimized using a binding affinity validation set (described in the Manual). The default scoring function (SF3) uses the repulsive polar term but not the desolvation term, while the solvation scoring function (SF5) does the opposite. The default SF3 is slightly faster and works better for proteins while the solvation term is generally better for nucleic acids. More importantly, the weighting terms of the scoring function can be re-optimized with larger or more focused validation sets to improve its performance.

Sampling

rDock uses a combination of stochastic and deterministic search techniques to generate low energy ligand poses. The standard docking protocol to generate a single ligand pose uses 3 stages of Genetic Algorithm search (GA1, GA2, GA3), followed by low temperature Monte Carlo (MC) and Simplex minimization (MIN) stages. The GA stages are interdependent and are designed to be used sequentially. Several scoring function parameters are varied between the stages to promote efficient sampling of the starting poses, whilst minimising the likelihood that the poses become trapped early in the search. The variations are in the functional form of the S^{inter} vdW potential (switched from 4–8 potential in GA1/GA2 to 6–12 potential in GA3/MC/MIN), the tolerances on the polar distance and angular functions (relaxed in GA1 and progressively tightened in GA2/GA3/MC), and the weight of the ligand dihedral potential (reduced in GA1 and progressively increased in GA2/GA3/MC). All scoring function parameters are at their final reported values for the final MC/MIN stages. The GA chromosome consists of the ligand centre of mass (COM), the ligand orientation, as represented by the Euler angles (heading, attitude, bank) required to rotate the ligand principal axes from the Cartesian reference axes, the ligand rotatable dihedral angles, and the receptor rotatable dihedral angles. The initial population is generated such that the ligand COM lies on a randomly selected grid point within the defined docking volume, and the ligand orientation and all dihedral angles are randomised. Mutations are applied to a randomly selected degree of freedom and the magnitude of the mutation is selected from rectangular distributions of defined width. A generation is considered to have passed when the number of new individuals created is equal to the population size. Instead of having a fixed number of generations, the GA is allowed to continue until the population converges (scoring improvement <0.1 units over the last three generations). This allows early termination of poorly performing runs for which the initial population is not able to generate a good solution. Once the GA converges, a low temperature Monte Carlo simulation is used to refine the pose, followed by Simplex routine to generate a minimised solution. A more detailed description of the sampling protocol can be found in the Manual. In a typical docking calculation, the whole process is repeated 10 to 100 times and the overall lowest scoring pose is taken as the correct solution (see below for discussion on convergence), but it is also possible to access the minimisation stage directly or simply score a pre-docked pose.

Biased docking

The main limitation in molecular docking is the quality of the scoring functions. It is therefore usual to introduce empirical bias, which can improve the quality of the results and also reduce the search space, thus improving performance. rDock implements several pseudo-energy scoring functions that are added to the total scoring function under optimisation, and a restricted search protocol.

Pharmacophoric restraints. This feature ensures that pharmacophores (derived from known ligands or hot-spot mapping methods) are satisfied by all generated poses. rDock recognizes nine feature types: neutral hydrogen bond acceptor, neutral hydrogen bond donor, hydrophobic, hydrophobic aliphatic, hydrophobic aromatic, negatively charged, positively charged, and any heavy atom. Each pharmacophore restraint is defined by a combination of feature type and position, specified as a tolerance sphere with coordinate (x,y,z), and radius (r). Restraints are classified as either mandatory or optional, where the user can specify how many optional restraints (N_{opt}) should be met. Ligands that have insufficient quantities of the defined restraint feature types are removed prior to docking. The penalty score for a single pharmacophore restraint is proportional to the square of the distance from the nearest ligand feature of the required type to the surface of the tolerance sphere, and is zero when the nearest ligand feature is within the tolerance sphere. The total pharmacophore restraint score, S^{ph4}, is the sum of all the mandatory restraints plus the N_{opt} lowest scoring optional restraints.

Tethered template. Tethered template docking can be used to enforce partial binding modes obtained from crystal structures of related molecules or constituent fragments. The template is defined by a reference bound ligand structure and a SMARTS query string defining the substructure to be tethered. The *sdtether* utility prealigns molecules with matching substructures with the reference substructure coordinates prior to docking. Non-matching molecules are rejected. Molecules that have more than one substructure match with the query are replicated within the library of compounds to be docked, and each replicate prealigned and docked individually, thus ensuring that all possible substructure alignments are examined. In this mode, the centre of mass and principal axes of the tethered substructure, rather than the whole molecule, define the ligand position and orientation. Dihedral angle mutations operate exclusively on the free (untethered) end of each ligand rotatable bond, ensuring the tethered substructure coordinates remain unchanged. Some movement of the tethered region is allowed up to user-defined maximum deviations from the reference coordinates for ligand translation (typically 0.1 Å) and ligand rotation (typically 1°). For greater sampling efficiency, tethering in rDock is enforced absolutely during pose generation by restricting the randomisation and mutation functions for the tethered degrees of freedom, rather than through the use of an external penalty function.

Other. 1) To ensure that all poses are contained wholly within the defined docking volume, a cavity penalty function (S^{cavity}) is calculated over all non-hydrogen ligand atoms. If the atom is within the docking volume this term is zero, else, it is proportional to the square of the distance to the nearest docking volume grid point.2) When experimental NMR distance limits (NOE or STD) are known for a specific ligand, restraints can be used to ensure that a minimum distance is fulfilled between an atom (or group of atoms) of the ligand and an atom (or group of atoms) of the receptor.

Results

Benchmarking

The performance of rDock was compared with that of Glide (version 57111 [19]) and AutoDock Vina [20] for database enrichment and binding mode prediction for various test sets. As detailed in Supporting Information Text S1, all receptors, docking cavities and ligands were prepared in the same manner and running parameters modified to ensure exhaustive sampling by all programs.

Table 2. Percentage of top-ranked poses with an RMSD below 2 Å.

	% Correct (top 1)	% Correct (all)
rDock	76±3[1]	99±0.2[1]
Glide	67.6	83.8
Vina	81.2±2[1]	97±0.5[1]

[1]Average and standard deviation taking 100 random sets of 100 docking poses out of a pool of 1000 solutions.
doi:10.1371/journal.pcbi.1003571.t002

Protein-ligand binding mode predictions. The CCDC-Astex Diverse Set of 85 diverse protein-ligand complexes was selected for comparing binding mode prediction [21]. The results, represented by percentage of correct predictions (ligand RMSD below 2 Å) can be seen in Table 2. rDock calculations converge after 20–50 GA runs (Figure S2; convergence also discussed below). The predicted binding mode is correct in approximately 80% of cases for rDock and Vina, while Glide's performance is close to 70%. Failures for rDock and Vina are due to scoring errors, as a correct pose is nearly always generated (99% and 97% of times, respectively). However, Glide fails to sample the correct binding mode in 16% of cases. Figure S3 shows the docking outcome for each system and program. Although no obvious trend can be identified, it would seem that rDock and Vina have a higher coincidence in the type of systems for which they succeed or fail.

RNA-ligand binding mode predictions. We selected 56 RNA-ligand complexes from the original RiboDock [6] and DOCK6 [22] sets to assess the performance of rDock with RNA as the receptor. RNA structures are more challenging than proteins (less closed cavities, less hydrophobic, featureless) and the ligands themselves are larger and more flexible (7.7±4.3 rotatable bonds vs. 5.1±3.1 for the Astex set). For this reason the success cut-off criterion is an RMSD below 2.5 Å, relative to the crystal structure. The scoring function SF5, which includes a solvation term, is better for RNA than SF3, as independently assessed [23]. After 50 GA runs, the top-ranked docking solution is correct in 54±3% of the systems (Figure S4), and at least one correct pose is generated in 98% of cases, confirming that as with proteins, errors are attributable to scoring rather than sampling problems. However, both SF3 and SF5 have been primarily optimized for proteins suggesting that development of an RNA-specific scoring function could result in improvements. Vina and Glide can work with but have not been optimised for ligand docking to RNA. On the same set of complexes, we obtain success rates of 29±2 for Vina and 17.8 for Glide.

Virtual screening (DUD). VS enrichment was assessed using the DUD benchmark set [24] which consists of 39 protein-ligand complexes with crystal structure, with an average of about 100 known active ligands per complex and 36 decoys per active ligand. The decoys are physically similar but topologically dissimilar to the ligands in order to avoid bias. The DUD-E benchmark set [25] was published recently, adding more protein-ligand complexes. For our test set, 20 of the original DUD sets were substituted with DUD-E data with more ligands and decoys per system. Figures S5 and S6 show the ROC curves for all systems and the most relevant parameters are summarized in Table S1. The results are summarised in Table 3. Using most metrics, Glide outperforms the other programs in ~70% of the systems, while rDock is better in ~20% of systems and Vina in the remaining 10%. On average, rDock AUC is 11% lower than Glide and 5% better than Vina. In terms of logAUC, on average, Glide outperforms rDock by 30%, while rDock outperforms Vina by 8%.

Sampling exhaustiveness and computing performance

A distinctive feature of rDock is that the GA converges very quickly. This behaviour was designed for VS, where it is important to discard poor ligands early on. Multiple docking runs (which includes GA optimisation followed by MC and Simplex minimisation) are necessary to reach the global minimum score (S_{min}), but few docking runs are necessary to reach a similar score (Figure 1). For instance, after 5 runs, approximately 80% of ligands reach a score of $0.8*S_{min}$, and the median value is $0.94*S_{min}$. Convergence depends on the dimensionality of the problem and fewer docking runs are necessary when the ligands contain fewer rotatable bonds (Figure 1) or when the cavity has a smaller size (Figure S7). System-specific multi-step HTVS

Table 3. Average values of different VS performance metrics over the 39 DUD/DUD-E systems.

Program	AUC[1]	logAUC[2]	EFmax[3]	EF 1%[4]	EF 20%[4]
rDock	0.69	0.26	98.7	11.4	2.5
	(18%)	(18%)	(33%)	(19%)	(18%)
Glide	0.78	0.37	334.6	22.6	3.2
	(69%)	(72%)	(41%)	(69%)	(72%)
Vina	0.66	0.24	124.3	8.9	2.2
	(13%)	(10%)	(26%)	(11%)	(10%)

The values in parentheses indicate the percentage of systems for which the program provides the optimal performance on a given metric.
[1]Area Under the ROC Curve.
[2]Area Under the semilogarithmic ROC Curve.
[3]Maximal Enrichment Factor.
[4]Enrichment Factor when the top x% of the virtual collection is selected.
doi:10.1371/journal.pcbi.1003571.t003

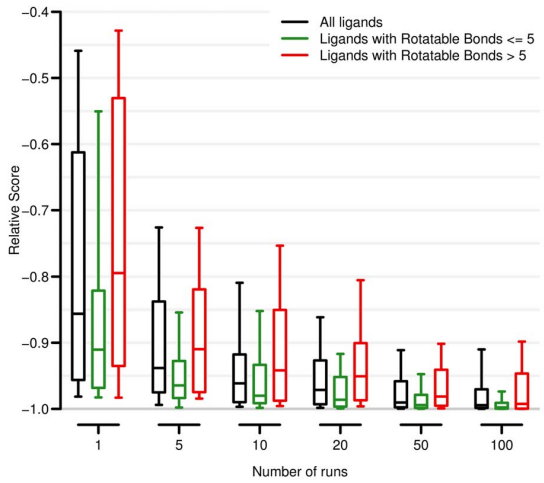


Figure 1. Relative score vs. the number of docking runs for all the protein-ligand complexes in the CCDC-Astex set. The boxplot indicates the median value (out of 1000 possible solutions) and the first and last quartile, while the whiskers span the 10% to 90% range. The whole set (black) has been sub-divided into ligands with 5 or fewer rotatable bonds (green) and the rest (red).
doi:10.1371/journal.pcbi.1003571.g001

protocols (see section below and Manual) achieve optimal performance with an average of 8–10 runs per ligand. Table 4 shows the average computing times per ligand on 4 DUD systems [24]. Precalculating the van der Waals potentials on a grid saves 20% to 40% of docking computing time, depending on the system. For exhaustive docking, rDock is approximately 5-fold faster than Vina, but still 8-fold slower than Glide SP. HTVS protocols achieve a further reduction of 5 to 8-fold in computing time, bringing the performance of rDock to be very similar to Glide SP with no negative impact on the results (Table S3). Using a relatively modest 100-core computing facility, a VS campaign of 1 million compounds can be completed in less than 1 day and the 21 million commercially accessible compounds compiled in ZINC database [26] could be screened in 10 to 20 days for most systems.

Considerations for real VS applications

Design of multi-step HTVS protocols. Different docking protocols are required for different applications. For detailed docking, where the user is interested primarily in high accuracy, a suggested rDock protocol is to allow receptor flexibility, bypass the pre-calculation of van der Waals potentials and perform exhaustive sampling (50–100 GA runs). For HTVS applications, where computing performance is important, the recommended rDock protocol is to limit the search space (i.e. rigid receptor), apply the grid-based scoring function and to use a multi-step protocol to stop sampling of poor scorers as soon as possible. An example is for the DUD system COMT, where the computational time can be reduced by 7.5-fold without affecting performance by: 1) 5 GA runs for all ligands; 2) ligands achieving a score of −20 or lower run 10 further GAs; 3) for those ligands achieving a score of −25 or lower, continue until 50 GAs. The optimal protocol is specific for each particular system and parameter-set, but can be identified with a purpose-built script (see Manual).

Guided docking. Usually, VS applications exploit existing information to optimize the cavity definition (e.g. choice of protein conformation, displaceable water molecules) and to bias the docking protocol with empirical restraints (e.g. pharmacophoric points, shape similarity). This is an essential step common to all successful docking-based VS undertakings [3,27]. For this reason, we have compared the outcome of VS on Hsp90, a DUD system for which we have developed and used optimal docking protocols [7,8,28]. The cavity includes 2 interstitial water molecules and two pharmacophoric points. As shown in Table 5 and Figures S8 and S9, all VS performance metrics improve significantly, particularly those related to early enrichment (logAUC, EF1%). As scoring functions are supplemented with empirical information, performance increases and the difference between programs reduce (Table S2).

Availability and future directions

The program is released under the Lesser General Public License and the source code, scripts, manuals, and test sets are available at http://rdock.sourceforge.net/. The current version has prototype code to sample fully the degrees of freedom and occupancy of interstitial water molecules, as previously described for GOLD [29], or to dock simultaneously to an ensemble of receptor coordinates to simulate receptor flexibility in an efficient way. These features require further validation. Future develop-

Table 4. Average computing times (in seconds per ligand) on 4 DUD systems.

	Vina[1]	Glide SP[1]	rDock			
			Grid-based SF		Indexed SF	
			VS[2]	Full[1,3]	VS[2]	Full[1,3]
ADA	86.4	4.2	4.2	27.0	5.4	33.0
COMT	77.4	3.0	3.0	22.5	5.0	31.8
PARP	54.0	1.5	3.9	16.5	5.7	29.1
Trypsin	372.0	6.0	14.1	53.1	20.1	82.5
Average	**147.5**	**3.7**	**6.3**	**29.8**	**9.1**	**44.1**

[1]Default program parameters were used.
[2]On HTVS mode, the average number of docking runs needed for these 4 systems is 10.
[3]50 docking runs are used for default docking.
All figures were obtained on Intel Xeon X5660 CPUs at 2.80 GHz.
doi:10.1371/journal.pcbi.1003571.t004

Table 5. VS performance metrics for Hsp90 using an unbiased protocol with default parameters (rDock, Glide & Vina) or an optimized cavity definition and empirical pharmacophoric restraints (rDock-guided & Glide-guided).

Program	AUC	logAUC	EFmax	EF 1%	EF 20%
rDock	0.63	0.20	3.9	0.0	1.5
	(0.8)	*(0.7)*	*(0.5)*	*(1.0)*	*(0.7)*
Glide	0.77	0.28	7.4	0.0	2.1
	(1.0)	*(1.0)*	*(1.0)*	*(1.0)*	*(1.0)*
Vina	0.55	0.16	1.4	0.0	0.75
	(0.7)	*(0.6)*	*(0.2)*	*(1.0)*	*(0.4)*
rDock-guided	0.92	0.46	36.9	12.3	4.3
	(1.2)	*(1.6)*	*(5.0)*	*(–)*	*(2.0)*
Glide-guided	0.90	0.46	17.4	6.9	4.6
	(1.2)	*(1.6)*	*(2.3)*	*(–)*	*(2.2)*

Note that Vina does not support pharmacophoric restraints. The numbers in parentheses indicate performance relative to the best non-guided result (Glide).
doi:10.1371/journal.pcbi.1003571.t005

ments will aim at improving the scoring functions for both protein-ligand and RNA-ligand interactions.

Supporting Information

Figure S1 Workflow summary of an rDock docking job. Shapes in gray background are not covered with any rDock program and must be carried out with independent software.
(TIF)

Figure S2 Binding mode prediction in the protein-ligand set (CCDC-Astex): Percentage of top-ranked poses with RMSD below 2.0 Å as a function of the number of docking runs. The boxplot indicates the median value (out of 100 possible solutions) and the first and last quartile, while the whiskers span the 10% to 90% range. The whole set (black) has been sub-divided into ligands with 5 or fewer rotatable bonds (green) and the rest (red).
(TIF)

Figure S3 Matrix representation of the docking outcome for each system in the CCDC-Astex set for the three programs evaluated. A black area indicates that the best-scoring pose for a particular system-program combination has an RMSD below 2.0 Å.
(TIF)

Figure S4 Binding mode prediction in the RNA-ligand set: Percentage of top-ranked poses with RMSD below 2.5 Å as a function of the number of GA runs. The boxplot indicates the median value (out of 100 possible solutions) and the first and last quartile, while the whiskers span the 10% to 90% range.
(TIF)

Figure S5 Receiver Operating Characteristic (ROC) Curves of all DUD systems. In the Y-axis, the true positive rate is the fraction of true positives out of the total actual positives and, in the X-axis, the false positive rate is the fraction of false positives out of the total actual negatives. In gray, ROC curve in case of random results.
(TIF)

Figure S6 Semilogarithmic Receiver Operating Characteristic (ROC) Curves of all DUD systems. In the Y-axis, the true positive rate is the fraction of true positives out of the total actual positives and, in the X-axis in logarithmic scale, the false positive rate is the fraction of false positives out of the total actual negatives. In gray, semilogarithmic ROC curve in case of random results.
(TIF)

Figure S7 Relative score vs. the number of docking runs for all the protein-ligand complexes in the CCDC-Astex set. The boxplot indicates the median value (out of 100 possible solutions) and the first and last quartile, while the whiskers span the 10% to 90% range. The whole set (black) has been sub-divided into systems with relatively small cavities (green) and the rest (red).
(TIF)

Figure S8 ROC curve of HSP90 without pharmacophoric restraints in normal (A) or semilogarithmic scale (B).
(TIF)

Figure S9 ROC curve of HSP90 with pharmacophoric restraints in normal (A) or semilogarithmic scale (B). It should be noted that using these settings, Glide only produces an output for 13 actives (out of 24) and 451 decoys (out of 864).
(TIF)

Software S1 Compressed file with the source code of the rDock software for ligand docking to Proteins and Nucleic Acids.
(GZ)

Table S1 Summary of statistics for all DUD systems and averages for each and all programs.
(DOCX)

Table S2 Spearman's rank correlation coefficient (ρ) between programs on the Hsp90 DUD set.
(DOCX)

Table S3 AUC for the 4 DUD systems used for calculating the time performance.
(DOCX)

Text S1 Supporting Methods: Test set preparation, execution and analysis.
(DOCX)

Text S2 Full Acknowledgements.
(DOCX)

Acknowledgments

We thank the users at Vernalis (and RiboTargets) who drove the development of the program and helped with validation as well as students who have helped to maintain and assess the program at York (see full acknowledgements in Supporting Information Text S2).

Author Contributions

Conceived and designed the experiments: XB REH SDM. Performed the experiments: SRC DAG ABGD SJ PS XB SDM. Analyzed the data: SRC DAG ABGD SJ PS XB REH SDM. Wrote the paper: XB REH SDM. Programming C++: ABGD SJ SDM. Programming Perl: SDM. Programming Python: DAG. Code Maintenance and sourceforge site maintenance: SRC DAG PS. Contributed to the initial drafting and assessment of the performance of the program: NF REH SDM.

References

1. Barril X. (2012) Druggability predictions: Methods, limitations, and applications. Wiley Interdisciplinary Reviews: Computational Molecular Science. Doi: 10.1002/wcms.1134.
2. Brooijmans N, Kuntz ID. (2003) Molecular recognition and docking algorithms. Annu Rev BiophysBiomolStruct 32: 335–373.
3. Barril X, Hubbard RE, Morley SD. (2004) Virtual screening in structure-based drug discovery. Mini Rev Med Chem 4: 779–791.
4. Jorgensen WL. (2004) The many roles of computation in drug discovery. Science 303: 1813–1818.
5. Shoichet BK. (2004) Virtual screening of chemical libraries. Nature 432: 862–865.
6. Morley SD, Afshar M. (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using ribodock. J Comput Aided Mol Des 18: 189–208.
7. Barril X, Brough P, Drysdale M, Hubbard RE, Massey A, et al. (2005) Structure-based discovery of a new class of Hsp90 inhibitors. Bioorg Med ChemLett 15: 5187–5191.
8. Brough PA, Barril X, Borgognoni J, Chene P, Davies NG, et al. (2009) Combining hit identification strategies: Fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. J Med Chem 52: 4794–4809.
9. Williamson DS, Borgognoni J, Clay A, Daniels Z, Dokurno P, et al. (2009) Novel adenosine-derived inhibitors of 70 kDa heat shock protein, discovered through structure-based design. J Med Chem 52: 1510–1513.
10. Foloppe N, Fisher LM, Howes R, Kierstan P, Potter A, et al. (2005) Structure-based design of novel Chk1 inhibitors: Insights into hydrogen bonding and protein-ligand affinity. J Med Chem 48: 4332–4345.
11. Foloppe N, Fisher LM, Howes R, Potter A, Robertson AG, et al. (2006) Identification of chemically diverse Chk1 inhibitors by receptor-based virtual screening. Bioorg Med Chem 14: 4792–4802.
12. Richardson CM, Williamson DS, Parratt MJ, Borgognoni J, Cansfield AD, et al. (2006) Triazolo[1,5-a]pyrimidines as novel CDK2 inhibitors: Protein structure-guided design and SAR. Bioorg Med ChemLett 16: 1353–1357.
13. Richardson CM, Nunns CL, Williamson DS, Parratt MJ, Dokurno P, et al. (2007) Discovery of a potent CDK2 inhibitor with a novel binding mode, using virtual screening and initial, structure-guided lead scoping. Bioorg Med ChemLett 17: 3880–3885.
14. Garmendia-Doval AB, Morley SD, Juhos S. (2004) Post docking filtering using cartesian genetic programming. In: Anonymous Artificial Evolution. Volume 2936 of Lecture Notes in Computer Science. : Springer. pp. 189–200.
15. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, et al. (2011) Open babel: An open chemical toolbox. J Cheminform 3: 33-2946-3-33.
16. Schrödinger L. (2011) Suite 2011: LigPrep, version 2.5.
17. Wang J, Wang W, Huo S, Lee M, Kollman PA. (2001) Solvation model based on weighted solvent accessible surface area. The Journal of Physical Chemistry B 105: 5055–5067.
18. Hasel W, Hendrickson TF, Still WC. (1988) A rapid approximation to the solvent accessible surface areas of atoms. Tetrahedron Computer Methodology 1: 103–116.
19. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, et al. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. J Med Chem 47: 1739–1749.
20. Trott O, Olson AJ. (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31: 455–461.
21. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WT, et al. (2007) Diverse, high-quality test set for the validation of protein-ligand docking performance. J Med Chem 50: 726–741.
22. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, et al. (2009) DOCK 6: Combining techniques to model RNA-small molecule complexes. Rna15: 1219–1230.
23. Chen L, Calin GA, Zhang S. (2012) Novel insights of structure-based modeling for RNA-targeted drug discovery. J ChemInf Model 52: 2741–2753.
24. Huang N, Shoichet BK, Irwin JJ. (2006) Benchmarking sets for molecular docking. J Med Chem 49: 6789–6801.
25. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. (2012) Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. J Med Chem 55: 6582–6594.
26. Irwin JJ, Shoichet BK. (2005) ZINC–a free database of commercially available compounds for virtual screening. J ChemInf Model 45: 177–182.
27. Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH. (2012) Structure-based virtual screening for drug discovery: A problem-centric review. Aaps j 14: 133–141.
28. Barril X, Morley SD. (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J Med Chem 48: 4432–4443.
29. Verdonk ML, Chessari G, Cole JC, Hartshorn MJ, Murray CW, et al. (2005) Modeling water molecules in protein-ligand docking using GOLD. J Med Chem 48: 6504–6515.

SUPPLEMENTARY MATERIAL

rDock: a fast, versatile and open source docking program for proteins and nucleic acids.

Text S1. Supporting Methods: Test set preparation, execution and analysis.

DUD and ASTEX sets:

- Protein Preparation

The receptor structure files in DUD and Astex sets were processed using Preparation Wizard tool from Maestro (from Schrödinger), and were then used as input for the three programs. To define the cavity, rDock was run using the crystallographic ligand provided as reference with the "reference ligand method" and the following parameter values (if not in the list, the default value was considered): radius=6.0, small_sphere=1.0 and max_cavities=1. The coordinates obtained for the center and the size of the binding site were applied for Glide and Vina to ensure the least dissimilar cavities between each program.

- Ligand Preparation

The structure of the ligands in DUD set was converted to smiles format and processed with LigPrep software (from Schrödinger) applying the following filters: maximum atoms=100, maximum stereoisomers=8, maximum tautomers=6 and ionizing at pH=7 with a tolerance of +-1.

The results in sdf format, compatible for running rDock, were converted to mae and pdbqt formats for running Glide and Vina, respectively.

In case of Astex set, the ligands had already been manually prepared, thus no need of LigPrep processing was needed. Hence, the process was the same as for DUD set after the ligands had been processed with LigPrep.

- Docking

The Molecular Docking process was defined to be the most similar as possible. The exhaustiveness of all programs was set higher than default to try to obtain less sources of error than usual (sample minimum?).

For DUD set, rDock was run with a receptor flexibility=3, scoring function "dock.prm" and 100 docking runs. Glide was run with expanded sampling and with the following options increased with respect to the default values to avoid filtering of intermediate poses and bad scored ligands which facilitated analysis of the results: postdock_npose=5000, poses_per_lig=5000 and nreport=(5000*number of ligands). Vina had all parameters as default but the following

ones, for the same reason as Glide: exhaustiveness=16, num_modes=100 and energy_range=30.
For Astex set, all the parameters were the same as in DUD but the number of runs in rDock and Vina, which were set to 1000 and to 50 jobs per ligand, respectively.

- Results analysis
In case of DUD set, ROC curves were generated using ROCR package for R (ref) and several statistical values, such as AUC and Enrichment Factors, were calculated.
In case of Astex set, the RMSD of each predicted binding mode with respect to the crystallized ligand was calculated using Open Babel toolkit (ref). Random sets of 100 ligands were selected from all the resulting binding modes (if more than 100 ligands were available) and the percentage of the top-scored binding mode with an RMSD below 2Å was calculated.

RNA:

The structure of the RNA-ligand complex was downloaded from the PDB and prepared using MOE (ref chemcomp). The cavity was defined using the crystallographic ligand in the PDB as reference with the “reference ligand method” from rDock and the following parameters different from default: radius=4.0, small_sphere=1.0 and max_cavities=1.
The docking jobs were run with receptor flexibility=3, scoring function “dock_solv.prm” and 1000 docking runs, for statistical purposes in analysis of results.
Like in Astex set, the RMSD of each predicted binding mode with respect to the crystallized ligand was calculated and random sets of 100 ligands were selected for calculating the percentage of top-scored binding modes with an RMSD below 2Å.

Pharmacophoric restraints:

Based on the knowledge available on the DUD systems and on their pharmacophoric properties, we selected HSP90. Three structural waters were added near to ASP78, the volume around residues TRP147 and GLY93 was excluded and hydrogen-bonds between ASP78 and the ligand and between one of the structural waters added and the ligand were added as pharmacophoric restraints.
rDock and Glide were run with the same parameters as in the same DUD system without any pharmacophoric restraint (Vina cannot use pharmacophoric restraints).
The results were processed the same way as in DUD set.

Text S2: Full Acknowledgements

The development of the initial RiboDock program at RiboTargets was directed by Mohammad Afshar and he (with managerial support from Rod Hubbard, David Knowles and Simon Sturge) oversaw the further development into the program rDock. Many expert users at RiboTargets (subsequently Vernalis) provided ideas and testing of many aspects of the program. In particular:

- I-Jen Chen performed the initial validation experiments

- Ben Davis and Fareed Aboul-ela helped with development of NMR restrained docking protocols

- Michael Brunsteiner helped with improvements to the Simplex minimization algorithm

- Alba Macias tested the docking protocols with explicit water molecules

- Christine Richardson provided user feedback on feature developments

Maintenance and distribution of rDock was transferred to the University of York (Rod Hubbard) in 2006. There, a number of students helped maintain and validate the software and generate an initial website with user interfaces. Most recently, Sanjana Sood and Paul Bond made substantial contributions to this work.

Supplementary Figures

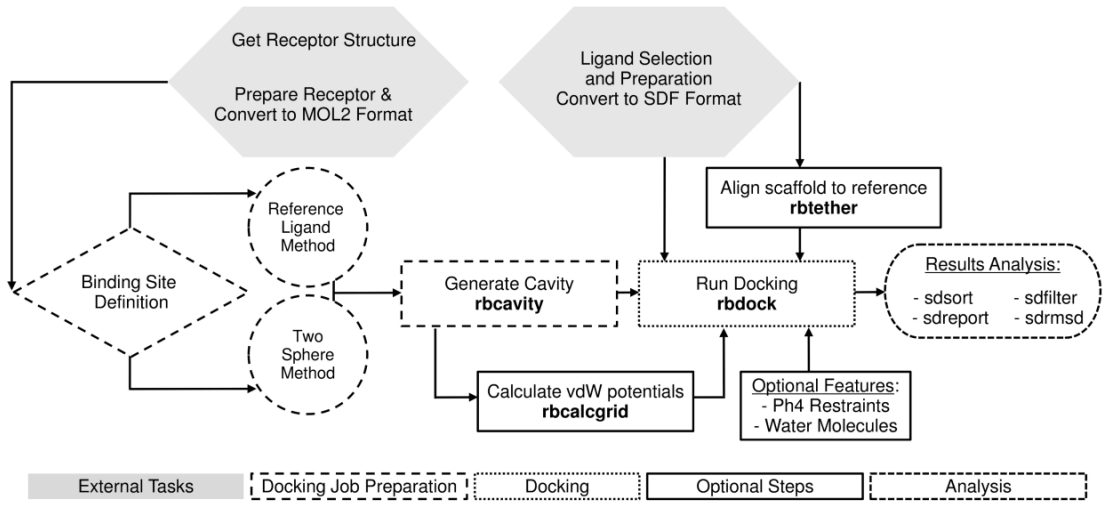


Figure S1. Workflow summary of an rDock docking job. Shapes in gray background are not covered with any rDock program and must be carried out with independent software.

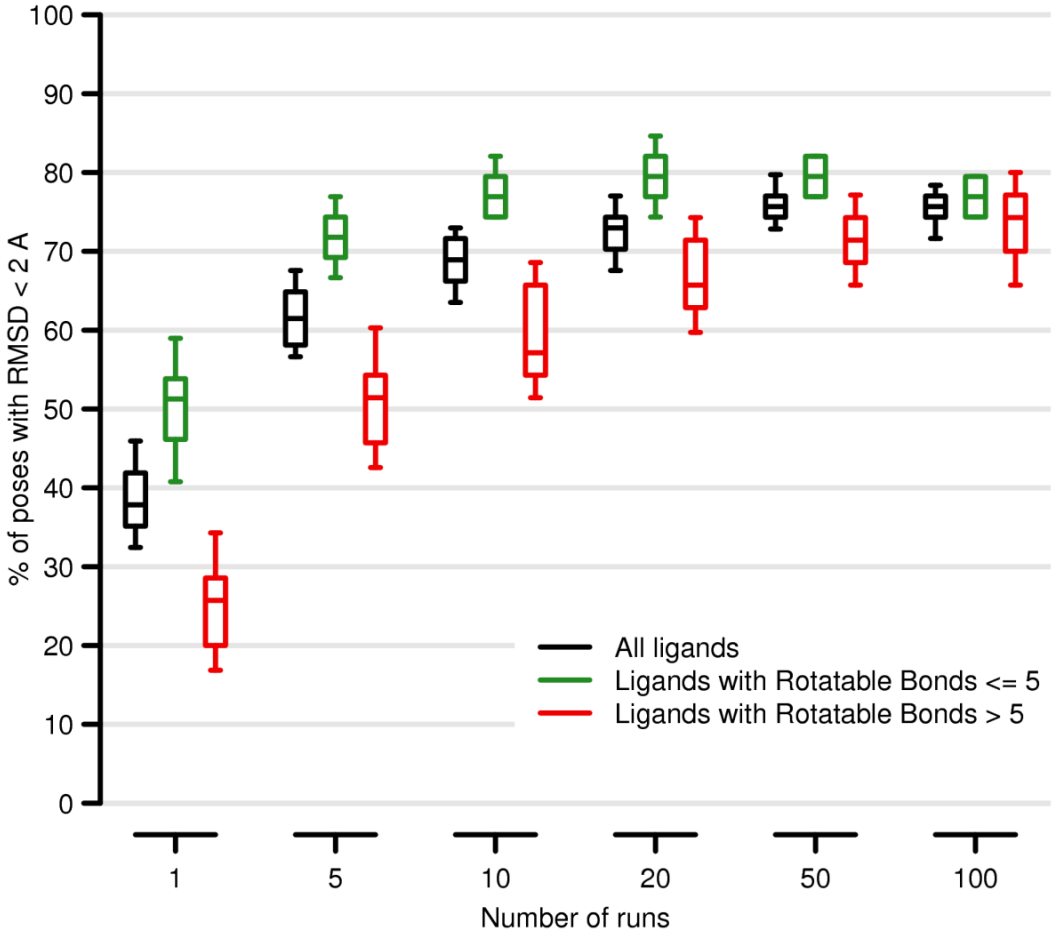


Figure S2. Binding mode prediction in the protein-ligand set (CCDC-Astex): Percentage of top-ranked poses with RMSD below 2.0Å as a function of the number of GA runs. The boxplot indicates the median value (out of 100 possible solutions) and the first and last quartile, while the whiskers span the 10% to 90% range. The whole set (black) has been sub-divided into ligands with 5 or fewer rotatable bonds (green) and the rest (red).

Figure S3. Matrix representation of the docking outcome for each system in the CCDC-Astex set for the three programs evaluated. A black area indicates that the best-scoring pose for a particular system-program combination has an RMSD below 2.0Å.

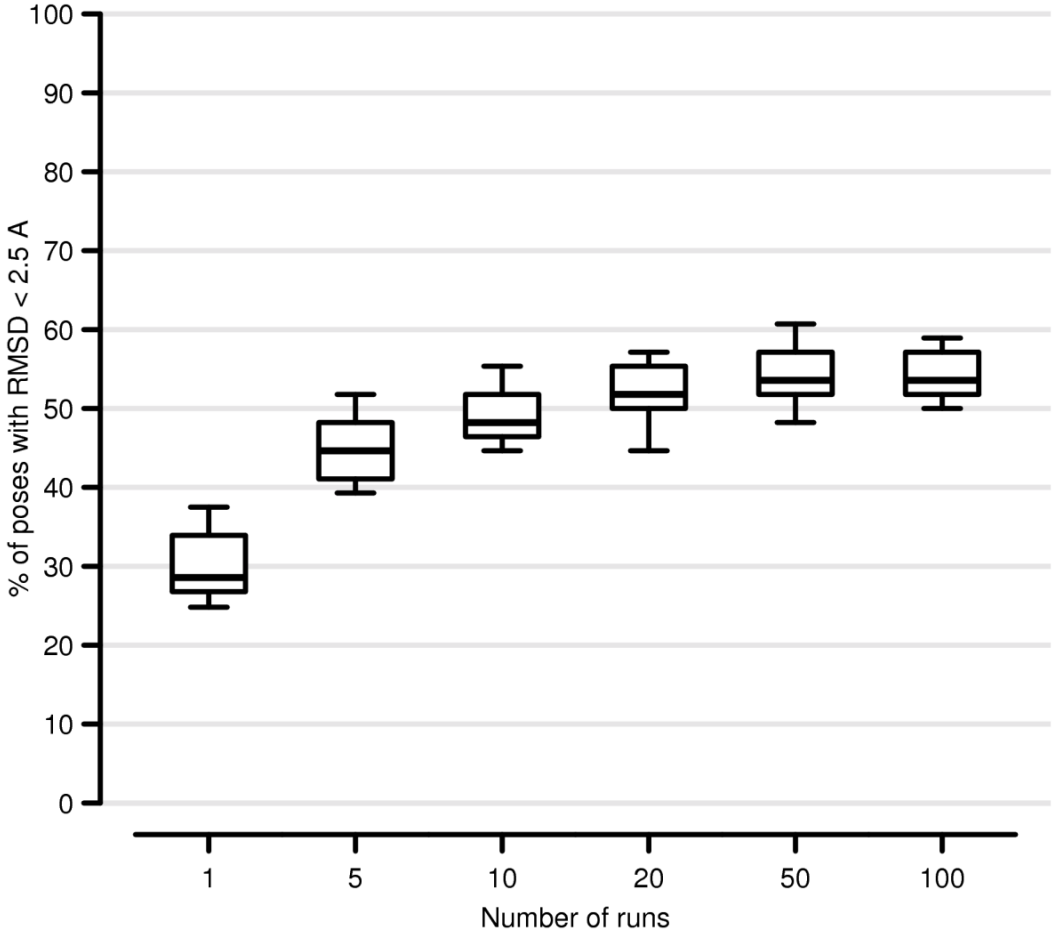


Figure S4. Binding mode prediction in the protein-RNA set: Percentage of top-ranked poses with RMSD below 2.5Å as a function of the number of GA runs. The boxplot indicates the median value (out of 100 possible solutions) and the first and last quartile, while the whiskers span the 10% to 90% range.

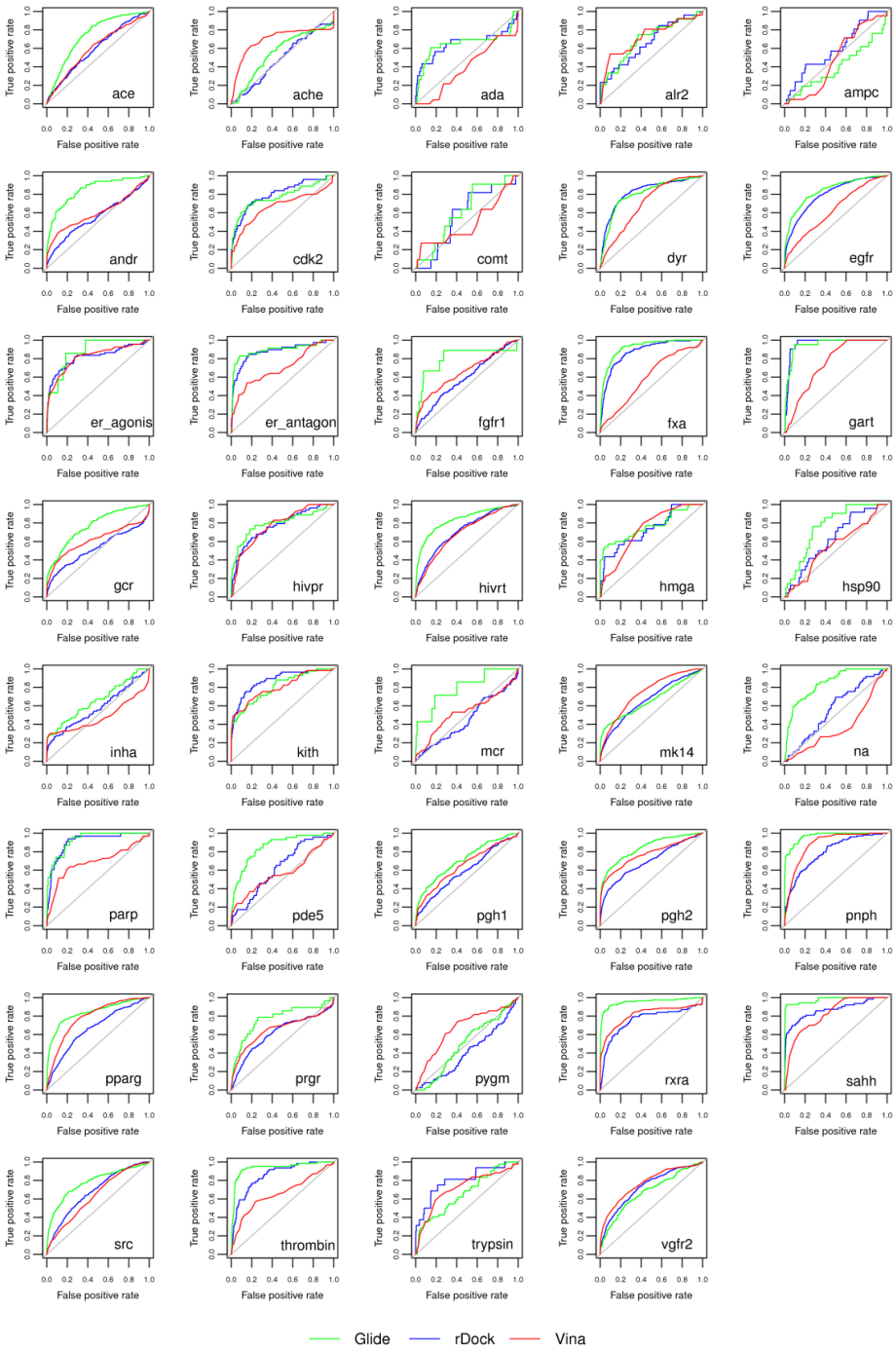


Figure S5. Receiver Operating Characteristic (ROC) Curves of all DUD systems. In the Y-axis, the true positive rate is the fraction of true positives out of the total actual positives and, in the X-axis, the false positive rate is the fraction of false positives out of the total actual negatives. In gray, ROC curve in case of random results.

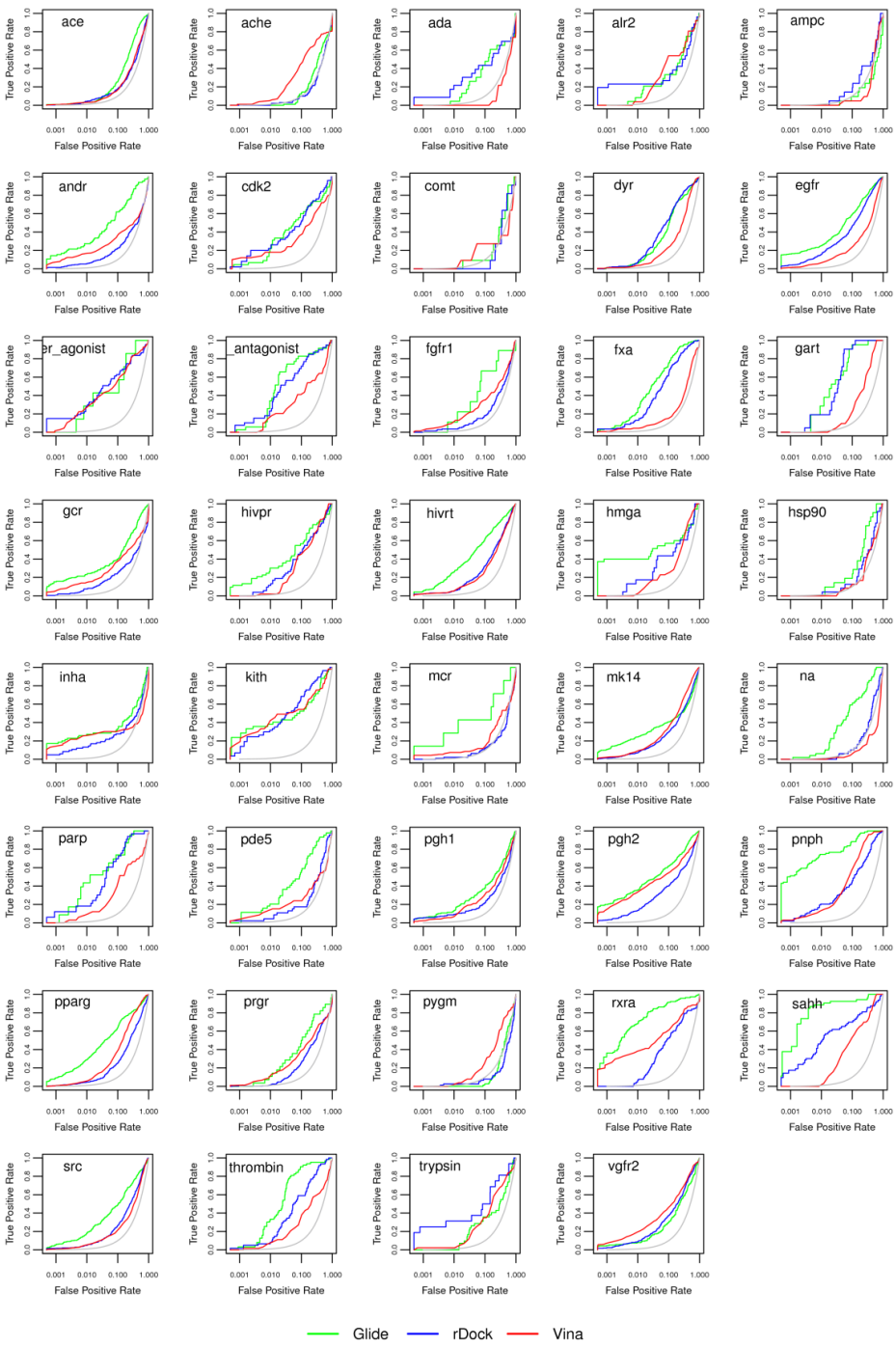


Figure S6. Semilogarithmic Receiver Operating Characteristic (ROC) Curves of all DUD systems. In the Y-axis, the true positive rate is the fraction of true positives out of the total actual positives and, in the X-axis in logarithmic scale, the false positive rate is the fraction of false positives out of the total actual negatives. In gray, semilogarithmic ROC curve in case of random results.

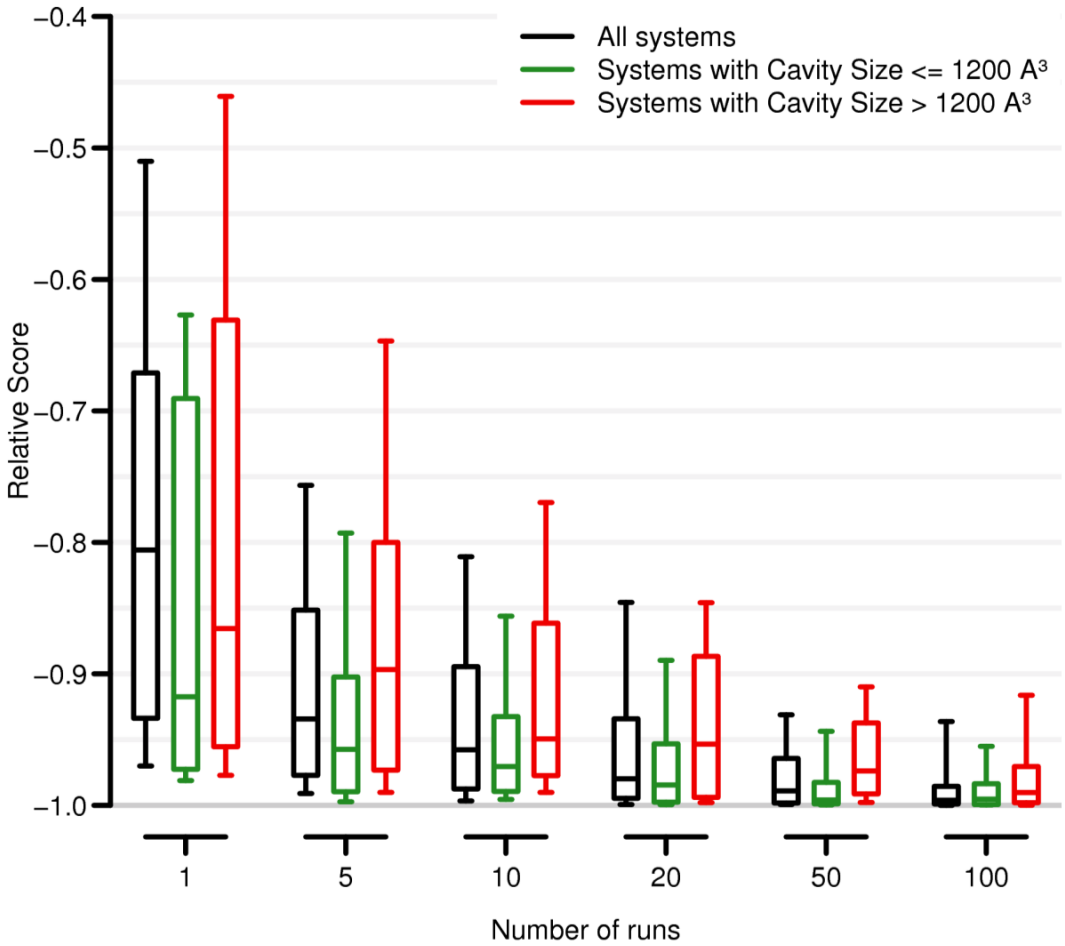
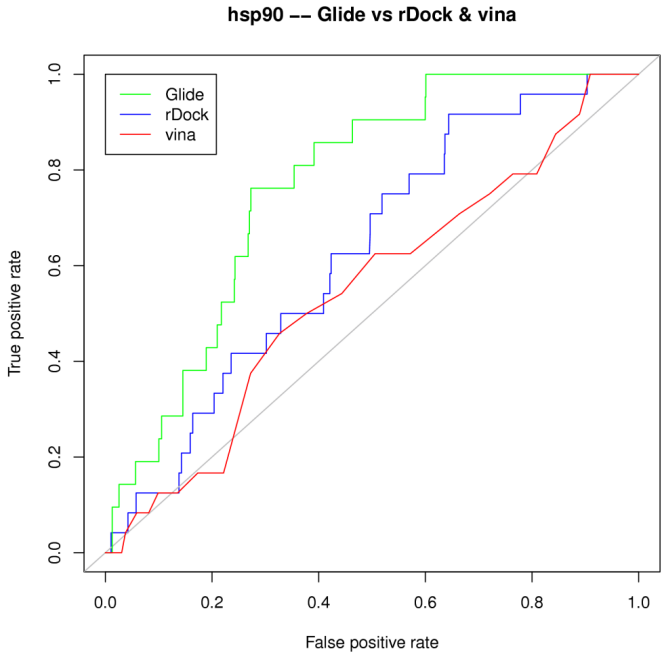


Figure S7. Relative score vs. the number of docking runs for all the protein-ligand complexes in the CCDC-Astex set. The boxplot indicates the median value (out of 1000 possible solutions) and the first and last quartile, while the whiskers span the 10% to 90% range. The whole set (black) has been sub-divided into systems with relatively small cavities (green) and the rest (red).



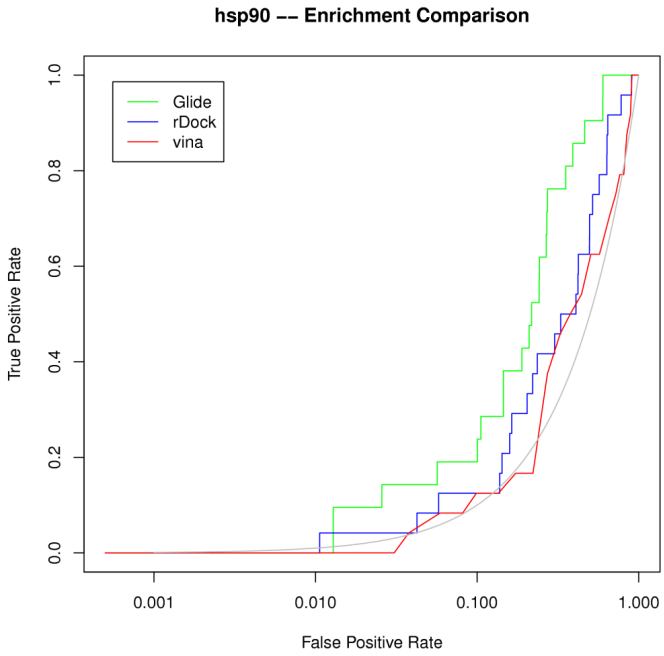
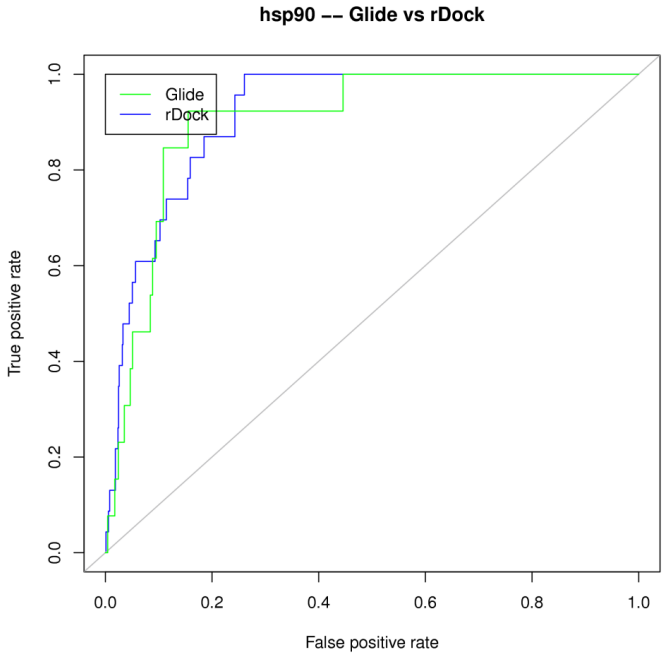


Figure S8.ROC curve of HSP90 without pharmacophoric restraints in normal (top) or semilogarithmic scale (bottom).



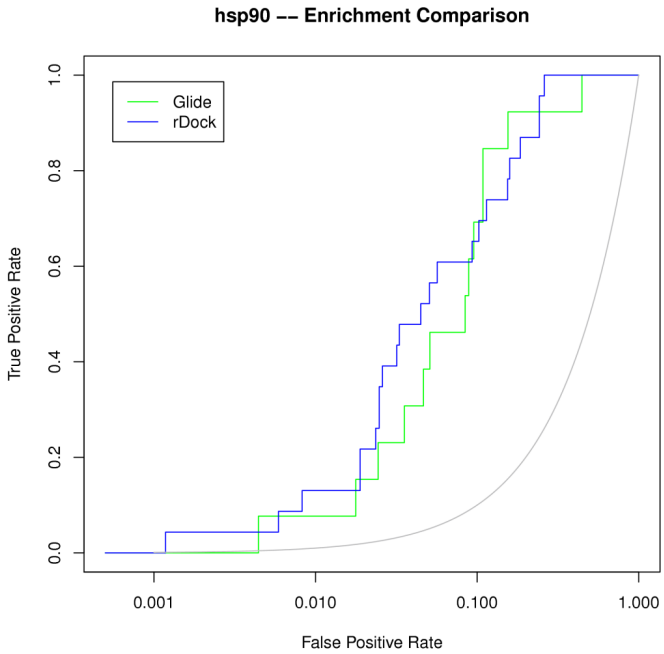


Figure S9.ROC curve of HSP90 withpharmacophoric restraints in normal (top) or semilogarithmic scale (bottom). It should be noted that using these settings, Glide only produces an output for 13 actives (out of 24) and 451 decoys (out of 864).

Table S1. Summary of statistics for all DUD systems and averages for each and all programs.

DUD system	Program	AUC	logAUC	EFmax	EF1	EF20
ace	rDock	0.6	0.14	11.95	4.24	1.47
	Glide	0.74	0.19	60.09	3.22	2.42
	Vina	0.6	0.15	8.59	3.08	1.62
ache	rDock	0.49	0.12	1.84	0.97	0.88
	Glide	0.55	0.14	1.48	0	1.33
	Vina	0.69	0.28	11.63	5.23	2.78
ada	rDock	0.66	0.35	70.35	19.54	2.82
	Glide	0.65	0.3	8.3	4.32	3.02
	Vina	0.39	0.09	1	0	0.19
alr2	rDock	0.67	0.35	207.92	20.79	2.11
	Glide	0.69	0.29	13.45	7.77	2.29
	Vina	0.72	0.31	6.71	3.15	2.44
ampc	rDock	0.59	0.2	2.66	0	1.9
	Glide	0.4	0.12	1.22	0	0.95
	Vina	0.48	0.12	1.22	0	0.39
andr	rDock	0.56	0.16	53.28	5.55	1.71
	Glide	0.85	0.4	797.68	27.18	3.69
	Vina	0.6	0.23	586.11	12.63	2.08
cdk2	rDock	0.8	0.4	87.4	21.36	3.5
	Glide	0.78	0.39	69.29	21.65	3.44
	Vina	0.67	0.32	174.8	17.48	2.38
comt	rDock	0.57	0.16	1.78	0	0.45
	Glide	0.62	0.21	4.76	0	1.34
	Vina	0.49	0.24	5.44	0	1.24
dyr	rDock	0.83	0.27	14.83	7.33	3.66
	Glide	0.8	0.25	14.55	5.82	3.69
	Vina	0.7	0.17	18.57	3.12	1.92
egfr	rDock	0.8	0.25	322.88	15.45	3.1
	Glide	0.84	0.35	2514.17	24.99	3.67
	Vina	0.63	0.16	129.17	6.51	1.74
er_agonist	rDock	0.82	0.44	326.57	28.2	3.73
	Glide	0.88	0.44	36.04	28.03	4.28
	Vina	0.82	0.4	45.72	24.19	3.51
er_antagonist	rDock	0.88	0.49	107.15	22.96	4.22
	Glide	0.89	0.52	36.23	33.44	4.28
	Vina	0.7	0.3	17.86	16.67	2.43
fgfr1	rDock	0.59	0.16	12.41	3.57	1.55
	Glide	0.81	0.35	18.09	11.06	3.33
	Vina	0.66	0.24	28.21	11.76	2.16
fxa	rDock	0.88	0.35	89.57	12.9	3.88
	Glide	0.92	0.42	53.09	24.99	4.49
	Vina	0.6	0.17	35.83	3.98	1.28
gart	rDock	0.96	0.56	44.13	18.91	4.96
	Glide	0.95	0.57	42.03	27.02	4.74
	Vina	0.74	0.26	2.6	0	1.99
gcr	rDock	0.53	0.16	57.91	7.33	1.74
	Glide	0.77	0.31	327.16	22.93	2.92
	Vina	0.62	0.24	144.77	13.36	2.32
hivpr	rDock	0.77	0.31	14.4	11.12	3.2
	Glide	0.8	0.43	186.91	29.32	3.67
	Vina	0.77	0.28	5.7	1.76	2.69
hivrt	rDock	0.71	0.2	111.18	5.03	2.48
	Glide	0.84	0.33	312	25.84	3.74
	Vina	0.69	0.19	37.29	5.91	2.23
hmga	rDock	0.73	0.34	29.87	15.93	2.81
	Glide	0.77	0.53	492	37.85	2.99
	Vina	0.73	0.28	6.09	2.72	2
hsp90	rDock	0.63	0.2	3.93	0	1.45
	Glide	0.77	0.28	7.4	0	2.13
	Vina	0.55	0.16	1.41	0	0.75
inha	rDock	0.58	0.23	144.82	12.85	1.86
	Glide	0.66	0.34	438.17	26.68	2.21
	Vina	0.48	0.28	325.84	19.91	1.62

kith	rDock	0.88	0.46	140	31.03	3.94
	Glide	0.8	0.48	437.86	36.87	3.09
	Vina	0.8	0.47	350	32.5	3.25
mcr	rDock	0.46	0.11	4.55	2.1	0.9
	Glide	0.8	0.45	126.29	28.06	3.57
	Vina	0.54	0.17	218.64	6.49	1.63
mk14	rDock	0.66	0.17	16.04	7.09	2.09
	Glide	0.66	0.26	807.95	24.16	2.34
	Vina	0.74	0.2	124.11	8.84	2.54
na	rDock	0.57	0.15	1.63	0	1.22
	Glide	0.84	0.36	17.07	6.03	3.36
	Vina	0.37	0.09	1	0	0.64
parp	rDock	0.91	0.49	134.42	16.8	4.54
	Glide	0.93	0.58	68.43	42.77	4.33
	Vina	0.69	0.29	11.2	7.91	2.88
pde5	rDock	0.63	0.2	38.98	4.33	1.41
	Glide	0.85	0.4	100.97	20.19	3.74
	Vina	0.57	0.24	38.98	12.99	1.69
pgh1	rDock	0.6	0.19	275.23	9.17	1.69
	Glide	0.71	0.26	160.74	13.23	2.53
	Vina	0.65	0.2	41.48	7.67	1.97
pgh2	rDock	0.7	0.23	52.98	13.53	2.52
	Glide	0.84	0.42	1955.11	35.47	3.6
	Vina	0.77	0.34	637.02	31.1	2.92
pnph	rDock	0.8	0.3	134.95	20.24	3.05
	Glide	0.97	0.72	1720.44	73.02	4.88
	Vina	0.88	0.35	50.61	12.19	3.68
pparg	rDock	0.67	0.17	9.51	4.57	2.11
	Glide	0.84	0.34	202.49	28.86	3.84
	Vina	0.79	0.23	8.24	5.67	3.08
prgr	rDock	0.62	0.16	5.92	1.7	2.2
	Glide	0.77	0.28	12.83	10.57	3.21
	Vina	0.66	0.24	35.5	12.01	2.33
pygm	rDock	0.41	0.09	5.68	2.56	0.58
	Glide	0.5	0.1	1.18	0	0.53
	Vina	0.67	0.2	2.95	2.38	2
rxra	rDock	0.76	0.26	8.51	6.81	3.32
	Glide	0.95	0.65	654.83	67.62	4.7
	Vina	0.81	0.45	1111.72	29.6	3.41
sahh	rDock	0.87	0.51	328.38	48.48	3.97
	Glide	0.98	0.83	698.49	86.39	4.71
	Vina	0.82	0.29	7.63	2.67	3.26
src	rDock	0.68	0.17	262.56	4.77	2.16
	Glide	0.78	0.27	133.95	18.12	3.28
	Vina	0.64	0.15	32.84	5.16	1.63
thrombin	rDock	0.86	0.36	36.89	8.02	3.76
	Glide	0.94	0.5	40.36	32.25	4.75
	Vina	0.66	0.28	11.9	7.44	2.42
trypsin	rDock	0.8	0.45	320.75	30.84	3.43
	Glide	0.66	0.24	7.42	0	2.02
	Vina	0.72	0.27	35.35	2.21	2.69
vgfr2	rDock	0.73	0.22	365.4	8.73	2.55
	Glide	0.68	0.2	170.25	7.83	2.3
	Vina	0.77	0.27	426.3	17.46	2.84
	Program	**AUC**	**logAUC**	**EFmax**	**EF1**	**EF20**
Averages	rDock	0.70	0.27	98.95	11.66	2.54
	Glide	0.78	0.37	326.94	22.91	3.22
	Vina	0.66	0.25	121.54	9.12	2.17
	Total	0.71	0.30	182.48	14.56	2.64

AUC: Area Under the ROC Curve. **LogAUC**: Area Under the semilog ROC Curve. **EFmax**: Maximal Enrichment Factor. **EFX**: Enrichment Factor when the top x% of the virtual collection is selected.

Table S2. Spearman's rank correlation coefficient (ρ)* between programs on the Hsp90 DUD set.

Program	rDock	rDock-guided	Glide	Glide-guided	Vina
rDock	1	0,37	0,46	0,22	0,33
rDock-guided		1	0,37	0,52	0,13
Glide			1	0,41	0,31
Glide-guided				1	0,13
Vina					1

* The introduction of empirical information produces results very different to the default parameters restraints (ρ = 0.37 for rDock; ρ = 0.41 for Glide) while the output of different programs becomes more similar (ρ = 0.52, comparing rDock-guided with Glide-guided). ρ is calculated from the rank of 464 molecules for which Glide-guided produces an output.

Table S3. AUC for the 4 DUD systems used for calculating the time performance.

	Vina[1]	Glide SP[1]	rDock			
			Grid-based SF		Indexed SF	
			VS[2]	Full[1,3]	VS[2]	Full[1,3]
ADA	0.39	0.67	0.68	0.64	0.57	0.62
COMT	0.51	0.69	0.67	0.62	0.65	0.64
PARP	0.68	0.9	0.88	0.86	0.87	0.86
Trypsin	0.74	0.49	0.63	0.62	0.66	0.76
Average	**0.58**	**0.69**	**0.72**	**0.69**	**0.69**	**0.72**

[1] Default program parameters were used. [2] On HTVS mode, the average number of docking runs needed for these 4 systems is 10. [3] 50 docking runs are used for default docking.

Dynamic Undocking and the Quasi-Bound State as Tools for Drug Design

Sergio Ruiz-Carmona[1], Peter Schmidtke[2], F Javier Luque[1], Lisa Baker[3], Natalia Matassova[3], Ben Davis[3], Stephen Roughley[3], James Murray[3], Rod Hubbard[3,4] and Xavier Barril[1,5]

Nature Chemistry, 2016, Accepted

1 Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Facultat de Farmacia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain, **2** Discngine, 33 rue du Fauburg Saint-Antoine, 75011 Paris, France, **3** Vernalis (R&D) Ltd, Granta Park, Cambridge, CB21 6GB, UK, **4** YSBL, University of York, Heslington, York, YO10 5DD, UK, **5** Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

Dynamic Undocking and the Quasi-Bound State as Tools for Drug Design

Sergio Ruiz-Carmona,[1] Peter Schmidtke,[2] F. Javier Luque,[1] Lisa Baker,[3] Natalia Matassova,[3] Ben Davis,[3] Stephen Roughley,[3] James Murray,[3] Rod Hubbard,[3,4] Xavier Barril[1,5,*]

[1] Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain.

[2] Discngine, 33 rue du Fauburg Saint-Antoine, 75011 Paris, France

[3] Vernalis (R&D) Ltd, Granta Park, Cambridge, CB21 6GB, UK

[4] YSBL, University of York, Heslington, York, YO10 5DD, UK

[5] Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

* Send correspondence to: xbarril@ub.edu

There is a pressing need for new technologies that improve the efficacy and efficiency of drug discovery. Structure-based methods have contributed towards this goal but they focus on predicting the binding affinity of protein–ligand complexes, which is notoriously difficult. We adopt an alternative approach that evaluates structural, rather than thermodynamic, stability. Noting that bioactive molecules present a static binding mode, we devised Dynamic Undocking (DUck), a fast computational method to calculate the work necessary to reach a quasi-bound state, where the ligand has just broken the most important native contact with the receptor. This non-equilibrium property is surprisingly effective in virtual screening because true ligands form more resilient interactions than decoys. Notably, DUck is orthogonal to docking and other 'thermodynamic' methods. We demonstrate the potential of the docking–undocking combination in a fragment screening against the molecular chaperone and oncology target Hsp90, for which we obtain novel chemotypes and a hit rate approaching 40%.

Structural stability is a fundamental property of protein–ligand complexes. Though cases of dual binding modes have been reported,[1,2] they are generally not dynamic, or involve predominantly hydrophobic interactions,[3] which lack directionality and do not impose strict geometric constraints.[4] By contrast, hydrogen bonds are ideal to provide structural stability because they have sharp distance and angular dependencies.[4] Their contribution to the free energy of binding (ΔG_{bind}) is variable but can be substantial.[5] Importantly, they often act as anchoring points in protein–ligand complexes, providing the minimal binding unit through one or a few hydrogen bonds as demonstrated for fragment-sized ligands.[6,7] We have previously shown that certain hydrogen bonds present strong opposition to small structural distortions and can act as kinetic traps because the local environment hinders the transition from a direct hydrogen bond to a water-bridged interaction.[8] As an early unbinding event, rupture of the so-called water-shielded hydrogen bonds can influence the whole dissociation process.[8,9] Taken together, these observations suggest that hydrogen bonds are the main determinants of structural stability, and lead us to postulate that their resilience should provide information about the binding potential of candidate ligands. Thus, we set out to investigate whether the work required to disrupt intermolecular hydrogen bonds can be used to predict ligand binding.

We will introduce DUck, a simplified computational procedure to calculate the work needed to break a key native contact, reaching a quasi-bound state (W_{QB}). Then, we will show that active compounds are structurally stable and present higher W_{QB} values than inactive ones. Finally, we demonstrate the use of this

property in virtual screening (VS) applications, showing that DUck complements the thermodynamic perspective offered by existing methods.

Results and Discussion

Simplified simulation of the early dissociation stage

To assess the hypothesis, we have devised Dynamic Undocking (DUck) simulations, where a key intermolecular hydrogen bond is pulled from an initial distance of 2.5 Å (close contact) to 5.0 Å (broken contact). In order to focus on just one specific hydrogen bond, we use model receptors comprising only the protein residues that are within 6 Å of the given hydrogen bond (Figure 1A). The work necessary to carry out the steering process is monitored, and we define the quasi-bound (QB) state as the point along the simulation where the work profile presents the highest value. W_{QB} is the work necessary to depart from the ideal hydrogen bond configuration and reach the QB state (Figure 1B). Notably, this is a non-equilibrium property, and there is no reason why it should correlate with any measurement of binding affinity. What is more, as the unbound state is not considered, W_{QB} cannot inform about the binding free energy. Instead, this magnitude solely indicates if the interaction under investigation gives rise to a (local) minimum in the free energy landscape and estimates the depth of said minimum (Supplementary Figure 1).

Relationship between W_{QB} and binding affinity

As an initial proof of concept, we apply DUck to a set of 41 fragment-like ligands (<300Da) of the cyclin dependent kinase 2 (CDK2) with known binding mode and half maximal inhibitory concentration (IC_{50}) values. The hinge region of all kinases is a hot spot for binding, where the protein backbone offers privileged hydrogen-bonding opportunities.[10] For CDK2, the central hydrogen-bond donor (NH of Leu83) is the most conserved interaction site and was used to define the reaction coordinate. W_{QB} presents only a weak correlation with binding affinity (Supplementary Figure 2), but the distribution of W_{QB} values is clearly skewed (Figure 2A and Supplementary Figure 3). Thus, 65% of weak binders (IC_{50} > 1 μM) present W_{QB} values below 6 kcal/mol, while all strong binders (IC_{50} < 1 μM) pass this threshold. Ligand 3FZ1,[11] is the clear exception as it presents an almost flat dissociation profile (W_{QB} = 0.12 kcal/mol). This is explained by an unsuitably long (3.4 Å) interaction with the hinge region, involving a methoxy group, which is a poor hydrogen bond acceptor.[4] Instead, this unusual ligand forms two charge-reinforced hydrogen bonds with Lys33 and Asn132, from which it draws structural stability (Supplementary Figure 4). This shows that some ligands can use alternative or additional interaction points to attain structural stability, in which case, DUck calculations (as currently implemented) may underestimate the cost of breaking the native contacts.

To further examine the surprising relationship between binding affinity and W_{QB}, we use the bromodomain and extra-terminal (BET) BRD4-BD1 as additional test system. The side-chain N of Asn140 is a well-known pharmacophoric point of this epigenetic target,[12] and defines the key intermolecular hydrogen bond. Again, we observe the same trend, i.e. higher W_{QB} for more potent ligands, but with a large

114 dispersion that blurs correlation (Supplementary Figures 3 and 5). Interestingly,
115 the lowest W_{QB} values (0, 1.1 and 1.7 kcal/mol) correspond to three kinase
116 inhibitors with off-target activity for the BRD4-BD1.[13] Thus, achieving potency in
117 the absence of a robust anchoring interaction is possible, but rare, which suggests
118 that it is an ineffective strategy.
119
120 DUck is very effective in virtual screening
121 We then assess whether the approach can be used in virtual ligand screening by
122 testing the ability of DUck to distinguish true CDK2 ligands from a set of carefully
123 selected decoys[14] for which we had generated binding modes by docking. The
124 distribution of W_{QB} is strikingly different from the active set, with 61% of
125 molecules presenting values below 2 kcal/mol and 49% below 1 kcal/mol (Figure
126 2A). This indicates that, in spite of forming the key hydrogen bond, this interaction
127 is labile for most of the docking decoys, which would translate to an unstable
128 binding mode. We therefore propose that W_{QB} can distinguish true ligands from
129 inactive molecules, as shown in the receiver operating characteristics (ROC) curves
130 (Figure 2B). To demonstrate the wider applicability of the method, we conducted
131 similar experiments with the adenosine A2A receptor (AA2R) and Trypsin, as
132 representatives of G protein-coupled receptors (GPCR) and serine proteases,
133 respectively (Figure 2B). Together with kinases (such as CDK2) these protein
134 families include a large part of the current and investigational drug targets.[15] The
135 key hydrogen bonds tracked by the DUck simulations involve the side-chain
136 carbonyl of Asn253, in the case of AA2R, and the carboxylic acid of Asp189, for
137 Trypsin. As shown in Figure 2B, the results for these systems are even better than
138 for CDK2, demonstrating that DUck is surprisingly effective in virtual screening.

139 Importantly, the performance improves consistently as sampling increases, but
140 good enrichments can be obtained with as little as 2 DUck runs per ligand
141 (Supplementary Figure 6).
142
143 DUck is orthogonal to existing methods
144 These results position DUck as a new method for virtual screening. But, as it aims
145 to predict a property that is fundamentally different from thermodynamic stability,
146 we investigate its complementarity with molecular docking, a method with a long
147 and successful history of application in virtual screening.[16,17] Using the rDock
148 software,[18] we find that docking scores have no correlation with W_{QB}, and good
149 docking scorers are nearly as likely to present a low resistance to dissociation as
150 the rest of the decoys (Figures 2C, 2D and Supplementary Figure 7). As such,
151 molecular docking and dynamic undocking can be considered orthogonal (i.e.
152 perfectly complementary) and the intersection between both techniques defines a
153 region highly enriched in true ligands. We have also performed extensive
154 calculations with other virtual screening tools (Glide docking, MMPBSA and
155 MMGBSA re-scoring). The results, summarised in Supplementary Figures 8 and 9,
156 confirm that DUck is complementary to all of them. In fact, as we obtain low W_{QB}
157 values for many decoys with good scores by all other methods, DUck post-filtering
158 delivers several fold improvement even when applied to a consensus list by two
159 independent 'thermodynamic' approaches (Figures 2E, 2F and Supplementary
160 Figure 10). These results support the idea that structural stability of the binding
161 mode, just like good chemical complementarity, is a necessary – but not sufficient –
162 condition for binding. By imposing both conditions simultaneously, we can
163 multiply the effectiveness of structure-based virtual screening. At the same time,

using W_{QB} as a post-docking filter means that only the best-scoring subset of the virtual chemical collection needs to be reassessed by DUck simulations, thus improving computational efficiency.

Fragment discovery with in tandem docking-undocking calculations.

To demonstrate the power of the docking-undocking combination, we have applied the method prospectively for the identification of small molecules that bind the molecular chaperone, Heat Shock Protein 90KDa (Hsp90). This oncology target has been a test-bed and paradigm in fragment and structure-based drug design.[19] With hundreds of Hsp90-ligand complexes deposited in the Protein Data Bank (PDB), discovery of novel chemotypes is very challenging. We focused on fragment-like molecules, as this may be the most efficient way to discover new leads and to generate scaffold-hoping ideas.[20,21] A collection of 280000 fragment-sized molecules was docked to the ATP binding site of Hsp90. A diverse set of 139 molecules from the best 450 (top 0.16%) was then selected and each one was subjected to 100 DUck runs to obtain fully converged W_{QB} values (note that fewer DUck runs would have given similar results (Supplementary Figure 11)). The distribution of W_{QB} values (Figure 3A, Supplementary Figure 12) shows that even at the upper limit of the docking score distribution a large proportion of putative ligands present low resistance to dissociation, with 32%, 50% and 80% presenting W_{QB} below 3, 4 and 6 kcal/mol, respectively. We purchased all the molecules from the high stability set (W_{QB} > 6 kcal/mol) that were available (n=21). They were tested using three different ligand-observed Nuclear Magnetic Resonance (NMR) experiments, in the absence or presence of a known competitor to confirm that fragment hits bind at the target site.[19] Eight out of the 21 molecules (38%) were

189 confirmed as true hits (Table 1). Crucially, for the same system and screening
190 method, the hit rate obtained with a general fragment screening library is 4.4%.[22]
191 Therefore, the DUck-based virtual screening increases the efficiency by nearly an
192 order of magnitude. This is similar to optimal virtual fragment screening results
193 reported for other systems.[23]. In order to better assess the contribution of DUck to
194 the success rate, we also purchased and collected data for 15 molecules from the
195 medium stability set (W_{QB} between 3 and 6 kcal/mol) and 11 from the low stability
196 set (W_{QB} < 3 kcal/mol). Only one molecule from these sets was a hit and,
197 importantly, its W_{QB} value is very close to the upper threshold (5.6 kcal/mol). This
198 confirms that DUck false negatives (i.e. active molecules with low W_{QB}) are rare, an
199 ideal property for a screening method. Hit rates for the three categories are
200 summarized in Figure 3B.
201
202 To assess the value of the hits as starting points, we have compared their chemical
203 structures to existing Hsp90 ligands, finding low similarity in all cases (Table 1).
204 Binding mode determination and analysis of the main interactions that define the
205 chemical scaffold offers a more precise assessment of their novelty. Crystal
206 structures for 3 of the fragment hits were determined by X-ray crystallography
207 (Figure 4 and Supplementary Figure 13). This confirmed that the docking pose
208 used as starting position for the DUck experiments was correct, particularly
209 regarding the key interaction that was being monitored (side-chain of Asp93).
210 Compound **1** is the most potent fragment hit (dissociation constant K_D=77μM) and
211 has a ligand efficiency (LE) of 0.33 kcal/mol per non-hydrogen atom, similar to
212 other Hsp90 fragment hits that have been evolved into very efficient lead
213 compounds.[24] Many 2-aminopyrimidines have been described as Hsp90 ligands,[19]

214 confirming the potential of the fragment hit, but the relative lack of novelty would
215 advise against using this fragment as starting point at this stage. Compound **2** is
216 less potent (K_D=320μM) but equally efficient (LE=0.32) by virtue of having fewer
217 atoms. In this case, the key interaction with Asp93 is mediated by an aminothiazole
218 moiety, which is unprecedented and would constitute a good starting point to
219 develop new chemical entities. Compound **3** (K_D=700μM; LE=0.25) belongs to the
220 well-known family of resorcinol inhibitors, which includes the clinical candidate
221 NVP-AUY922,[19] but provides an interesting example of scaffold hopping, where the
222 oxime acts a bioisosteric replacement of the five-membered rings included as core
223 scaffold in the patents. Compounds **4**, **5** and **6** also represent completely novel
224 starting points, as their scaffold is unique amongst Hsp90 inhibitors. The binding
225 mode could not be confirmed experimentally, but is likely correct because two
226 independent methods deemed the molecules active based on the predicted
227 geometry (Their predicted binding modes can be found in Supplementary Figure
228 14).
229
230 **Conclusions**
231
232 In summary, we have demonstrated that the concept of structural stability can be
233 used very effectively in structure-based drug design, complementing the standard
234 focus on binding free energy. Hydrogen-bonding groups in the active site are
235 privileged structures to fix the ligand in place, particularly when they act as
236 binding hot spots and can form water-shielded hydrogen bonds.[8] The work needed
237 to break such interactions (W_{QB}) is very useful to detect true ligands even though it
238 is a non-equilibrium property that is not expected to correlate with ΔG_{bind}. This

intriguing fact may reflect the nature of proteins, which have been designed to bind their natural ligands not only with high affinity and selectivity, but also forming structurally stable complexes. Thus, it will be important to test the approach on other types of supramolecular assemblies. Dynamic Undocking (DUck), a particular implementation of steered molecular dynamics, allows us to calculate W_{QB} in a very efficient manner. DUck can be used in combination with existing 'thermodynamic' approaches to multiply their effectiveness. The docking-undocking combination has proven particularly useful for virtual fragment screening, delivering novel, diverse and suitable starting points with a hit rate of 38%. At present, we focus on a single key hydrogen bond to estimate W_{QB}, which requires previous knowledge and has a critical impact on the outcome. Future investigations should address the extension of the method to multiple sites and other interaction types to improve performance and avoid reliance on extrinsic decisions. DUck inherits the intrinsic limitations of structure-based methods (e.g. protein flexibility, quality of the force-field) and may have some of its own (e.g. long range effects, steering conditions). Further tests will reveal its true potential, but considering that it is orthogonal to existing methods and computationally very efficient, we expect that it will be rapidly adopted by the structure-based drug design community and adapted to other biotechnological applications involving non-covalent complexes.

METHODS

Dynamic Undocking

Dynamic Undocking (DUck) is a particular type of Steered Molecular Dynamics (SMD),[25] where we force the rupture of an intermolecular hydrogen bond formed between a pre-defined interaction point in the receptor and a complementary atom in the ligand. Additionally, we use a model receptor that includes only the minimal subset of the protein necessary to preserve the local environment around the hydrogen bond that is being monitored. This transformation minimizes the influence of peripheral interactions, thus simplifying the dissociation pathway and facilitating convergence (Supplementary Figure 15). As an added bonus, it speeds up the calculations by a factor of 5 (Supplementary Table 2). The first and essential step is to identify an atom of reference in the protein, which must form a hydrogen bond with all (or most) known ligands. For well-known systems, like the ones used here, it can be identified from a structural superimposition of all the available protein-ligand complexes. On novel binding sites, it may be identified with a quantitative hot spot identification method.[26]. Then, the model receptor is generated from a representative 3D structure of the protein by selecting all residues with at least one atom within 6 Å of the atom of reference. The selection is visually inspected and, if needed, additional residues that are deemed necessary to preserve the local environment are included in the selection. Unselected residues are eliminated and truncated side chains are acetylated or N-methylated, as needed. Interstitial water molecules, if present, are preserved. The PDB codes, reference interaction points and the list of protein residues and water molecules for each system are listed in Supplementary Table 3. Given the model receptor

286 (protein chunk) and a set of ligands properly oriented (docking poses or
287 superimposed X-ray geometries), a MOE[27] SVL script developed in house
288 automatically performs the following steps: 1) Calculates AM1-BCC charges for the
289 ligand.[28] 2) Assigns parm@Frosst[29] atom types and non-bonded parameters to the
290 ligand. 3) Identifies the ligand atom that is hydrogen-bonded to the protein's
291 reference atom (based on distance and type). 4) Writes input and execution files to
292 carry out the MD simulations with AMBER[30]. 5) Calls AMBER's tLeap to generate
293 valid topology and coordinate files for each individual receptor-ligand complex.
294 For the protein, the AMBER force field 99SB is used. Each system is placed in a
295 cuboid box spanning at least 12 Å more than the furthest atom in each direction.
296 The box is then filled with TIP3 water molecules to create periodic boundary
297 conditions. When needed, Na+ or Cl- ions are added to force the neutrality of the
298 whole system. MD simulation conditions (where non-default) are as follows: 1) At
299 all stages, harmonic restraints with a force constant of 1 kcal/mol·Å^2 are placed on
300 all non-hydrogen atoms of the receptor to prevent structural changes. 2)
301 Spontaneous rupture of the key hydrogen bond during non-steered simulations is
302 prevented with a gradual restraint for distances beyond 3 Å (parabolic with k=1
303 kcal/mol·Å^2 between 3Å and 4Å and linear with k=10 kcal/mol·Å beyond 4 Å). 3)
304 All equilibration and simulation steps were run using a Langevin thermostat with a
305 collision frequency of 4 ps^{-1} and the cutoff for non-bonded interactions was set to
306 9Å. 4) Bonds involving hydrogen are constrained using SHAKE.[31] In order to
307 equilibrate the system the following steps are executed: 1) Energy minimization
308 for 1000 cycles. 2) Assignment of random velocities at 100K and gradual warming
309 to 300K for 400 ps in the NVT ensemble. 3) Equilibration of the system for 1 ns in
310 the NPT ensemble (1 atm, 300K). At this stage, the first SMD simulations can be

311 executed. We run two SMDs from the same restart file, but at different
312 temperatures (300K and 325K) to ensure that the trajectories proceed differently.
313 The SMD lasts 500 ps, during which time the distance between the key hydrogen
314 bonds is steered from 2.5 Å to 5.0 Å (constant velocity of 5 Å/ns) with a spring
315 constant of 50 kcal/mol·Å^2. We have tested slower velocities and the results are
316 essentially unchanged (Supplementary Figure 16). The spring constant had little
317 influence and on a limited test set we obtained essentially identical results in the
318 range k=10 kcal/mol·Å^2 to k=1000 kcal/mol·Å^2. We have also investigated the
319 importance of the specific reaction coordinate by using the closest contact
320 between CDK2 Leu83:O and the ligand (instead of Leu83:N). The W_{QB} values
321 obtained with these different atoms of reference (located only 3 Å apart) present a
322 high correlation (r^2=0.75; Supplementary Figure 17). By contrast, when the atoms
323 of reference involve completely different part of the ligand, the results are
324 uncorrelated (Supplementary Figure 18). To generate diverse starting points for
325 SMD trajectories, we perform 1ns unbiased MD simulation and repeat the process
326 as many times as desired (e.g. 50ns unbiased MD simulations are needed to
327 execute 100 SMD trajectories). All simulations were performed with Amber 12
328 adapted for running in GPUs and executed either in-house with NVIDIA GeForce
329 TITAN X GPUs or at the Barcelona Supercomputing Center using NVIDIA Tesla
330 M2090 GPUs. The simulations took 24 minutes (unbiased MD) or 30 minutes
331 (SMD) of wallclock time per nanosecond (average values for the systems tested on
332 the TITAN GPUs). Work profiles outputted by the SMD simulations are processed
333 as explained in the main text to obtain W_{QB} values. Various methods could be used
334 to obtain free energies from the SMD work, but they have strict convergence
335 requirements, are computationally much more expensive and the results are only

336 valid if the reaction coordinate is mechanistically correct.[25] Instead, we simply
337 assume that W_{QB} is an upper limit to the equivalent magnitude in free energy
338 (ΔG_{QB}). In order to get as close as possible to ΔG_{QB}, we run multiple SMD replicas
339 and take the overall lower W_{QB} as the representative value. Note that we have used
340 very conservative settings, favouring sampling over computational efficiency.
341 Based on convergence analysis (Supplementary Figures 6 and 11) and other tests,
342 we propose the protocol shown in Supplementary Figure 19 for virtual screening.
343 Less than one GPU hour per ligand would be necessary to discard approximately
344 80% of candidate ligands and produce a reasonable estimate of W_{QB} for the
345 remaining ones. By comparison, a high-throughput implementation of MM-PBSA (1
346 ns of sampling) would require at least 3 GPU hours plus 20 CPU minutes per
347 ligand.
348
349 **Hsp90 virtual screening**
350 A collection of 280000 purchasable fragment-sized molecules (<250 Da), were
351 docked to the ATP binding site of Hsp90 with an optimized protocol, where the key
352 hydrogen bond with Asp93 is enforced.[18] Next, we grouped the 1000 top scoring
353 molecules into 400 clusters based on chemical similarity and visually inspected the
354 top-scoring molecule within each cluster to select 139 molecules that were
355 subjected to DUck simulations. Docking score was the main selection criterion,
356 with 90 molecules originating from the top 200 and all of them within the top 450.
357 Additional criteria included high predicted aqueous solubility and chemical
358 diversity. The selected molecules were subjected to 100 DUck calculations. We
359 divided the molecules in three categories according to their resistance to
360 dissociation: weak ($W_{QB} < 3$; N=44; 32%), medium ($3 < W_{QB} < 6$; N=67; 48%) and

361 strong (W_{QB} > 6; N=28; 20%). We tested all the molecules that we could buy from
362 the strong set. For comparison, we also purchased and tested 15 molecules of
363 medium and 11 from the low stability sets. The chemical structures of the 47
364 compounds are shown in Supplementary Figure 12.
365
366 **Screening by NMR**
367 Identification of compounds which bind to the ATP site of Hsp90α was performed
368 as described previously.[32,33] Briefly, a number of 1D ^{1}H NMR experiments (STD,
369 water-LOGSY, relaxation filtered) were used to identify interactions between
370 compounds and the protein; a potent competitor (PU3) was then added in order to
371 block the ATP binding site. Compounds which bound and were then displaced
372 were identified as interacting specifically with the protein.[34] Molecules active in all
373 experiments were considered *bona fide* hits, while those giving a positive response
374 in one or two experiments were considered unconfirmed hits because changes in
375 NMR signal are not necessarily related to binding. All NMR experiments were
376 performed on a BrukerAvIII HD 600 MHz NMR spectrometer at 298K; pulse
377 sequences included an excitation sculpting module in order to suppress bulk
378 water. Samples contained 500 μM ligand and 10 μM Hsp90α in 20mM tris pH 7.5,
379 50mM NaCl 1mM freshly prepared DTT and contained 10% D_2O.
380
381 **X-Ray crystallographic studies**
382 Protein was produced and crystallized as previously described.[35] For the
383 successful crystals, data were collected at 100K on an in-house Bruker D8 Venture
384 TXS Generator with a Bruker Photo 100 detector and were subsequently
385 processed using SAINT & SADABS. The crystals belong to the space groups I222.

386 The structures were solved by molecular replacement using a previously solved
387 Hsp90α protein model (PDB code: 1UY6; PU3 ligand and solvent removed) and the
388 program AMoRe.[36] Twenty cycles of rigid-body then restrained refinement were
389 carried out using the refinement program REFMAC5[37] followed by model building
390 and solvent addition using the molecular graphics program COOT.[38] The progress
391 of the refinement was assessed using R_{free} and the conventional R factor. Once
392 refinement was completed the structures were validated using various programs
393 from the CCP4i package.[39] Full data collection and refinement statistics are
394 presented in Supplementary Table 4.
395
396 Methodological details concerning the creation of the datasets, molecular docking,
397 MMPBSA and MMGBSA calculations, and surface plasmon resonance experiments
398 are provided as Supplementary Information.
399
400
401
402

ACKNOWLEDGEMENTS

We thank Carles Galdeano for helpful discussions and manuscript revision. We thank the Barcelona Supercomputing Center for access to computational resources. This work was financed by the Spanish Ministerio de Economia (SAF2012-33481, SAF2015-68749-R), the Catalan government (2014 SGR 1189) and the ICREA Academia program (FJL).

Correspondence and requests for materials should be addressed to X. B.

Author contributions

F.J.L, R.H. and X.B. conceived the overall strategy of the study. P.S. and X.B. conceived and implemented the DUck approach. S.R.C. and X.B. carried out and analysed all computational work. B.D. performed NMR experiments. L.B. performed X-ray crystallography experiments. N.M. performed SPR experiments. All authors discussed the results, designed experiments and wrote the paper.

Code Availability statement

A step-by-step guide on how to prepare, execute and analyse DUck simulations, along with scripts used to automate the process can be found on the following website: http://www.ub.edu/bl/undocking

REFERENCES

1. Kuhnert, M. *et al.* Tracing Binding Modes in Hit-to-Lead Optimization: Chameleon-Like Poses of Aspartic Protease Inhibitors. *Angew. Chem. Int. Ed. Engl.* **54,** 2849–2853 (2015).
2. Krohn, A., Redshaw, S., Ritchie, J. C., Graves, B. J. & Hatada, M. H. Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. *J. Med. Chem.* **34,** 3340–3342 (1991).
3. Smith, L. J., Gunsteren, W. F. Van & Allison, J. R. Multiple binding modes for palmitate to barley lipid transfer protein facilitated by the presence of proline 12. *Protein Sci.* **22,** 56–64 (2013).
4. Bissantz, C., Kuhn, B. & Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* 5061–5084 (2010). doi:10.1021/jm100112j
5. Klebe, G. Applying thermodynamic profiling in lead finding and optimization. *Nat. Rev. Drug Discov.* **14,** 95–110 (2015).

6. Ferenczy, G. G. & Keserű, G. M. Thermodynamics of fragment binding. *J. Chem. Inf. Model.* **52,** 1039–45 (2012).

7. Kozakov, D. *et al.* Ligand deconstruction: Why some fragment binding positions are conserved and others are not. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E2585-2594 (2015).

8. Schmidtke, P., Luque, F. J., Murray, J. B. & Barril, X. Shielded hydrogen bonds as structural determinants of binding kinetics: application in drug design. *J. Am. Chem. Soc.* **133,** 18903–10 (2011).

9. Colizzi, F., Perozzo, R., Scapozza, L., Recanatini, M. & Cavalli, A. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J. Am. Chem. Soc.* **132,** 7361–71 (2010).

10. Noble, M. E. M., Endicott, J. a & Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Science* **303,** 1800–5 (2004).

11. Anderson, D. R. *et al.* Benzothiophene inhibitors of MK2. Part 1: structure-activity relationships, assessments of selectivity and cellular potency. *Bioorg. Med. Chem. Lett.* **19,** 4878–81 (2009).

12. Filippakopoulos, P. & Knapp, S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discov.* **13,** 337–356 (2014).

13. Ember, S. W. J. *et al.* Acetyl-lysine binding site of bromodomain-containing protein 4 (BRD4) interacts with diverse kinase inhibitors. *ACS Chem. Biol.* **9,** 1160–1171 (2014).

14. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* **55,** 6582–6594 (2012).

15. Rask-Andersen, M., Almén, M. S. & Schiöth, H. B. Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov.* **10,** 579–90 (2011).

16. Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432,** 862–5 (2004).

17. Brooijmans, N. & Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **32,** 335–73 (2003).

18. Ruiz-Carmona, S. *et al.* rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **10,** e1003571 (2014).

19. Roughley, S., Wright, L., Brough, P., Massey, A. & Hubbard, R. E. Hsp90 inhibitors and drugs from fragment and virtual screening. *Top. Curr. Chem.* **317,** 61–82 (2012).

20. Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* **6,** 211–9 (2007).

21. Joseph-McCarthy, D., Campbell, A. J., Kern, G. & Moustakas, D. Fragment-based lead discovery and design. *J. Chem. Inf. Model.* **54,** 693–704 (2014).

22. Chen, I.-J. & Hubbard, R. E. Lessons for fragment library design: analysis of

484 output from multiple screening campaigns. *J. Comput. Aided. Mol. Des.* 603–
485 620 (2009). doi:10.1007/s10822-009-9280-5

486 23. Teotico, D. G. *et al.* Docking for fragment inhibitors of AmpC beta-lactamase.
487 *Proc. Natl. Acad. Sci. U. S. A.* **106,** 7455–7460 (2009).

488 24. Murray, C. W. *et al.* Fragment-based drug discovery applied to Hsp90.
489 Discovery of two lead series with high ligand efficiency. *J. Med. Chem.* **53,**
490 5942–55 (2010).

491 25. Chipot, C. & Pohorille, A. *Free Energy Calculations Theory and Applications in*
492 *Chemistry and Biology*. *Free Energy Calculations Theory and Applications in*
493 *Chemistry and Biology* (Springer Berlin Heidelberg, 2007).

494 26. Alvarez-Garcia, D. & Barril, X. Molecular simulations with solvent
495 competition quantify water displaceability and provide accurate interaction
496 maps of protein binding sites. *J. Med. Chem.* **57,** 8530–9 (2014).

497 27. Chemical Computing Group Inc. Molecular Operating Environment (MOE),
498 2014.09. (2015).

499 28. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality
500 atomic charges. AM1-BCC model: II. Parameterization and validation. *J.*
501 *Comput. Chem.* **23,** 1623–41 (2002).

502 29. Bayly, C. I., McKay, D. & Truchon, J. F. An Informal AMBER Small Molecule
503 Force Field: parm@Frosst. (2011).

504 30. Case, D. A. *et al.* AMBER 12. (2012).

505 31. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the
506 cartesian equations of motion of a system with constraints: molecular
507 dynamics of n-alkanes. *J. Comput. Phys.* **23,** 327–341 (1977).

508 32. Brough, P. a *et al.* Combining hit identification strategies: fragment-based
509 and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine
510 inhibitors of the Hsp90 molecular chaperone. *J. Med. Chem.* **52,** 4794–809
511 (2009).

512 33. Baurin, N. *et al.* Design and characterization of libraries of molecular
513 fragments for use in NMR screening against protein targets. *J. Chem. Inf.*
514 *Comput. Sci.* **44,** 2157–2166 (2004).

515 34. Davis, B. in *Protein-Ligand Interactions: Methods and Applications* (eds.
516 Williams, M. A. & Daviter, T.) **1008,** 389–413 (Springer Science+Business
517 Media, 2013).

518 35. Wright, L. *et al.* Structure-activity relationships in purine-based inhibitor
519 binding to HSP90 isoforms. *Chem. Biol.* **11,** 775–85 (2004).

520 36. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta*
521 *Crystallogr. Sect. A Found. Crystallogr.* **50,** 157–163 (1994).

522 37. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular
523 structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol.*
524 *Crystallogr.* **53,** 240–255 (1997).

38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

39. Collaborative, C. P. & others. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50,** 760 (1994).

530 **TABLES**
531
532 **Table 1.** Summary of results for the 9 Hsp90 NMR Class 1 hits. Chemical structures
533 of all compounds are shown in Supplementary Table 1.
534

ID	MW	Docking	DUck	SPR Kd (mM)	PDB Sim[b]	ChEMBL Sim.[b]
		Score (Rank[a])	Score (Rank[a])			
1*	248.7	-25.0 (79)	9.1 (10)	77	2XDX (0.37)	CHEMBL 1340447 (0.44)
2*	221.3	-25.0 (73)	8.2 (11)	320	2WI6 (0.29)	CHEMBL 1536318 (0.54)
3*	230.2	-26.7 (19)	11.3 (1)	700	4EFU (0.32)	CHEMBL 1458840 (0.51)
4	240.3	-26.4 (22)	7.4 (16)	730	3WHA (0.29)	CHEMBL 1542436 (0.37)
5	165.2	-23.8 (128)	8.1 (12)	-	4EFT (0.27)	CHEMBL 1313412 (0.28)
6	206.3	-23.3 (138)	9.5 (5)	-	3HHU (0.42)	CHEMBL 2103879 (0.42)
7	236.3	-25.4 (51)	7.8 (15)	-	3B24 (0.31)	CHEMBL 1375884 (0.36)
8	224.7	-25.3 (58)	7.0 (22)	-	2XDX (0.35)	CHEMBL 1383799 (0.37)
22	237.3	-28.3 (2)	5.6 (33)	-	3O0I (0.27)	CHEMBL 1834092 (0.33)

535
536 *Xray structure solved [a]Position within the list
537 of 149 molecules that were evaluated with
538 DUck. [b]Hsp90 structure in the PDB or
539 compound with Hsp90 activity in ChEMBL (as
540 of 23/03/2016) with the closest similarity to
541 the fragment hit. Similarity (values in
542 parentheses) was calculated with Open Babel
543 using the FP2 fingerprint.
544
545

546 **FIGURE CAPTIONS**
547

548

549 **Figure 1.** Calculation of W_{QB}. **a**. The receptor is idealized as a model system
550 containing only the local environment around a key intermolecular hydrogen
551 bond. **b**. Representative work profiles obtained from dynamic undocking
552 simulations for a strong (black) and a weak (grey) ligand. The quasi-bound state is
553 defined as the point with the highest energy relative to the ideal hydrogen bond
554 geometry.

555
556

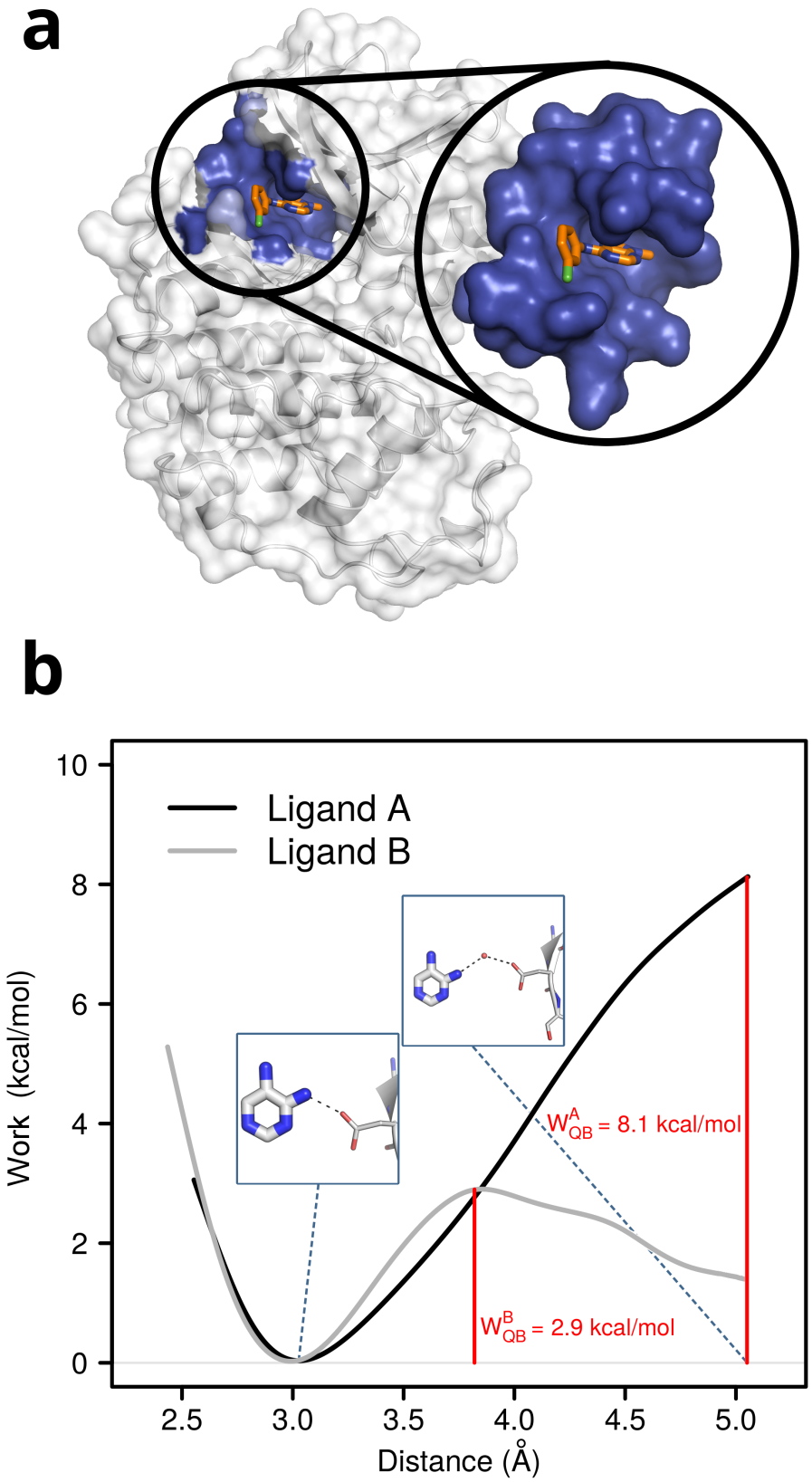
557 **Figure 2.** Application of the quasi-bound approximation to ligand ranking. **a.**
558 Distribution of W_{QB} values of potent CDK2 ligands (IC_{50} < 1μM; dark grey), weak
559 CDK2 ligands (IC_{50} > 1μM; light grey) and non-binding decoys (black). Points
560 indicate population values, from which the smooth lines are extrapolated. **b.** ROC
561 curves for the CDK2 (black), A2AR (red) and Trypsin (green) DUD sets. Plotted
562 results correspond to 2 DUck runs per ligand. AUC values are shown in
563 Supplementary Figure 6. **c.** Docking score vs. W_{QB} values for active (red) and
564 inactive (black or gray) compounds in the CDK2 retrospective virtual screening
565 dataset. The quadrant in orange highlights the area corresponding to top 25%
566 docking score and top 25% W_{QB} values, where optimal enrichment factors (EF) are
567 achieved. **d.** For the same set, distribution of W_{QB} values for the active compounds
568 (red), all decoys (black) and decoys in the top 25% docking score (gray). **e.**
569 Distribution of W_{QB} values of CDK2 actives (red) and decoys (gray) ranked in the
570 top 25% by two independent docking programs (rDock and Glide). **f.** Distribution
571 of W_{QB} values of CDK2 actives (red) and decoys (gray) ranked in the top 25% both
572 by MMPBSA and the rDock docking program.

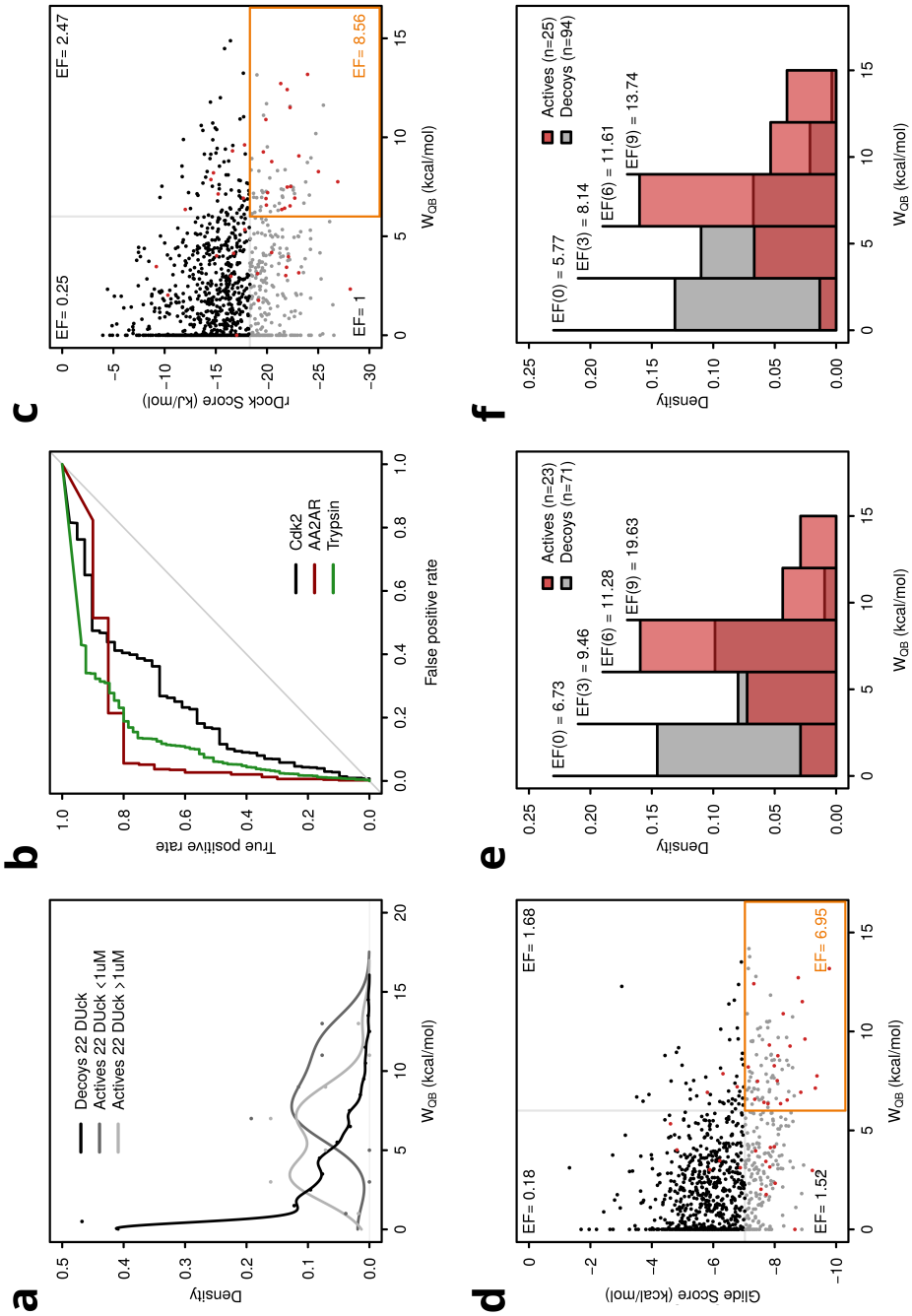
573

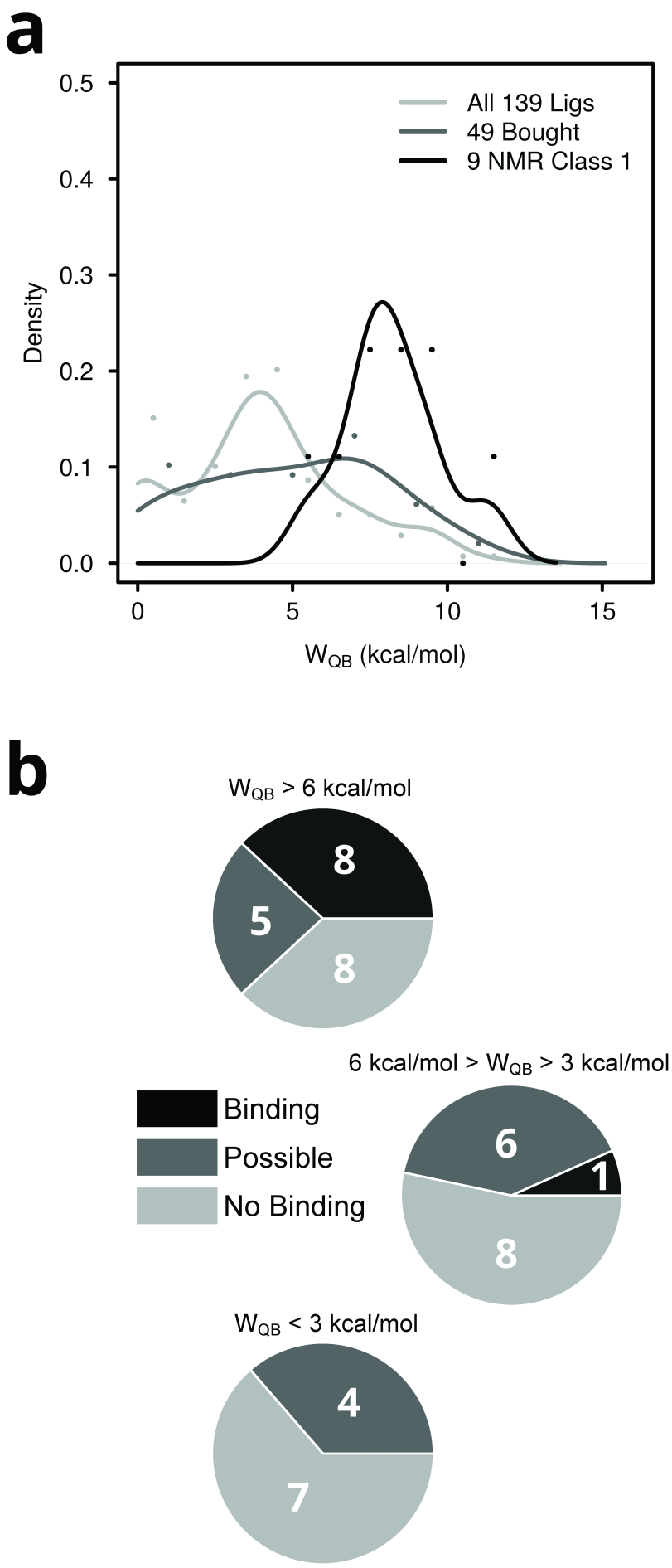
574
575 **Figure 3.** Additional analyses of the prospective application of DUck in Hsp90. **a**.
576 Distribution of W_{QB} values for 139 top docking scorers (pale gray), 47 compounds
577 within this set that were purchased (dark gray), and the 9 compounds detected as
578 active. **b**. Pie charts showing the hit rates for the set of compounds with high W_{QB}
579 (top), medium W_{QB} (middle) and low W_{QB} (bottom). The area in black corresponds
580 to bona fide hits, dark gray represents compounds that give a positive signal in 1
581 or 2 NMR experiments, pale gray corresponds to inactive compounds. Labels
582 indicate the number of compounds of each class. Chemical structures are shown in
583 Supplementary Figure 12.

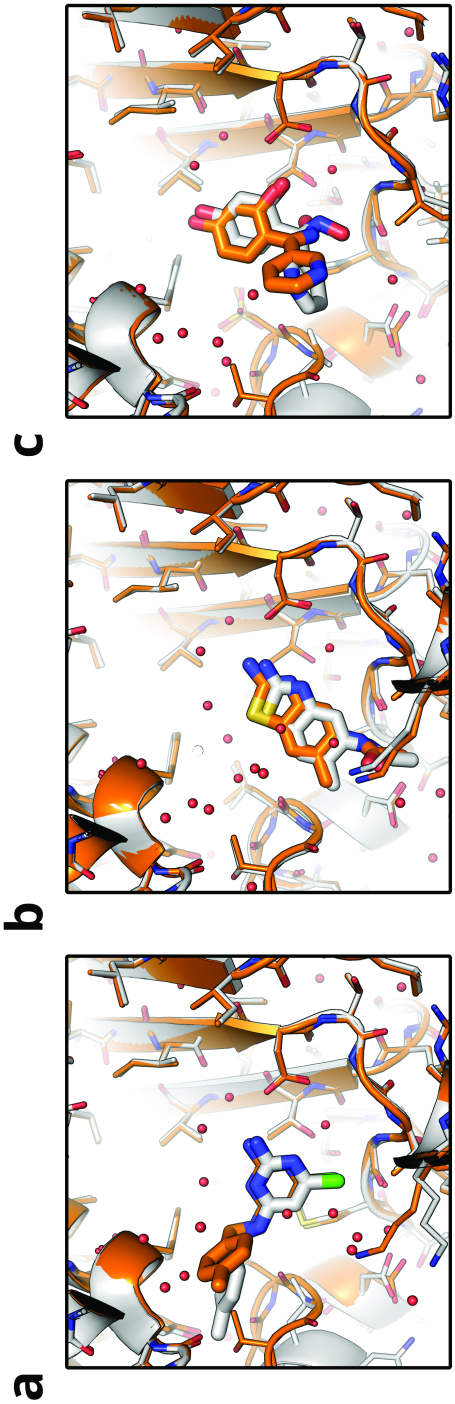
584

Figure 4. Experimental (grey) and predicted (orange) binding modes of the fragment hits. **a.** Compound **1**, the RMSD of the whole molecule is 2.58 Å due to a conformational change of the protein next to the p-toluene ring. The pyridine ring and bonded atoms, where the key interaction occur, have a RMSD of 0.54 Å **b.** Compound **2** has a RMSD of 0.54 Å **c.** Compound **3** has a RMSD of 1.55 Å, all hydrogen bond interactions are preserved.


a
b
Ligand A
Ligand B
Work (kcal/mol)
Distance (Å)
0
2
4
6
8
10
2.5
3.0
3.5
4.0
4.5
5.0
W^A_{QB} = 8.1 kcal/mol
W^B_{QB} = 2.9 kcal/mol


a
Decoys 22 DUck
Actives 22 DUck <1uM
Actives 22 DUck >1uM
Density
W_{QB} (kcal/mol)
b
Cdk2
AA2AR
Trypsin
True positive rate
False positive rate
c
EF= 0.25
EF= 2.47
EF= 1
EF= 8.56
rDock Score (kJ/mol)
W_{QB} (kcal/mol)
d
EF= 0.18
EF= 1.68
EF= 1.52
EF= 6.95
Glide Score (kcal/mol)
W_{QB} (kcal/mol)
e
Actives (n=23)
Decoys (n=71)
EF(0) = 6.73
EF(3) = 9.46
EF(6) = 11.28
EF(9) = 19.63
Density
W_{QB} (kcal/mol)
f
Actives (n=25)
Decoys (n=94)
EF(0) = 5.77
EF(3) = 8.14
EF(6) = 11.61
EF(9) = 13.74
Density
W_{QB} (kcal/mol)


a
All 139 Ligs
49 Bought
9 NMR Class 1
Density
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
W_QB (kcal/mol)
b
W_QB > 6 kcal/mol
8
5
8
6 kcal/mol > W_QB > 3 kcal/mol
6
1
8
W_QB < 3 kcal/mol
4
7
Binding
Possible
No Binding


a
b
c

Supplementary Information

Dynamic Undocking and the Quasi-Bound State as Tools for Drug Design

Sergio Ruiz-Carmona, Peter Schmidtke, F. Javier Luque, Lisa Baker, Natalia Matassova, Ben Davis, Stephen Roughley, James Murray, Rod Hubbard, Xavier Barril

Supplementary Information – Dynamic Undocking and the Quasi-Bound State

INDEX

SUPPLEMENTARY METHODS

Datasets

When possible, datasets were geared towards fragment-sized ligands because they present more scaffold diversity, make fewer peripheral interactions that could mask the main interactions and because Fragment-Based Drug Discovery (FBDD) approaches are increasingly important as hit identification strategy.[1,2] For CDK2, all ligands with molecular weight below 300 Da and known binding affinity (IC50) were extracted from the PDB.[3] To increase the diversity of the dataset, all ligands were clustered at 75% similarity using the MACCS fingerprints as implemented in MOE (Chemical Computing Group Inc., 2015) and only the centroids were used to define the active set. The composition of the dataset is described in Supplementary Table 5. It should be noted that this is a noisy dataset because data sources are very heterogeneous and IC_{50} values have an indirect relationship with dissociation constants.[4] As such, it should only be used to detect trends. In order to assess the significance of the correlation, we have also investigated the correlation between IC_{50} and molecular weight (Supplementary Figure 20). For retrospective VS experiments, a pool of 30 decoys per active fragment was obtained with the DUD-E decoy generator,[5] which puts together a set of putatively inactive molecules with physicochemical properties very similar to active ones. For BRD4, as it was designed to study the correlation between experimental binding affinity and W_{QB}, only the ligands with known binding mode and measured IC_{50} or K_D were considered (relationship with molecular weight reported in Supplementary Figure 21). The crystal structure of each ligand-protein complex was obtained from PDB and used as input for subsequent calculations. The composition of the dataset is described in Supplementary Table 6. In the case of AA2AR, as there are few structures in the PDB, the active fragments were taken from the DUD-E benchmark set.[5] The rest of the procedure is the same as described for CDK2. For Trypsin, we found that few ligands have a low molecular weight so we did not filter by size. Instead, a random subset of 2000 actives and decoys was selected from DUD-E. In the case of Hsp90, all candidate molecules originate from a unified collection generated in house from the commercial libraries of five preferred vendors (Specs, Enamine, Life Chemicals, Princeton Biomoleculars and Asinex). In this case we set an upper limit of 250 Da, obtaining 280000 candidate fragments. All ligands were

prepared for docking using Schrödinger's Ligprep[6]with the following options different than default: neutralize and ionize at pH 7 with a threshold of +- 1 with a maximum of 6 tautomers and 8 stereoisomers generated.

Molecular Docking with rDock

For CDK2, AA2AR and Trypsin, the 3D structure used to define the receptor was obtained from the DUD-E benchmark set.[5] MOE[7] was used to generate mol2 files that can be read by rDock, our docking engine.[8] For Hsp90, we use the same cavity definition and docking protocol described previously.[8] In all systems, pharmacophoric restraints were used to ensure that the key interaction point was matched by every molecule in the dataset, as defined in Supplementary Table 3. rDock was run with the default parameters for standard docking. 50 individual docking processes were executed per ligand, thus ensuring that the lowest-energy binding mode is identified. The best-scoring solution is accepted as the putative binding mode. Ligands that do not fulfill the pharmacophore are identified by the restraint penalty and eliminated from the dataset (i.e. not considered in the ROC curves or any other analysis).

Molecular Docking with Glide

In order to demonstrate that our methodology provides an advantage regardless of the docking program used, we also run the CDK2 system with Glide.[9] The generation of the cavity with Glide was performed using coordinates defined as in rDock docking and default parameters. Pharmacophoric restraints were defined to force all ligands to make a hydrogen bond with Leu77:N acting as donor. Glide docking was run with default parameters (Supplementary Figures 22, 23 and 24). The best docking pose for each ligand was selected and used as input for DUck.

MMGBSA and MMPBSA

MMGBSA and MMPBSA calculations using AMBER12 software were also performed and compared against the rest of methods. Each ligand was simulated for 5 ns with the full size receptor of CDK2 using the same MD configuration defined in the section above (Supplementary Figures 24 and 25). For each simulation, a total of 25 snapshots separated by 200 ps were used and the free

energies were averaged over the ensemble of conformations. All the calculations were performed with default parameters with the exception of the following: the GB model used is one of the developed by Onufriev et al.[10] (igb=2) and the atomic radii are set up according to the topology (radiopt=0).

Surface plasmon resonance

Surface plasmon resonance (SPR) experiments have been done mainly as described before.[11,12] All measurements were performed on a Biacore T200 instrument (Biacore GE Healthcare) at 20°C on Series S NTA chips. 25 mM HEPES pH7.4, 175 mM NaCl, 0.01% P-20, 0.025mM EDTA and 1% DMSO was used as a running buffer. HSP90 protein was produced as described previously. Chip surface was generated with multi-His-tagged Hsp90 protein as in reference.[11] The sensor surface was regenerated by 0.35 M EDTA and 45% DMSO with additional 60 sec injections of 0.1 mg/mL trypsin and 0.5 M imidazole. In some experiments, the protein was further stabilized on NTA surface by covalent amine coupling as advised by manufacturer. Screening of fragments was conducted in dose response titrations of nine two-fold diluted experimental points with the top concentration of 500 μM. Each fragment has been tested at least three times. Data processing was performed using BIAevaluation 2.1 (Biacore GE Healthcare Bio-SciencesCorp) or Scrubber2 (BioLogic) software. Sensorgrams were double referenced prior to global fitting of the concentration series to a Steady State Affinity model. Representative sensorgrams are shown in Supplementary Figure 26.

SUPPLEMENTARY REFERENCES

1. Joseph-McCarthy, D., Campbell, A. J., Kern, G. & Moustakas, D. Fragment-based lead discovery and design. *J. Chem. Inf. Model.* **54,** 693–704 (2014).

2. Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* **6,** 211–9 (2007).

3. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

4. Klebe, G. Applying thermodynamic profiling in lead finding and optimization. *Nat. Rev. Drug Discov.* **14,** 95–110 (2015).

5. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* **55,** 6582–6594 (2012).

6. Schrödinger. LigPrep, version 2.9. *Schrödinger, LLC* (2014).

7. Chemical Computing Group Inc. Molecular Operating Environment (MOE), 2014.09. (2015).

8. Ruiz-Carmona, S. *et al.* rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **10,** e1003571 (2014).

9. Friesner, R. a *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47,** 1739–49 (2004).

10. Onufriev, A., Bashford, D. & Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins Struct. Funct. Genet.* **55,** 383–394 (2004).

11. Meiby, E. *et al.* Fragment screening by weak affinity chromatography: comparison with established techniques for screening against HSP90. *Anal. Chem.* **85,** 6756–66 (2013).

12. Murray, J. B., Roughley, S. D., Matassova, N. & Brough, P. a. Off-rate screening (ORS) by surface plasmon resonance. An efficient method to kinetically sample hit to lead chemical space from unpurified reaction products. *J. Med. Chem.* **57,** 2845–2850 (2014).

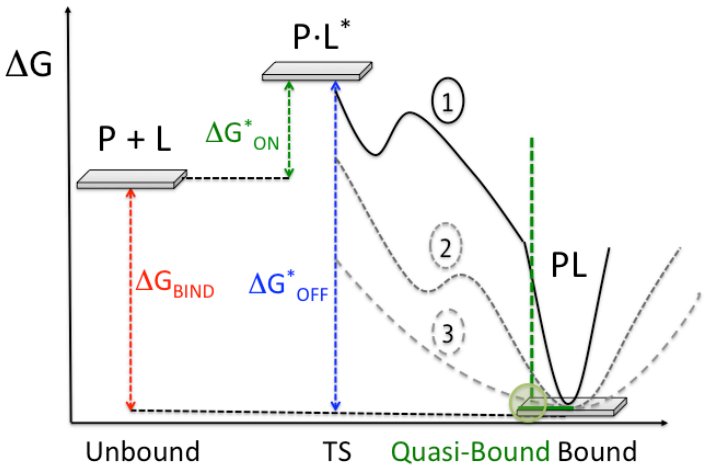
13. Wright, L. *et al.* Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms. *Chem. Biol.* **11,** 775–85 (2004).

14. Anderson, D. R. *et al.* Benzothiophene inhibitors of MK2. Part 1: structure-activity relationships, assessments of selectivity and cellular potency. *Bioorg. Med. Chem. Lett.* **19,** 4878–81 (2009).

SUPPLEMENTARY FIGURES

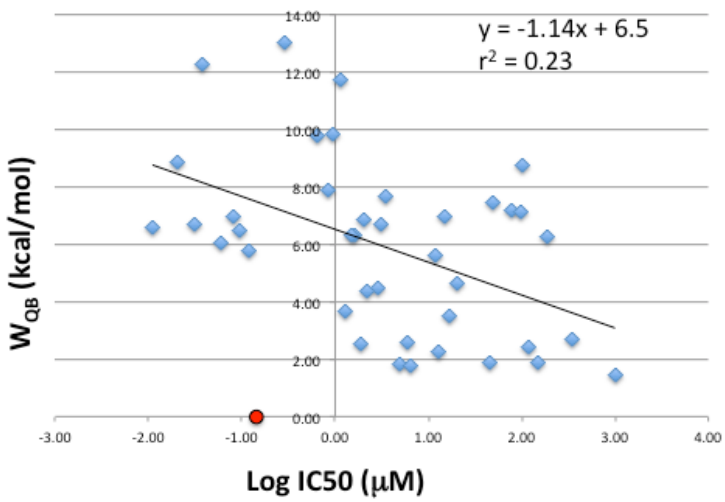
Supplementary Figure 1

Graphical representation of the quasi-bound state in relation to the dissociation process. The macroscopic constants describing the behavior of a non-covalent complex are determined by the relative free energies of three states (bound, transition state and unbound). States in-between are theoretically irrelevant, so molecules 1, 2 and 3 would have the same kinetic and thermodynamic constants. The Quasi-bound state is merely designed to probe the slope around the bound state, obtaining an approximation to the structural stability of the binding mode. We find that true ligands are more likely to have a profile like 1, whereas many decoys have profiles similar to 2 or 3.



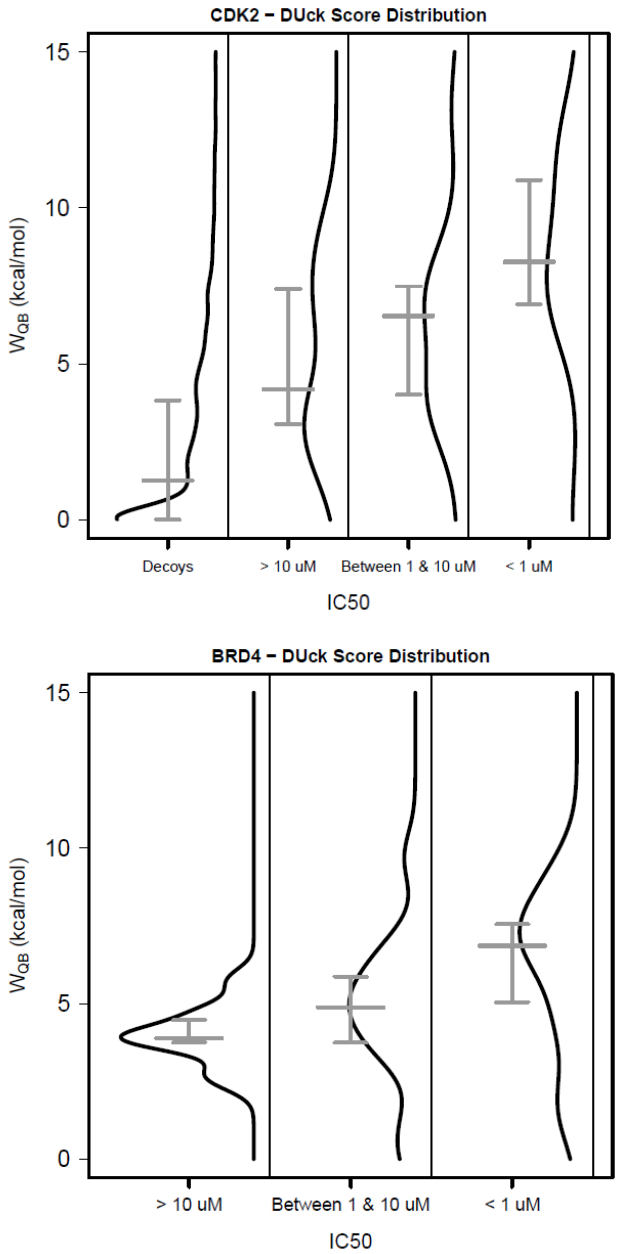
Supplementary Figure 2

W_{QB} values vs. experimentally determined activities (expressed as Log(IC_{50})), for a set of 41 Fragment-like CDK2 ligands taken from the PDB. Ligand 3FZ1 is shown in red and not included in the correlation. As shown below, this ligand does not fulfill the condition of using the hinge region as attachment point.



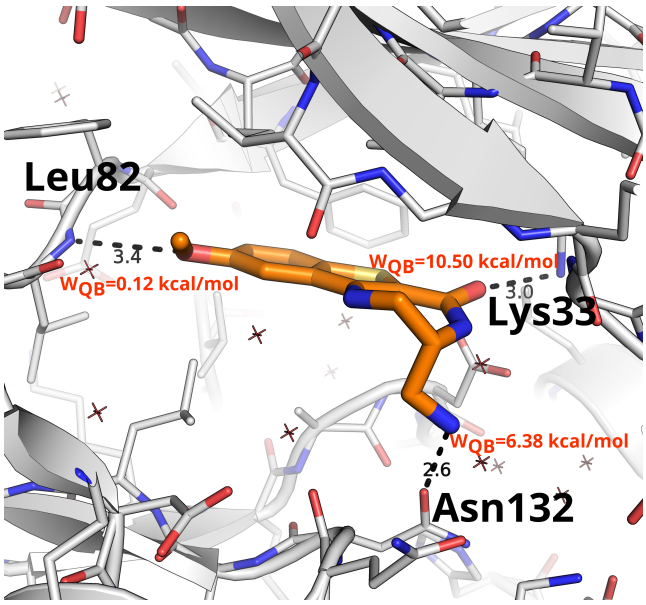
Supplementary Figure 3

Distribution of W_{QB} values as a function of binding affinity (IC_{50}), for the CDK2 (top) and BRD4 set (bottom). Compounds with the same binding affinity present a wide distribution of W_{QB} values, but there is a tendency towards higher values for more potent compounds. Most notably, very low W_{QB} values are rare for potent ligands.



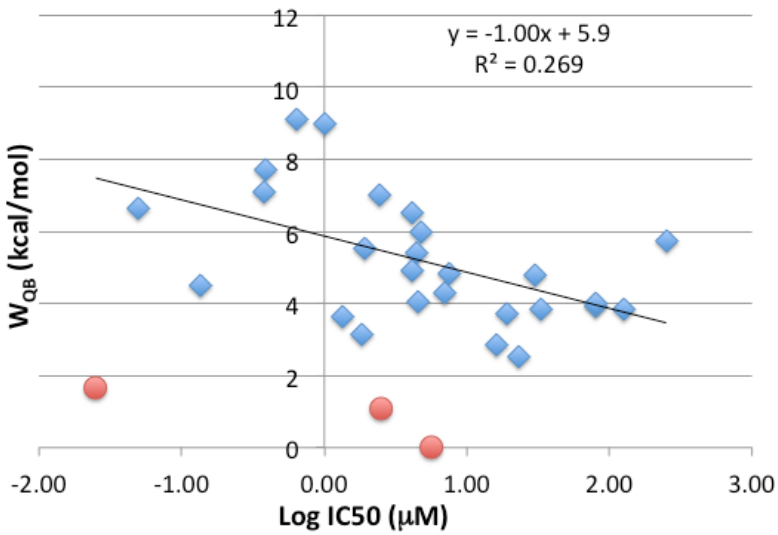
Supplementary Figure 4

Binding mode of ligand in PDB structure 3FZ1. This ligand is unusual because its interaction with the hinge region is labile. Structural and SAR data confirms that this interaction is not important for potency.[14] Instead, this ligand forms two charge-reinforced hydrogen bonds with Nζ of Lys33 and Oδ1 of Asn132, from which it draws structural stability. Note that the IC_{50} reported in the PDB for this compound is wrong. The correct value is 146 nM.[14]



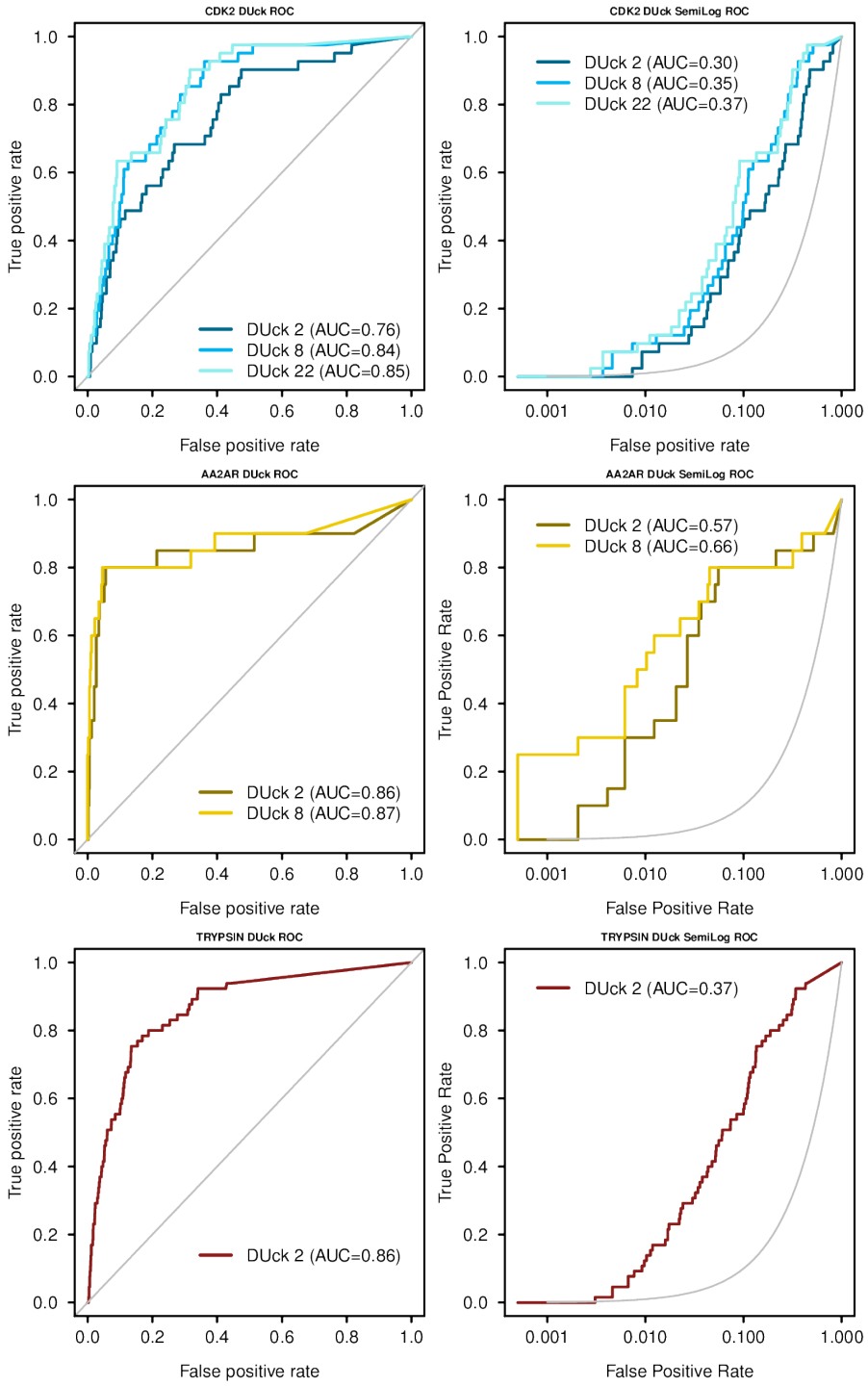
Supplementary Figure 5

W_{QB} values vs. experimentally determined activities (expressed as Log(IC_{50})), for a set of 30 BRD4 ligands taken from the PDB. The points in red have not been included in the correlation. They correspond to three kinase inhibitors that bind to BRD4 as an unintended secondary target and present extremely low resistance to breaking the interaction with Nδ2 of Asn120 (PDB codes 4O74, 4O77 & 4O7E).

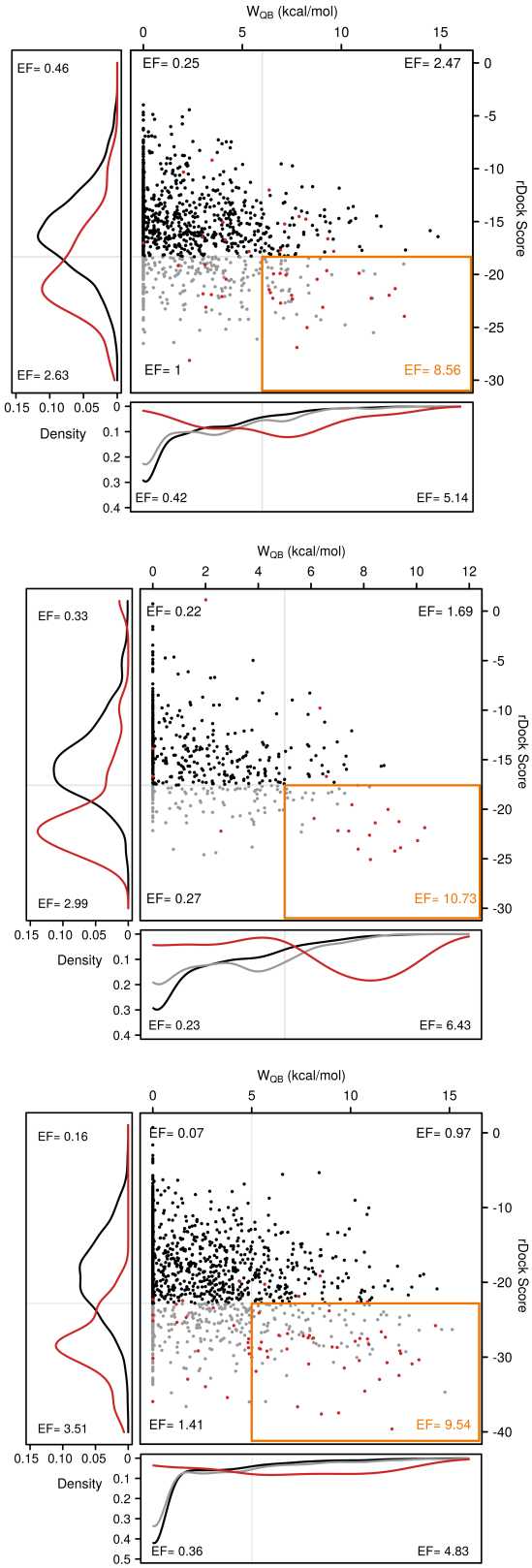


Supplementary Figure 6

ROC curves (left) and semilog-ROC curves (right) of the retrospective virtual screening experiments on CDK2 (top), AA2R (middle) and Trypsin (bottom). The grey line indicates the baseline (random selection). For CDK2, the results corresponding to 2, 8 and 22 DUck runs are reported. For AA2R, the results corresponding to 2, and 8 DUck runs are reported. For Trypsin, only 2 DUck runs were executed. AUC values are inset in the plots.

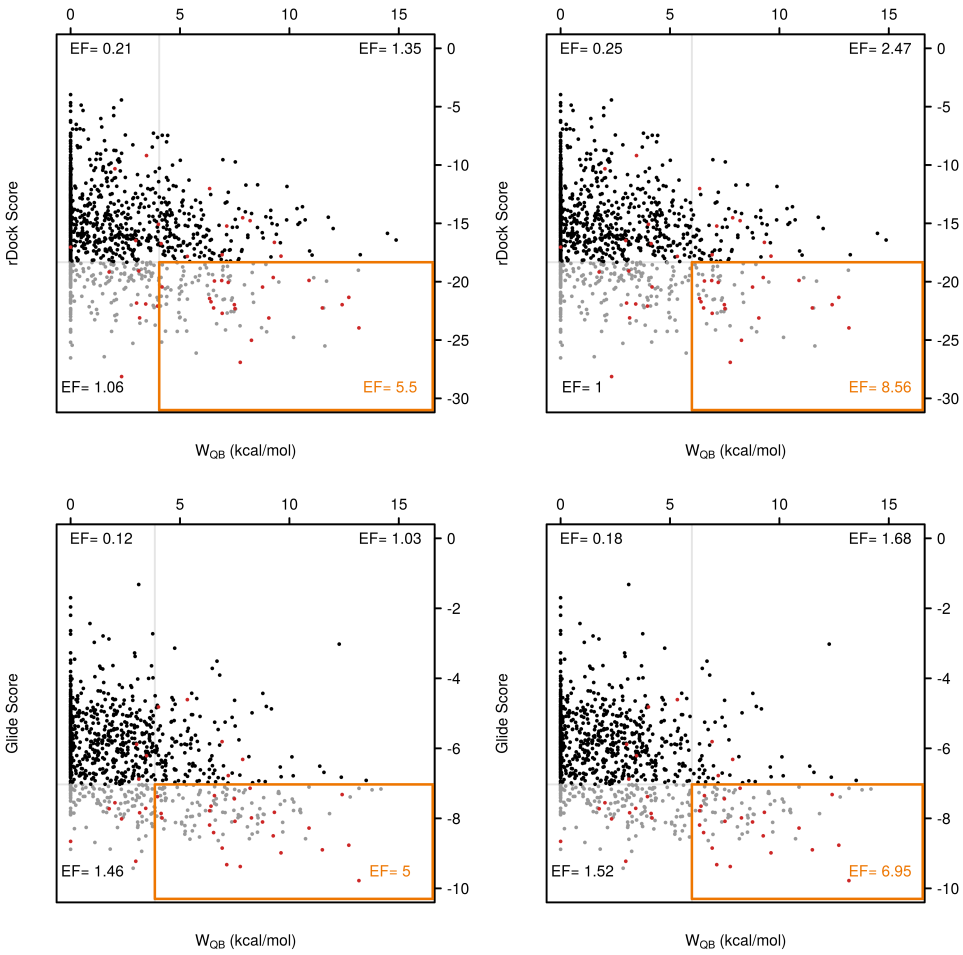


Supplementary Figure 7
Docking (rDock) score vs. W_{QB} values for active (red) and inactive compounds (black or gray) in the retrospective virtual screening datasets for CDK2 (top), AA2AR (middle)and Trypsin (bottom). The side panels show the distribution of active (red) and inactive (black) compounds for each individual method (docking to the left, DUck at the bottom). Gray points (central panel) and gray line (inferior panel) represent the decoys with a docking score within the top 25%. The orange square highlights the area corresponding to top 25% docking score and top 25% W_{QB} values, where optimal enrichment factors (EF) are achieved.



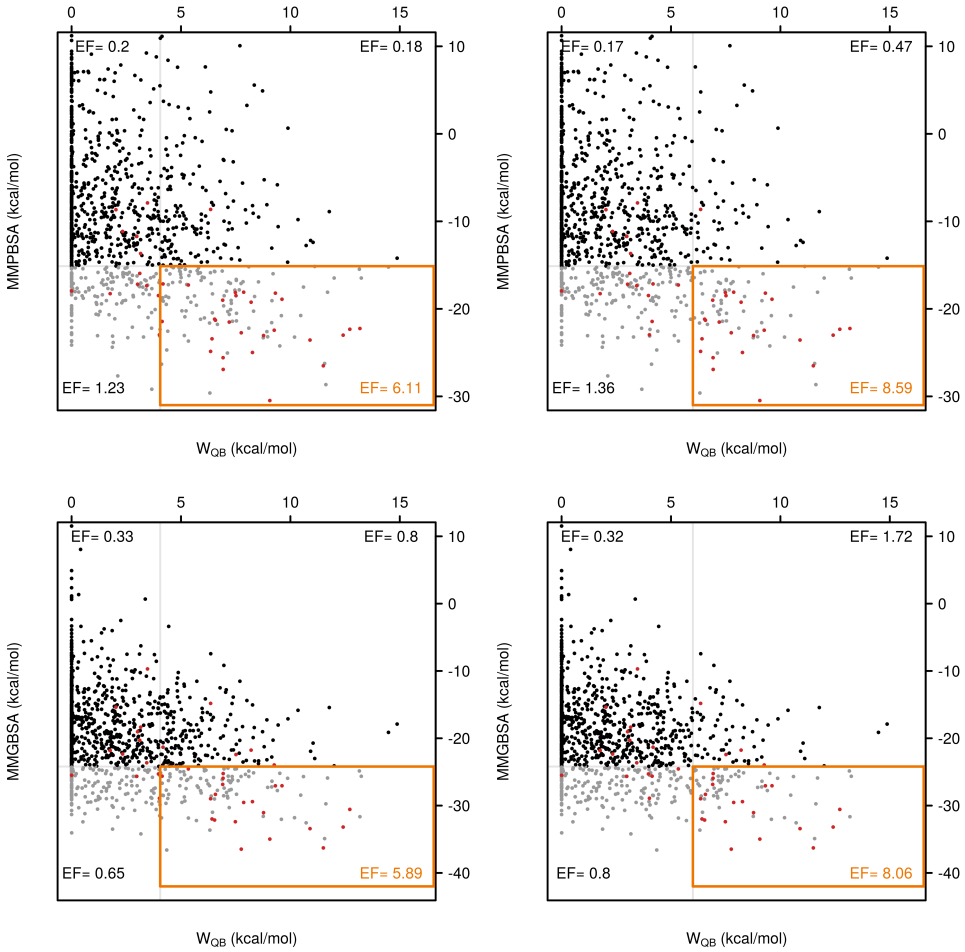
Supplementary Figure 8

Docking score vs. W_{QB} obtained for two different programs on the CDK2 test set. Each molecule was docked with rDock (top) or Glide (bottom) and the binding mode generated by each program was used as starting geometry for DUck simulations. In both cases, docking scores are orthogonal to W_{QB} and a high proportion of good scorers have very low W_{QB} values. The intersection between methods defines a subset highly enriched in active molecules. Two intersecting levels are presented per program:top25% docking+ top 25% DUck (left);and top25% docking + top 12% DUck (right).



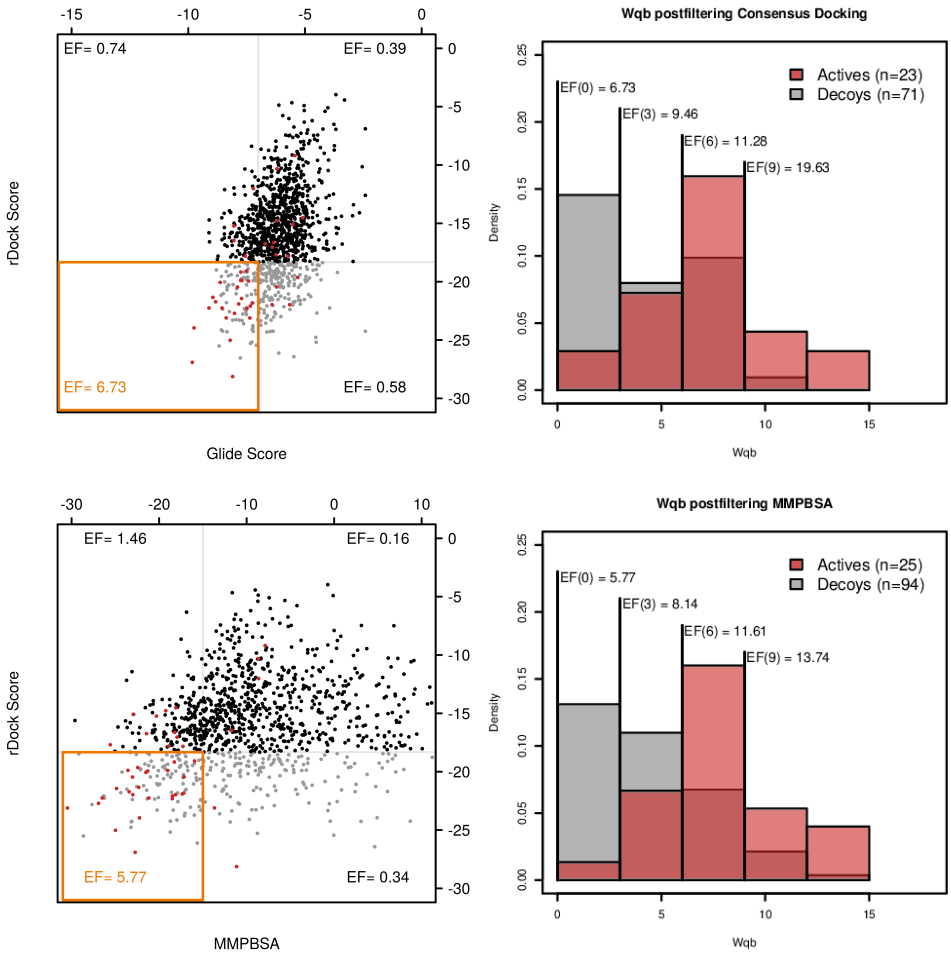
Supplementary Figure 9

MMPBSA and MMGBSA-calculated ΔG_{bind} vs. W_{QB} on the CDK2 test set. The rDock-generated binding mode was used as starting point for molecular dynamics simulations, which where then processed to obtain MMPBSA and MMGBSA binding free energies. In both cases, the calculated ΔG_{bind} values are orthogonal to W_{QB} and a high proportion of good MM(PB/GB)SA scorers have very low W_{QB} values. The intersection between methods defines a subset highly enriched in active molecules. Two intersecting levels are presented per method: top25% MM(PB/GB)SA + top 25% DUck (left); and top25% MM(PB/GB)SA + top 12% DUck (right).



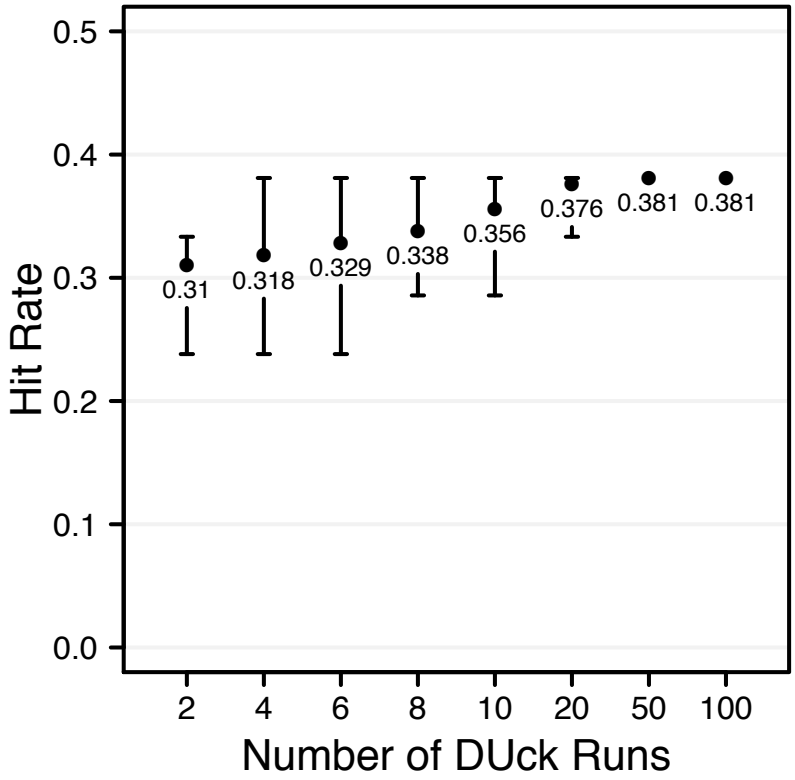
Supplementary Figure 10

Filtering by W_{QB} increases performance even after consensus scoring (CDK2 test set). The left panels show a scatter plot of rDock score vs. Glide score (top) and rDock score vs. MMPBSA-calculated ΔG_{bind} (bottom). Molecules ranked in the top 25% by both methods (highlighted area) are then binned according to their W_{QB} (right panels, also shown in the main text). Filtering by W_{QB} would increase the enrichment factor in a cut-off dependent manner.



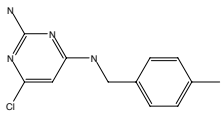
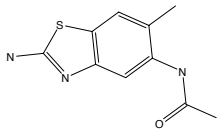
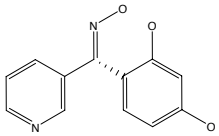
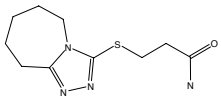
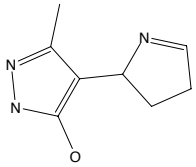
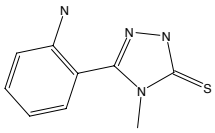
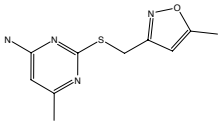
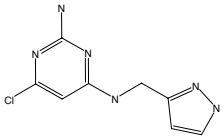
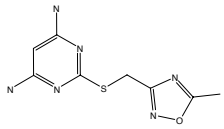
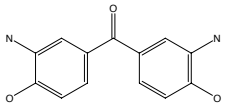
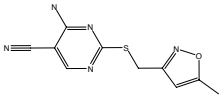
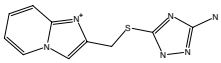
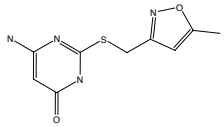
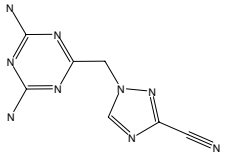
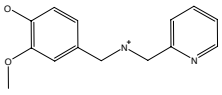
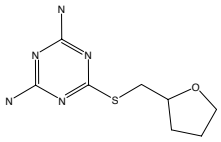
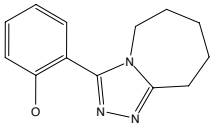
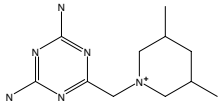
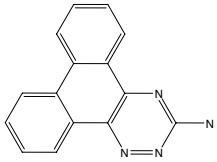
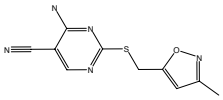
Supplementary Figure 11

Percentage of active molecules in the top 21 (out of 47 compounds tested) as a function of the number of DUck runs. At the screening stage we carried out 100 DUck simulations per ligand, obtaining a hit rate of 38%. Retrospectively, we took 50 random combinations of N={2,4,6,8,10,20,50} DUck runs and calculated the hit rates that would have been obtained. Averages are represented as filled circles and labeled with their actual values. The bars span from the maximum to the minimum values.



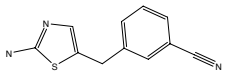
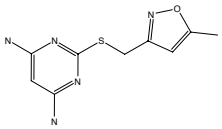
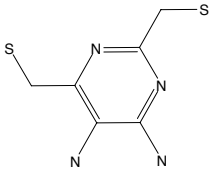
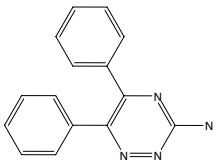
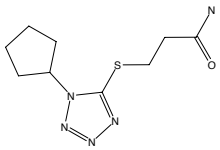
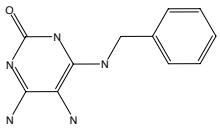
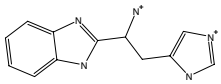
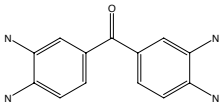
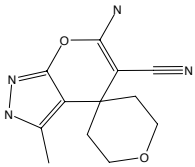
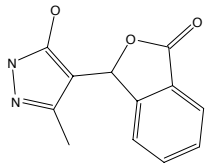
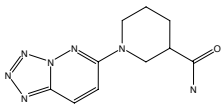
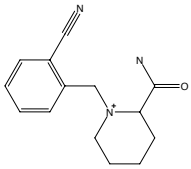
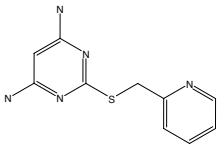
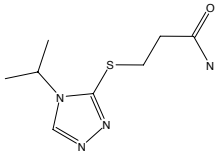
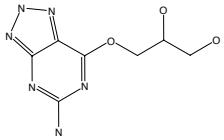
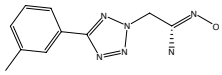
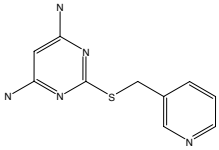
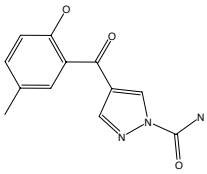
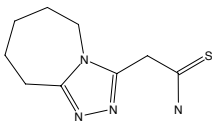
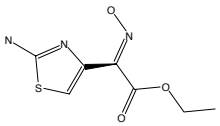
Supplementary Figure 12

Chemical structure of the tested compounds. Duck Class refers to strong, medium and weak binders (1, 2 & 3, respectively). NMR Class 1 are true binders. The rest are considered inactive. The real numbers correspond to rDock score (left) and W_{QB} (right).

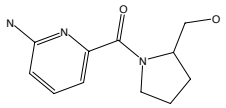
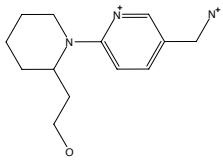
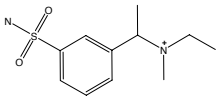
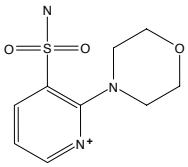
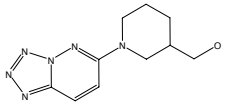
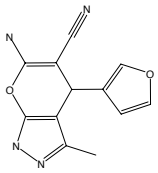
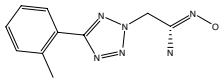
DUck Class: 1  **NMR Class:** 1 -24.9700 9.1000 ID: 1	**DUck Class:** 1  **NMR Class:** 1 -25.0300 8.2000 ID: 2	**DUck Class:** 1  **NMR Class:** 1 -26.6200 11.3000 ID: 3	**DUck Class:** 1  **NMR Class:** 1 -26.4500 7.4000 ID: 4
DUck Class: 1  **NMR Class:** 1 -23.7700 8.2000 ID: 5	**DUck Class:** 1  **NMR Class:** 1 -23.2600 9.5000 ID: 6	**DUck Class:** 1  **NMR Class:** 1 -25.4500 7.8000 ID: 7	**DUck Class:** 1  **NMR Class:** 1 -25.3500 7.0000 ID: 8
DUck Class: 1  **NMR Class:** 2 -28.0400 7.3000 ID: 9	**DUck Class:** 1  **NMR Class:** 2 -27.1200 6.4000 ID: 10	**DUck Class:** 1  **NMR Class:** 2 -26.1400 9.8000 ID: 11	**DUck Class:** 1  **NMR Class:** 3 -25.4000 7.0000 ID: 12
DUck Class: 1  **NMR Class:** 3 -25.0100 6.4000 ID: 13	**DUck Class:** 1  **NMR Class:** nb -26.0000 6.5000 ID: 14	**DUck Class:** 1  **NMR Class:** nb -25.9100 6.7000 ID: 15	**DUck Class:** 1  **NMR Class:** nb -24.7300 6.5000 ID: 16
DUck Class: 1  **NMR Class:** nb -24.4600 10.3000 ID: 17	**DUck Class:** 1  **NMR Class:** nb -28.4400 7.2000 ID: 18	**DUck Class:** 1  **NMR Class:** nb -25.9200 9.2000 ID: 19	**DUck Class:** 1  **NMR Class:** nc -26.3400 7.3000 ID: 20

Supplementary Information – Dynamic Undocking and the Quasi-Bound State

Supplementary Figure 12 (cont)

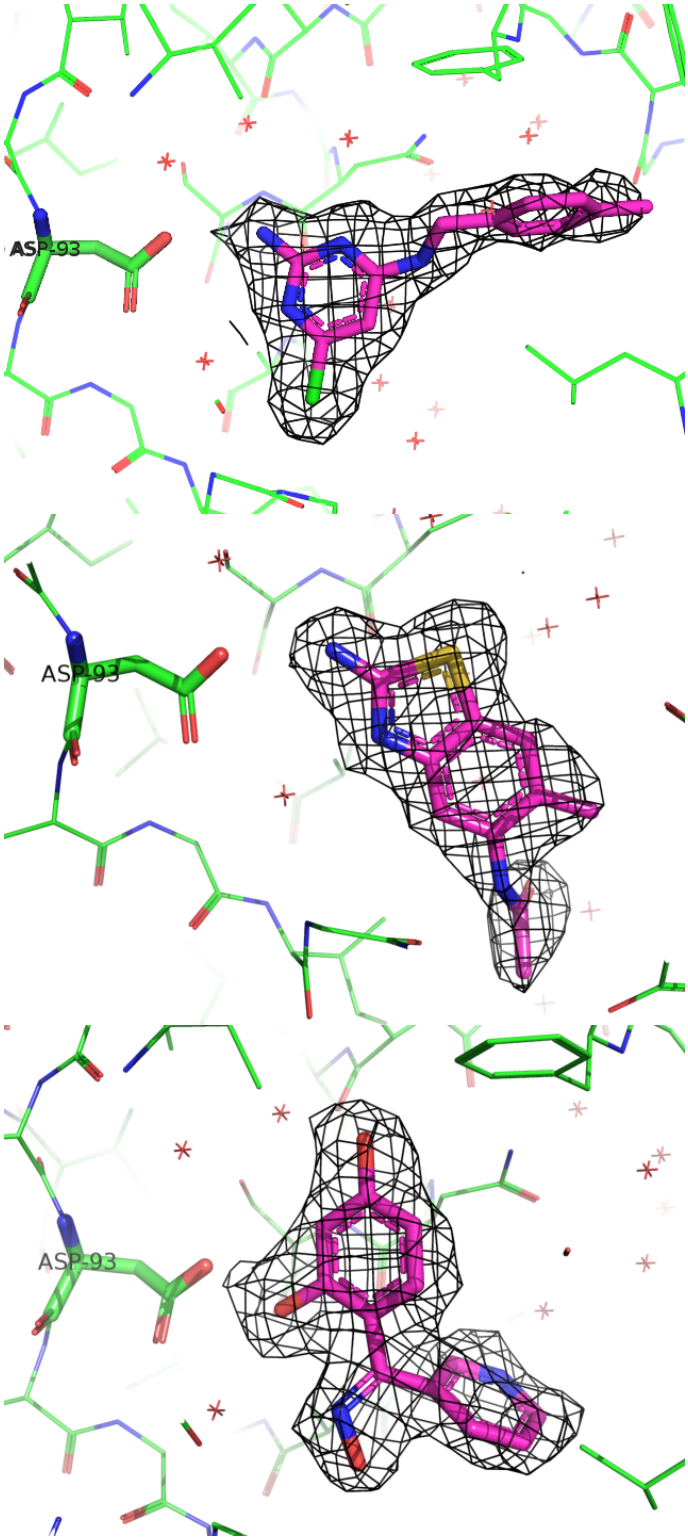
DUck Class: 1  **NMR Class:** r -24.0100 8.0000 ID: 21	**DUck Class:** 2  **NMR Class:** 1 -28.2700 5.6000 ID: 22	**DUck Class:** 2  **NMR Class:** 2 -26.8600 4.2000 ID: 23	**DUck Class:** 2  **NMR Class:** 2 -25.3300 3.6000 ID: 24
DUck Class: 2  **NMR Class:** 2 -25.3200 3.9000 ID: 25	**DUck Class:** 2  **NMR Class:** 3 -25.0900 5.5000 ID: 26	**DUck Class:** 2  **NMR Class:** 3 -27.3400 3.5000 ID: 27	**DUck Class:** 2  **NMR Class:** 3 -27.5500 3.2000 ID: 28
DUck Class: 2  **NMR Class:** nb -25.4600 4.7000 ID: 29	**DUck Class:** 2  **NMR Class:** nb -25.0900 5.5000 ID: 30	**DUck Class:** 2  **NMR Class:** nb -26.7600 4.4000 ID: 31	**DUck Class:** 2  **NMR Class:** nb -26.4700 3.1000 ID: 32
DUck Class: 2  **NMR Class:** nb -25.1000 4.9000 ID: 33	**DUck Class:** 2  **NMR Class:** nb -24.8300 4.8000 ID: 34	**DUck Class:** 2  **NMR Class:** nb -24.3200 4.4000 ID: 35	**DUck Class:** 2  **NMR Class:** nb -26.8200 3.0000 ID: 36
DUck Class: 3  **NMR Class:** 2 -27.9100 0.8000 ID: 37	**DUck Class:** 3  **NMR Class:** 2 -26.2500 1.4000 ID: 38	**DUck Class:** 3  **NMR Class:** 2 -25.4700 0.5000 ID: 39	**DUck Class:** 3  **NMR Class:** 3 -25.3700 1.0000 ID: 40

Supplementary Figure 12 (cont)

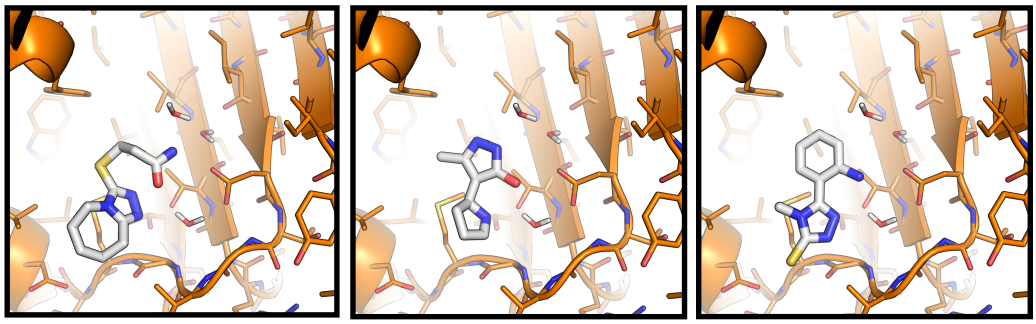
DUck Class: 3  **NMR Class:** nb -26.4800 2.3000 ID: 41	**DUck Class:** 3  **NMR Class:** nb -25.7700 2.5000 ID: 42	**DUck Class:** 3  **NMR Class:** nb -24.9900 0.0000 ID: 43	**DUck Class:** 3  **NMR Class:** nb -24.9700 0.0000 ID: 44
DUck Class: 3  **NMR Class:** nb -24.9100 0.0000 ID: 45	**DUck Class:** 3  **NMR Class:** nb -23.2300 2.0000 ID: 46	**DUck Class:** 3  **NMR Class:** nb -26.1100 1.7000 ID: 47	

Supplementary Figure 13

Crystal structure of Hsp90 in complex with compounds **1** (top), **2** (middle) and **3** (bottom). The 2fofc electron density maps are displayed at the 1.0 Sigma level (Carve = 1.7).

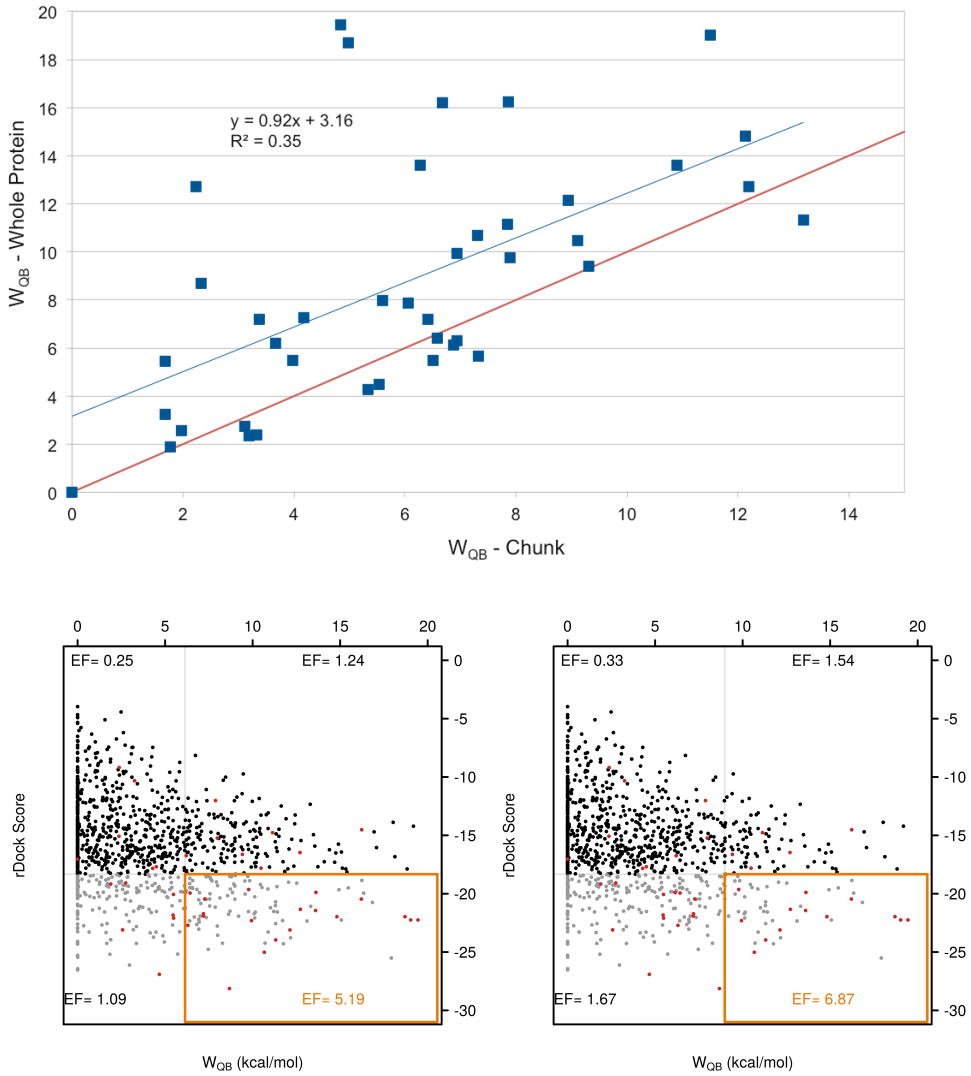


Supplementary Figure 14
Predicted binding modes for compounds **4**, **5** and **6** (from left to right).



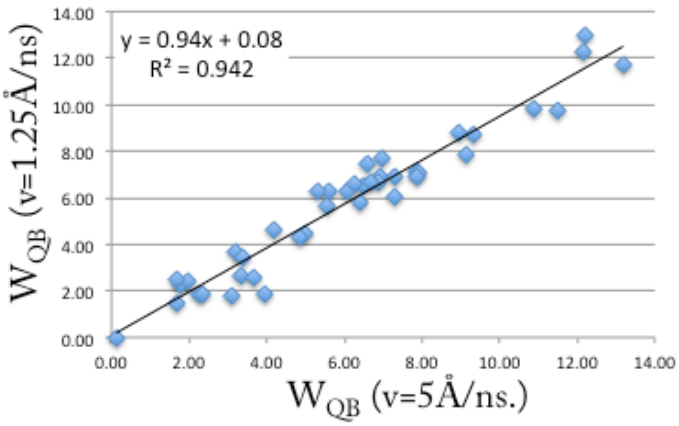
Supplementary Figure 15

Dependence of the results on the size of the receptor. W_{QB} values of CDK2 ligands were calculated using the whole protein as receptor and plotted against the results obtained with a truncated system (top). W_{QB} values obtained with the truncated system represent a lower bound to those obtained with the full system. This indicates that when the whole system is included, W_{QB} may not reflect the contribution of the interaction under investigation. Potentially, this may give rise to false positives. Noteworthy, the virtual screening results are comparable to those obtained with the truncated system (bottom; compare with Supplementary Figure 8).



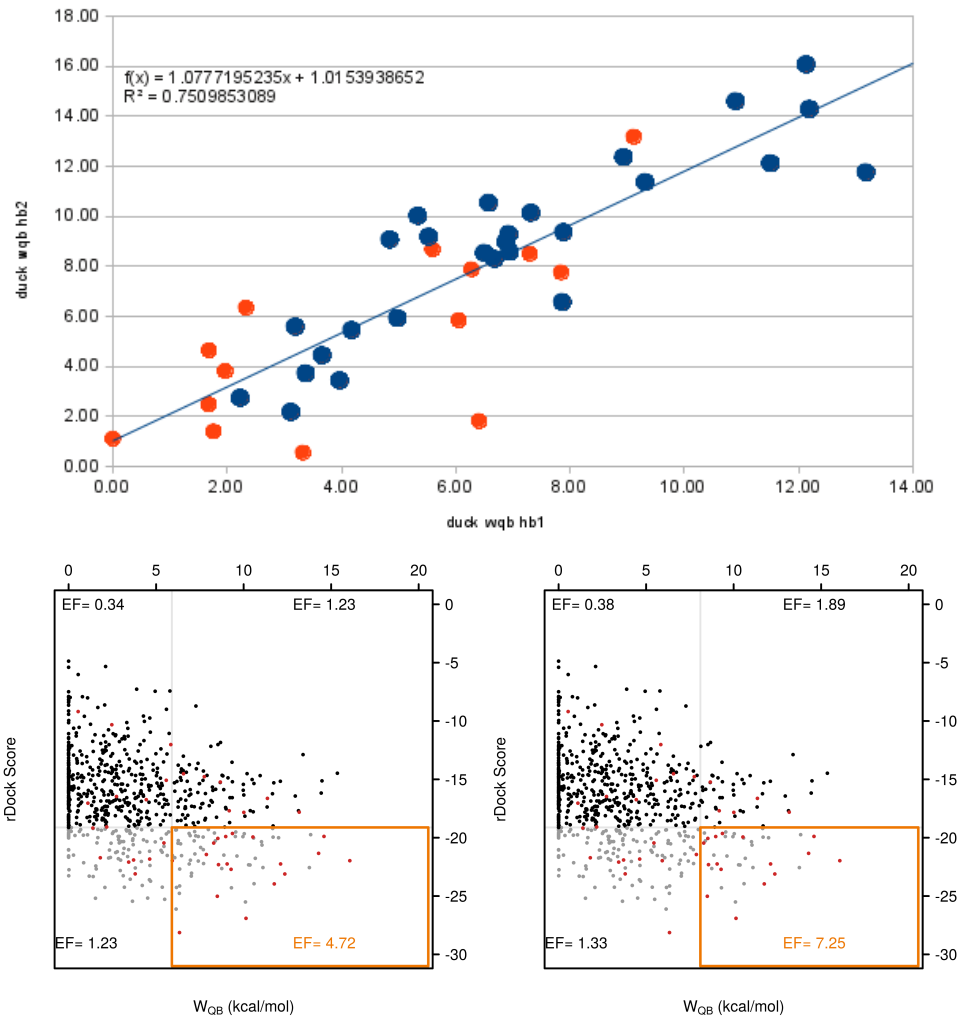
Supplementary Figure 16

Dependence of the results on the steering velocity. Two different velocities are compared: 5 Å·ns^{-1} (used through this work) and 1.25 Å·ns^{-1}. Slower velocities mean more sampling and, potentially, lower W_{QB} values. The high correlation (r^2=0.94) indicates that the standard conditions (v=5 Å·ns^{-1}) produce converged results.



Supplementary Figure 17

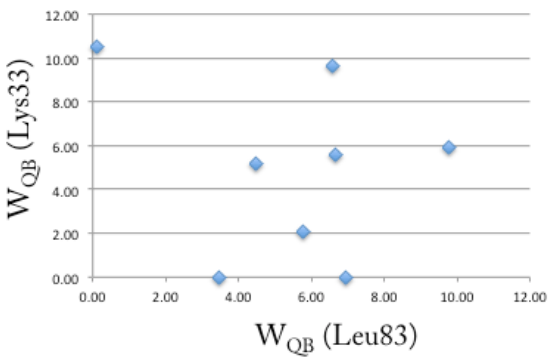
Dependence of the results on the choice of reaction coordinate. W_{QB} values obtained using two different atoms of reference in the hinge region of CDK2 are highly correlated (top) and afford similar enrichment factors in retrospective virtual screening (bottom; compare with Supplementary Figure 8). The atoms used as reference (Leu83:N in the x-axis and Leu83:O in the y-axis) are part of the hinge and located in close proximity (3Å). Most ligands form a hydrogen bond with both atoms at the same time. Points in red represent ligands that only form a hydrogen bond with Leu83:N.



Supplementary Figure 18

W_{QB} of CDK2 ligands pulling the amine of Lys33 and comparison with W_{QB} values obtained for the hinge region (in kcal/mol). Only those ligands capable of forming a hydrogen bond with NZ of Lys33 have been considered. It should be noted that this part of the active site presents large conformational diversity between structures. In consequence, the DUck results may be less reliable than for the hinge region.

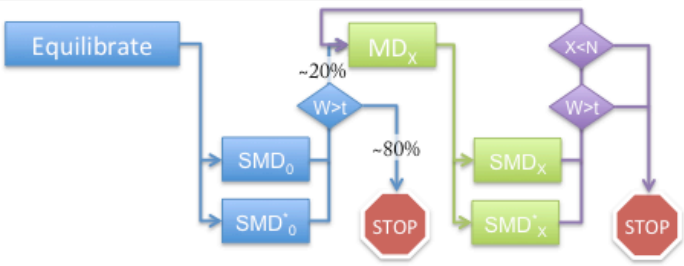
PDB Code	W_{QB} (O Leu83)	W_{QB} (Nζ Lys33)
1OIQ	4.48	5.18
3BHT	6.59	9.65
3BHV	6.94	0.00
3EJ1	5.77	2.06
3FZ1	0.12	10.50
3QTQ	6.66	5.56
3QTW	9.76	5.91
3TIY	3.47	0.00



Supplementary Figure 19

Proposed protocol for DUck-based virtual screening and comparison with MMPBSA. The smaller size of the system speeds up calculations by a factor or 5 (Supplementary Table 2), also permitting shorter equilibration times. Each ligand undergoes equilibration and at least two SMDs (45 GPU minutes). Molecules with W_{QB} above a given threshold (e.g. t=6 kcal/mol) would then proceed to N cycles of unbiased MD + SMD simulations (42 GPU minutes per cycle). A similar protocol for MMPBSA would require at least 2 GPU hours of equilibration followed by N cycles of 1ns MD simulation and MMPBSA calculation of representative snapshots (1 GPU hour + 20 CPU minutes per cycle).

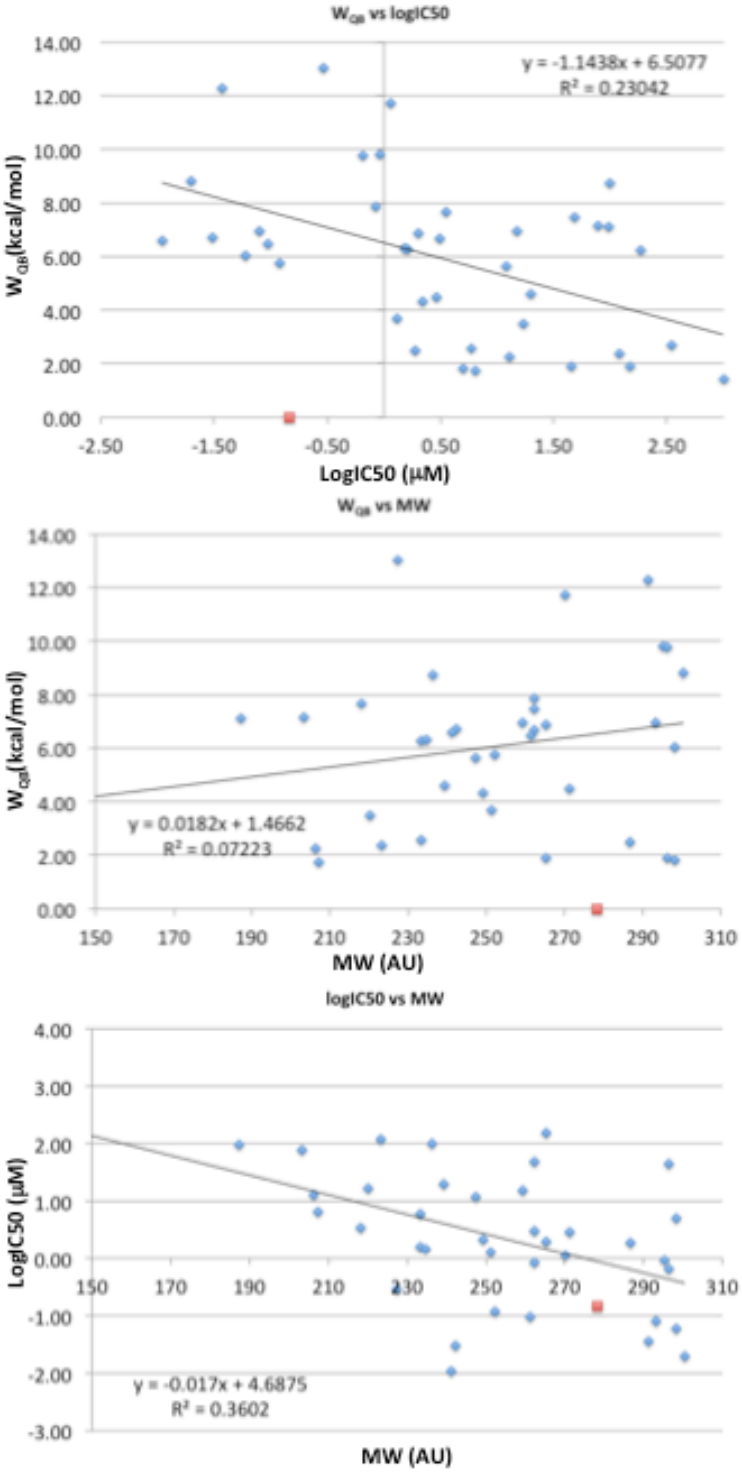
Simulated time	0.6ns	0.5ns x 2	0.5ns	0.5ns x 2
Wall clock time	15′	15′	12′	15′



Simulated time	1.5ns	1ns	n.a.
Wall clock time	120′	60′	20′

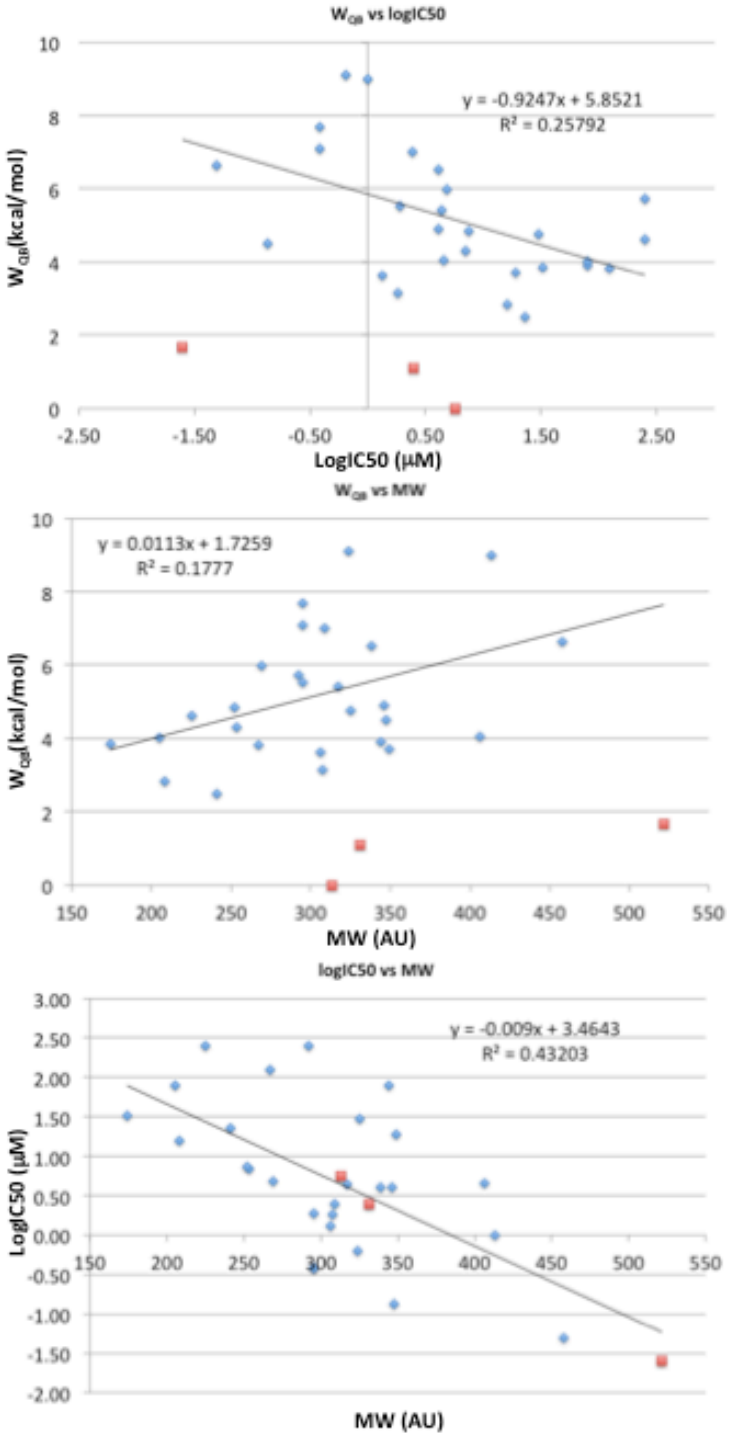
Supplementary Figure 20

CDK2 test set: correlation between LogIC50 and W_{QB} is not caused by Molecular Weight. The correlation between W_{QB} and MW (r^2=0.07) is lower than the correlation between W_{QB} and LogIC50 (r^2=0.23) or between LogIC50 and MW (r^2=0.36). Red points (discussed in Supplementary Figure 2) are excluded from all correlations.



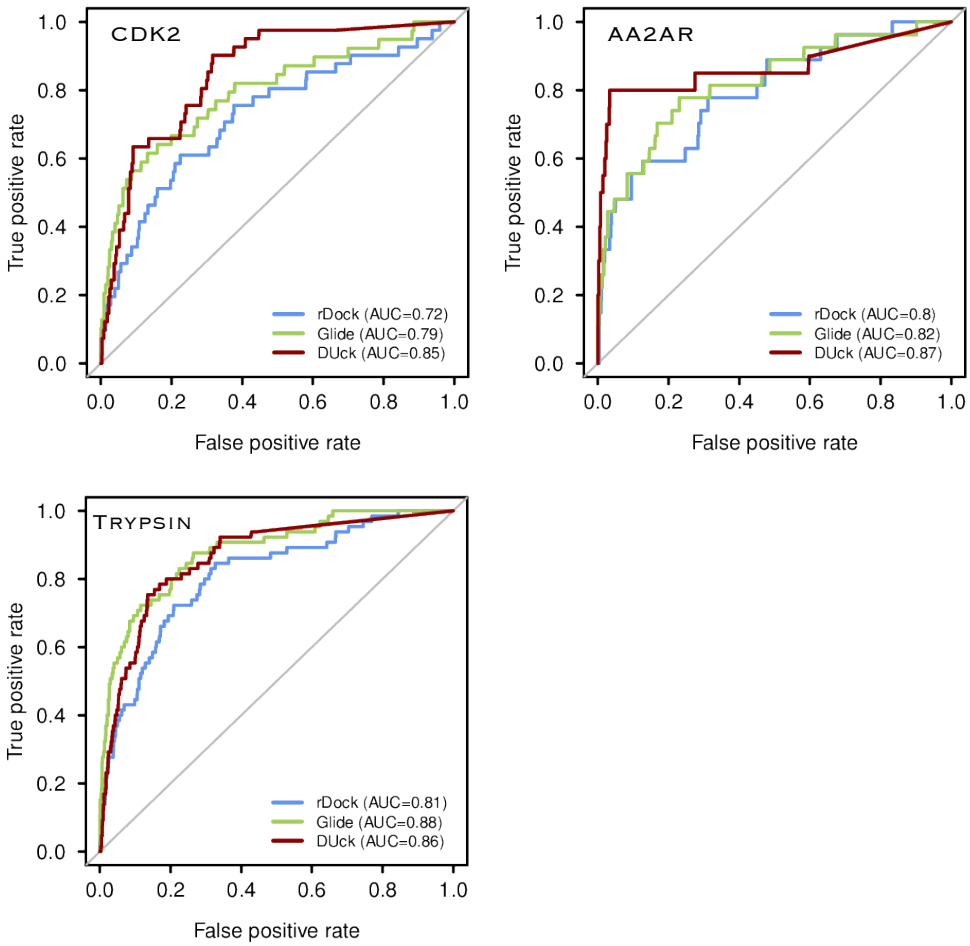
Supplementary Figure 21

BRD4 test set: correlation between LogIC50 and W_{QB} is not caused by Molecular Weight. The correlation between W_{QB} and MW (r^2=0.17) is lower than the correlation between W_{QB} and logIC50 (r^2=0.26) or between LogIC50 and MW (r^2=0.43). Red points (discussed in Supplementary Figure 5) are excluded from all correlations.



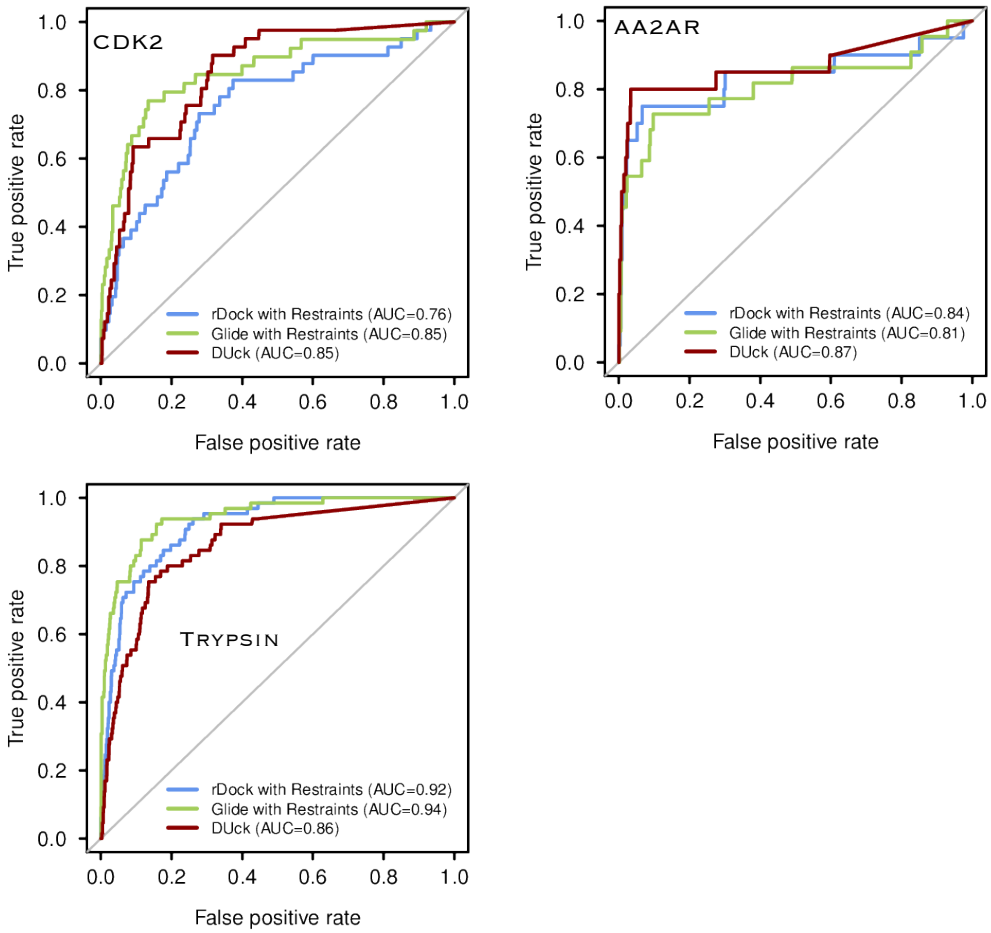
Supplementary Figure 22

ROC curves comparison of DUck (in standalone mode) with unbiased docking with Glide and rDock for the three test systems: CDK2, AA2AR and Trypsin.



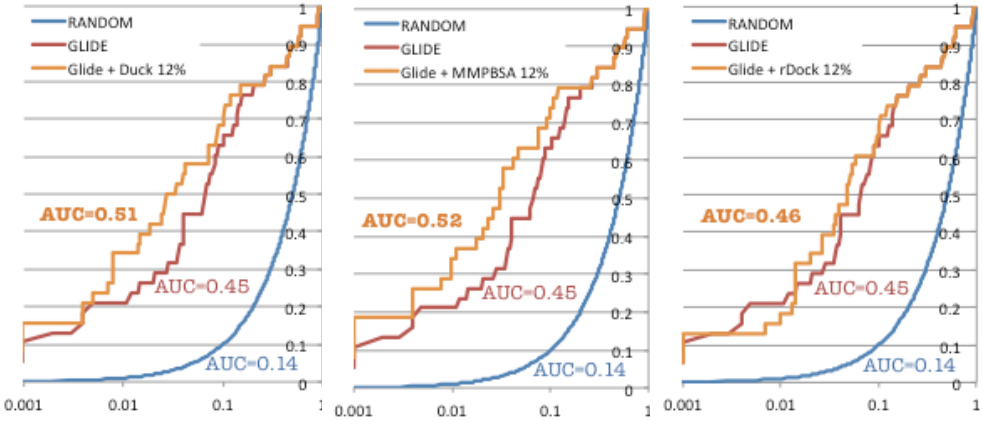
Supplementary Figure 23

ROC curves comparison of DUck (in standalone mode) with pharmacophore-guided docking with Glide and rDock for the three test systems: CDK2, AA2AR and Trypsin.



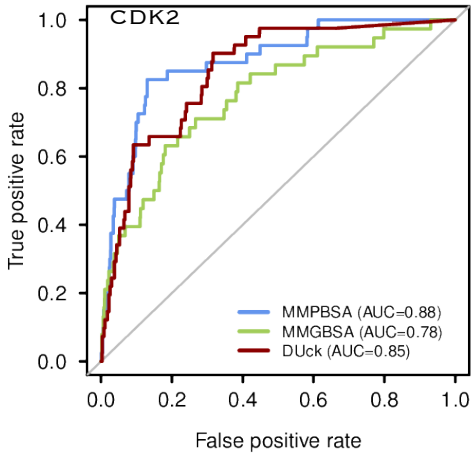
Supplementary Figure 24

DUck postfiltering improves early enrichment. Semilogarithmic ROC curves for the retrospective virtual screening of CDK2, obtained with the best-performing program for this test set (Glide),alone or in combination with three different postfiltering methods: DUck (left); MMPBSA (middle) and rDock (right). Ligands were initially ranked according to Glide's score. Then, moved to the back of the list if they were not in the top 12% of the rescoring method. This shows that the Glide-DUck combination is superior to Glide alone. For this test set the effect is most prominent in the top 1% to 5% of the library. Glide-MMPBSA combination is provided for comparison and affords very similar results. The Glide-rDock combination does not improve early enrichment.



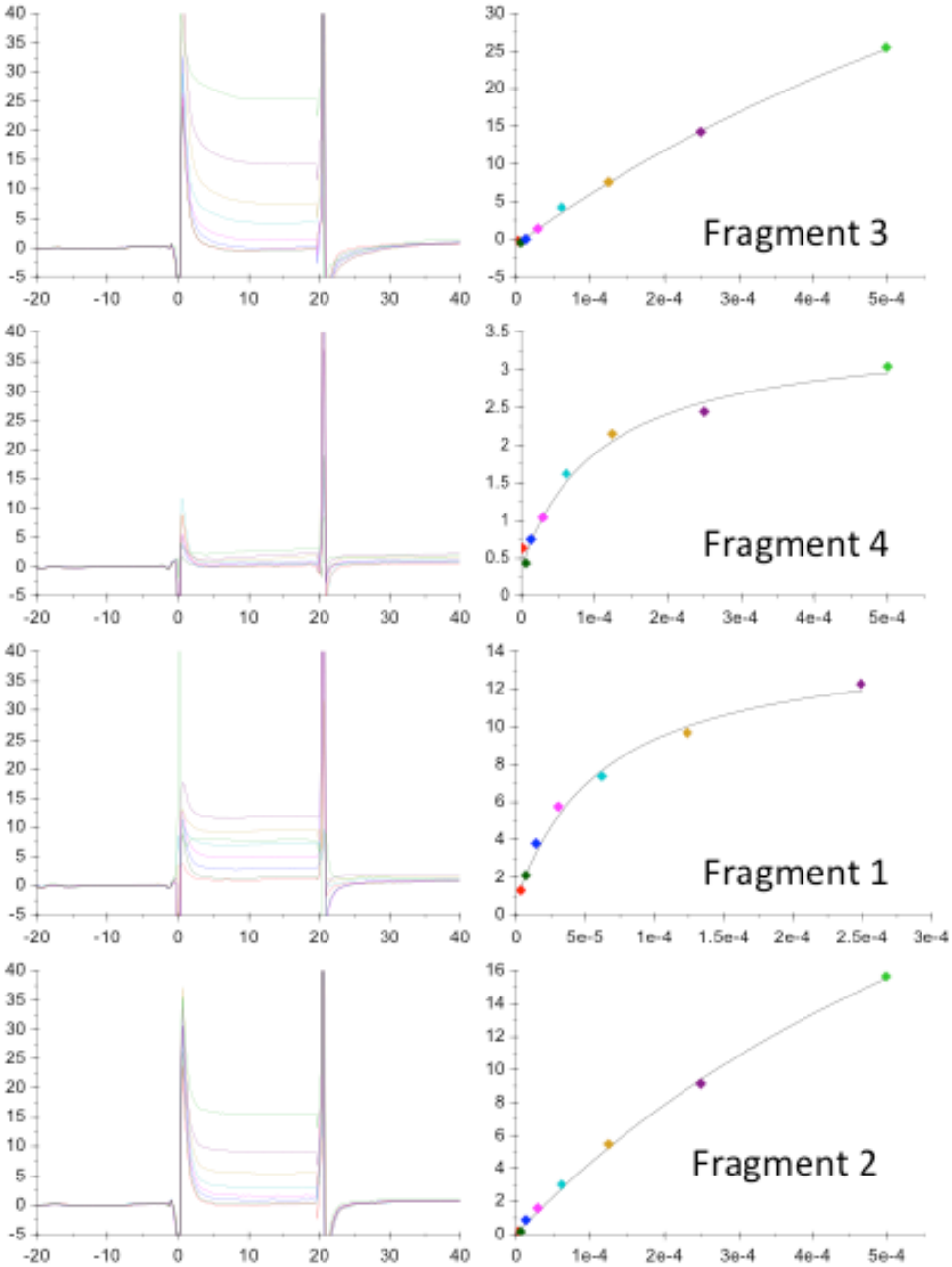
Supplementary Figure 25

ROC curves comparison of DUck (in standalone mode) with MMPBSA and MMGBSA for CDK2



Supplementary Figure 26

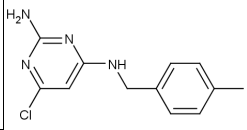
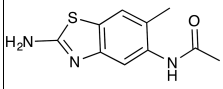
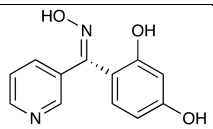
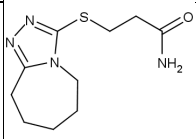
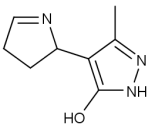
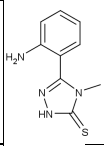
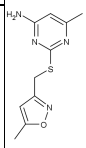
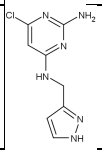
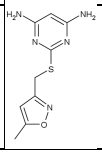
Examples of typical sensorgrams (left column) and steady state plots (right column) for the binding of the fragment hits to Hsp90. Fragments were tested in a 2-fold dilution series starting at 500uM or 250uM concentrations. Steady state values were calculated 4seconds before the injection stopped and plotted against the concentration. The K_D value was calculated by fitting the data to a steady state affinity model (Biacore T200 evaluation software GE Healthcare)



SUPPLEMENTARY TABLES

Supplementary Table 1

Chemical structures and summary of results for the 9 Hsp90 NMR Class 1 hits.

ID	Structure	MW	Docking		DUck		Xray	SPR Kd (mM)	PDB Sim[b]	ChEMBL Sim[b]
			Score	*Rank*[a]	W_{QB}	*Rank*[a]				
1		248,72	-24,97	*79*	9,1	*10*	Yes	77	2XDX (0.37)	CHEMBL 1340447 (0.44)
2		221,29	-25,03	*73*	8,2	*11*	Yes	320	2WI6 (0.29)	CHEMBL 1536318 (0.54)
3		230,22	-26,62	*19*	11,3	*1*	Yes	700	4EFU (0.32)	CHEMBL 1458840 (0.51)
4		240,33	-26,45	*22*	7,4	*16*	-	730	3WHA (0.29)	CHEMBL 1542436 (0.37)
5		165,20	-23,77	*128*	8,1	*12*	-	-	4EFT (0.27)	CHEMBL 1313412 (0.28)
6		206,27	-23,26	*138*	9,5	*5*	-	-	3HHU (0.42)	CHEMBL 2103879 (0.42)
7		236,30	-25,45	*51*	7,8	*15*	-	-	3B24 (0.31)	CHEMBL 1375884 (0.36)
8		224,66	-25,35	*58*	7,0	*22*	-	-	2XDX (0.35)	CHEMBL 1383799 (0.37)
22		237,29	-28,27	*2*	5,6	*33*	-	-	3O0I (0.27)	CHEMBL 1834092 (0.33)

[a] Position within the list of 149 molecules that were evaluated with DUck. [b]Hsp90 structure in the PDB or compound with Hsp90 activity in ChEMBL (as of 23/03/2016) with the closest similarity to the fragment hit. Similarity (values in parentheses) was calculated with Open Babel using the FP2 fingerprint.

Supplementary Table 2

Number of atoms of the investigated systems. On average, using a protein chunk with explicit solvation produces a system 20% in size relative to the whole protein. As computational times scale linearly with the number of particles, this represents a 5-fold gain in efficiency.

Number of Atoms

	Full System		Protein Chunk for DUck	
System	Protein	Periodic Box[a]	Protein[b]	Periodic Box[a,b]
Hsp90	3291	30387	527 (16,0%)	9415 (31,0%)
Cdk2	4578	46803	345 (7,5%)	9110 (19,5%)
AA2AR	4603	73039	525 (11,4%)	8815 (12,1%)
Trypsin	3231	26721	335 (10,4%)	9696 (36,3%)
Average	3926	44238	433 (11,0%)	9259 (20,9%)

[a] Protein solvated with TIP3 water molecules using Amber's tleap program. In all cases, the periodic system is a truncated octahedral box, the distance parameter is 12.0 and the closeness parameter is 0.65.[b] Values in parentheses are percentage of atoms relative to the full system.

Supplementary Table 3

Detail of the receptor definition used in DUck simulations. Water and residue numbers were taken from the corresponding PDB file.

System	Reference Atom	PDB Code (Chain)	Protein residues included as receptor	Water Molecules
CDK2	LEU 83 NH	1CKP (A)	ILE10 VAL18 LYS20 ALA21 VAL29 VAL30 ALA31 LEU32 VAL64 PHE80 GLU81 PHE82 LEU83 HIS84 GLN85 ASP86 LEU133 LEU134 ILE135 ASN136 ALA144	-
AA2AR	ASN 253 ND2	3EML (A)	LEU167 PHE168 GLU169 VAL172 PRO173 MET174 MET177 VAL178 ASN181 PHE182 TRP246 LEU247 PRO248 LEU249 HIS250 ILE251 ILE252 ASN253 CYS254 PHE255 THR256 PHE257 HIS264 ALA265 PRO266 LEU267 MET270 TYR271 LEU272 ALA273 ILE274	-
Trypsin	ASP189 OD1	2AYW (A)	HIS57 LEU99 ASP102 ASP189 SER190 CYS191 GLN192 GLY193 ASP194 SER195 VAL213 SER214 TRP215 GLY216 SER217 GLY219 CYS220 ALA221A GLN221 LYS224 PRO225 GLY226 VAL227 TYR228 THR229	1017 1096 1098 1101
Hsp90	ASP93 OD2	2YED (A)	GLU47 LEU48 ILE49 SER50 ASN51 SER52 SER53 ASP54 ALA55 LEU56 ASP57 LYS58 ILE78 ILE91 VAL92 ASP93 THR94 GLY95 ILE96 GLY97 MET98 GLY137 PHE138 VAL150 ILE151 THR152 GLY183 THR184 LYS185 VAL186	2043 2045 2049 2105 2107
BRD4	ASN140 ND2	3U5L (A)	TRP81 PRO82 PHE83 GLN84 GLN85 PRO86 VAL87 ASP88 ALA89 LYS91 LEU92 ASN93 LEU94 TYR97 ILE101 PRO104 MET105 THR131 ASN135 CYS136 TYR137 TYR139 ASN140 ASP144 ASP145 ILE146 MET149	-

Supplementary Table 4

Data collection and refinement statistics for Hsp90 in complex with Compounds **1**, **2** and **3**.R_{free} is the R factor calculated using 5% of the reflection data chosen randomly and omitted from the refinement process, whereas R_{cryst} is calculated with the remaining data used in the refinement. Rms bond lengths and angles are the deviations from ideal values; the rms deviation in B factors is calculated between covalently bonded atoms.

Compound	**1**	**2**	**3**
Data collection statistics			
Resolution (Å)	2.20	2.00	2.10
Space group	I222	I222	I222
Cell dimensions (Å) a = b = c =	 66.87 90.29 98.33	 64.96 88.41 99.06	 68.98 88.18 96.90
No. molecules/asymmetric unit	1	1	1
Solvent content (%)	57.25	54.73	57.41
Measured reflections	66152	66886	62479
Unique reflections	15401	19011	17526
Completeness: Overall / in hrb[a] (%)	99.5 / 98.5	96.7 / 90.9[b]	99.4 / 99.9
Mean I/σI: Overall / in hrb	11.2 / 2.8	11.1 / 1.3	8.33 / 0.95
R_{merge}: Overall / in hrb (%)	0.083 / 0.315	0.048 / 0.412	0.074 / 0.555
Refinement statistics			
R_{free} (%)	24.0	30.8[b]	27.6
R_{cryst} (%)	19.1	22.1	22.4
Rms Deviations: Bonds (Å) Angles (°) B Factor (Å^2)	 0.018 1.920 4.679	 0.019 1.958 5.536	 0.019 2.046 6.415
PDB Code	5FNC	5FND	5FNF

[a]hrb: highest resolution bin. [b]Diffraction data for this structure was collected from a crystal that did not cryo-freeze correctly therefore the data in some of the resolution bins was of a lesser quality than the equivalent data collected from the other two crystals. This is likely to have impacted the refinement statistics for this structure.

Supplementary Table 5

List of ligands in the CDK2 test set. Ligands highlighted in red are not included in the correlation plotted in Supplementary Figure 2 and Supplementary Figure 25.

PDB	No.Atoms	MW	IC50 (uM)	Log IC50	W_{QB} (kcal/mol)
1E1V	21	247.303	12.00	1.08	5.53
1E1X	22	251.292	1.30	0.11	3.19
1JSV	23	265.293	2.00	0.30	6.93
1JVP	19	233.274	1.60	0.20	5.59
1OIQ	23	271.325	2.90	0.46	4.98
1PF8	21	242.26	0.03	-1.51	6.88
1PXJ	16	206.267	13.00	1.11	1.76
1PXK	19	249.293	2.20	0.34	4.84
1PXM	23	298.365	0.06	-1.22	7.32
1VYW	24	291.355	0.04	-1.43	12.13
1VYZ	19	227.268	0.29	-0.54	12.19
1W0X	25	298.351	5.00	0.70	3.97
1WCC	10	129.55	350.00	2.54	3.33
2BTR	19	261.344	0.10	-1.02	6.50
2BTS	24	300.417	0.02	-1.70	8.94
2C4G	22	270.294	1.15	0.06	13.18
2C5O	17	207.275	6.50	0.81	3.12
2CLX	21	218.221	3.50	0.54	6.94
2EXM	17	203.249	78.00	1.89	7.84
2R3H	19	239.282	20.00	1.30	4.18
2VTA	10	118.139	185.00	2.27	6.05
2VTH	18	223.249	120.00	2.08	1.97
2VTJ	22	286.739	1.90	0.28	1.68
2VTL	16	187.203	97.00	1.99	7.89
2VTM	11	144.137	1000.00	3.00	1.68
2VTN	22	262.246	0.85	-0.07	9.11
2VTR	16	234.67	1.50	0.18	5.34
3BHT	20	241.255	0.01	-1.96	6.28
3BHV	26	293.291	0.08	-1.10	7.30
3EJ1	20	252.281	0.12	-0.92	6.41
3FZ1	24	278.352	0.15	-0.84	0.12
3PXY	22	233.233	5.90	0.77	3.66
3QQK	21	259.328	15.00	1.18	7.87
3QTQ	21	262.332	3.10	0.49	6.67
3QTR	24	295.361	0.93	-0.03	10.90
3QTW	24	296.349	0.65	-0.19	11.51
3R8Z	21	262.332	49.00	1.69	6.57
3RZB	20	236.294	100.00	2.00	9.31
3TIY	20	220.185	17.00	1.23	3.38
3TIZ	23	265.314	150.00	2.18	2.23
4EZ3	25	296.305	45.00	1.65	2.33

Supplementary Table 6

List of ligands in the BRD4 test set. Ligands highlighted in red are not included in the correlation plotted in Supplementary Figure 5 and Supplementary Figure 26.

PDB	No.Atoms	MW	IC50 or Kd (nM)	Log IC50	W_{QB} (kcal/mol)
3MXF	31	458.00	49.00	-1.31	6.63
3U5J	22	308.77	2460.00	0.39	7.00
3U5L	23	323.78	640.00	-0.19	9.12
4A9L	22	325.38	30000.00	1.48	4.78
4C66	23	343.85	79400.00	1.90	3.92
4CFK	23	307.35	1830.00	0.26	3.13
4CFL	23	306.36	1330.00	0.12	3.63
4E96	24	347.39	136.00	-0.87	4.51
4HBV	13	241.09	23000.00	1.36	2.51
4HBW	18	269.32	4800.00	0.68	5.98
4HBX	20	295.36	1900.00	0.28	5.53
4HBY	22	317.36	4400.00	0.64	5.42
4HXR	21	338.41	4100.00	0.61	6.54
4HXS	23	346.42	4100.00	0.61	4.90
4J0R	22	295.34	386.00	-0.41	7.70
4J0S	22	295.34	382.00	-0.42	7.10
4LR6	13	174.20	33000.00	1.52	3.85
4LZS	15	208.26	16000.00	1.20	2.84
4MEN	20	267.33	125000.00	2.10	3.84
4MEO	22	292.34	250000.00	2.40	5.73
4MEQ	17	225.25	250000.00	2.40	4.62
4O72	30	413.49	1000.00	0.00	9.00
4O74	38	521.66	25.00	-1.60	1.67
4O77	25	331.35	2500.00	0.40	1.10
4O78	30	406.44	4600.00	0.66	4.07
4O7A	23	349.17	19000.00	1.28	3.72
4O7E	24	313.36	5700.00	0.76	0.00
4PCE	19	253.34	7000.00	0.85	4.30
4PCI	19	252.31	7500.00	0.88	4.84
4UYD	15	205.22	79400.00	1.90	4.03

Docking-Undocking Combination Applied to the D3R Grand Challenge 2015

Sergio Ruiz-Carmona[1,2] **and Xavier Barril**[1,2]

Journal of Computer-Aided Molecular Design, 2016, Accepted

1 Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Departament de Fisicoquimica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain, **2** Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010, Barcelona, Spain.

Docking-undocking combination applied to the D3R Grand Challenge 2015

Sergio Ruiz-Carmona[1], Xavier Barril[1,2,*]

[1]Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain.

[2]Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

* Send correspondence to: xbarril@ub.edu

Abstract

Novel methods for drug discovery are constantly under development and independent exercises to test and validate them for different goals are extremely useful. The Drug Discovery Data Resource (D3R) Grand Challenge 2015 offers an excellent opportunity as an external assessment and validation experiment for Computer-Aided Drug Discovery methods. The challenge comprises two protein targets and prediction tests: binding mode and ligand ranking. We have faced both of them with the same strategy: pharmacophore-guided docking followed by dynamic undocking (a new method tested experimentally here) and, where possible, critical assessment of the results based on pre-existing information. In spite of using methods that are qualitative in nature, our results for binding mode and ligand ranking were amongst the best on Hsp90. Results for MAP4K4 were less positive and we track the different performance across systems to the level of previous knowledge about accessible conformational states. We conclude that docking is quite effective if supplemented by dynamic undocking and empirical information (e.g. binding hot spots, productive protein conformations). This setup is well suited for virtual screening, a frequent application that was not explicitly tested in this edition of the D3R Grand Challenge 2015. Protein flexibility remains as the main cause for hard failures.

Introduction

Computer-Aided Drug Discovery (CADD) methods are constantly under development anda wide spectrum of options is available to the scientific community to address each specific situation at every stage of the drug discovery process[1-3].

Independent validation experiments are extremely useful to test the different methods, try them out under different circumstances and validate them for a specific goal. For instance, there have been experiments to help the development of protein structure modeling software [4],the prediction of protein-protein interactions [5]or certain physico-chemical properties of small molecules [6].

In this direction, the D3R Grand Challenge2015 provides an independent exercise to assess and validate CADDtools related with protein-ligand interactions. Two proteins (Hsp90 and MAP4K4) with datasets comprising different ligands with measured affinities and crystal structures are provided as blind sets. Different measures for each of the datasets were used to evaluate the performance of different methods in two situations that are common in drug discovery projects: ligand ranking and binding mode prediction.

Docking, scoring and free energy methods have been widely applied in structure-based drug discovery [7–12] as they provide an excellent assistance particularly in early stages of the development of new drugs. Docking is a very common method that can be used both for predicting the binding mode of a protein-ligand complex and for virtually assaying thousands to millions of drug-like molecules in a relatively short amount of time, speeding up the finding of promising candidates and dramatically decreasing the cost in comparison with the experimental alternative.[13] However, the scoring functions employed in docking have been trained to reproduce specific data sets and are qualitative in nature. As such they are not expected to correlate with binding free energies.[14] Further limitations include receptor flexibility or the presence of water molecules that can be wither trapped or displaced by the ligand.

In our particular approach, docking is a central component tackling the D3R Grand Challenge 2015, but we aim to overcome some of its limitations with complementary tools and, whenever possible, guiding the calculations with previous knowledge about the systems. Specifically, we have used rDock software [15] as the docking engine, using pharmacophoric restraints to ensure that the predicted ligand poses fulfil certain key

interaction points.[16–18] In the case of Hsp90, they correspond to a hydrogen bond with the carboxylate of Asp93 and, in case of MAP4K4, a hydrogen bond with the nitrogen atom of Cys108 in the hinge region (a short linear sequence that acts as a hinge between the N-terminal and C-terminal domains in kinases).These interaction points can be identified merely by superimposing all the available crystal structures of protein-ligand complexes for each system in the PDB and obtaining a pharmacophore definition as detailed in the Methods, which can be supplemented to rDock in order to increase its efficiency as shown in previous studies [15]. Hsp90 presents at least one water molecule that can be displaced by certain ligand classes. By excluding this water molecule, we make the receptor definition valid for all chemotypes[19]. Then, to address the protein flexibility, we took a knowledge-based approach. We investigated the effect of protein flexibility on docking performance using Hsp90 as a test set, so we are familiar with the different conformations the protein can adopt upon ligand binding. We selected the We selected the most common conformation amongst all known Hsp90 protein-ligand complexes (namely, closed lid) for running docking and revised the quality of the predictions knowing that certain chemotypes can induce a conformational change of the lid to the open or helical states.[20] In contrast, MAP4K4 is a much less well characterized proteinand we took a best guess based on our previous knowledge about other kinases. As we will discuss below, the different degree of previous knowledge for each system has had a major effect on the outcome and highlights the importance of the human factor, which remains essential even as the computational tools improve.

Finally, we have introduced the use of Dynamic Undocking (DUck), a new tool used to assess the structural stability of protein-ligand complexes.[21] Here we have experimentally adopted a consensus approach, where the docking poses are re-evaluated and re-ranked based on their resistance to break the key hydrogen bonding interaction. This approach allowed us to detect not only false positives but also false negatives from docking results. DUck has been shown to be orthogonal to docking, as it evaluates structural stability as opposed to binding affinity.[21] For some ligands, re-scoring by DUck has allowed us to identify good binding poses which are *apriori* discarded due to bad docking scores. In other cases, docking and DUck selected the same pose, increasing our confidence on predicted binding modes that would be deemed doubtful if they had been backed up only by docking.

In the next sections, we will discuss in detail the methodology and the results obtained in the D3R Grand Challenge, drawing some conclusions to explain the failures and successes, as well as some recommendations for future editions of this challenge.

Methods

Selection of Cavity

The D3R Grand Challenge 2015 has two differentiated objectives: predict the crystallographic poses and the affinities or rankings for a series of ligands. Both of these objectives rely on a good definition of the system and a reliable characterization of the ligand-receptor interaction is crucial. For Hsp90, 4 receptor structures from the PDB were proposed by the organizers (2JJC, 2XDX, 4YKR and 4YKY). All of them were in the so-called closed conformation of the lid with the exception of 2XDX, which had the lid in open conformation.

As most of the known ligand-Hsp90 complexes have the lid in closed conformation, 2XDX was discarded. 2JJC was also discarded because, unlike the ligands in the test set, it is a very small and may be unable to modulate the cavity for better docking performance.[22, 23] Structures 4YKY and 4YKR are very similar in all respects (both bind a ligand of the resorcinol family) and were considered equivalent. The former was selected as reference structure. In a previous study [15] we demonstrated the improvement in virtual screening applications when guiding the docking process by adding previous knowledge, with a specific example for Hsp90. Additionally, it is known that three interfacial water molecules have an important role mediating the protein-ligand contacts. For this reason, they have been included in all docking runs as structural waters in the binding site. Some ligand types (e.g. adenine) interact with a fourth interfacial water molecule, but it is displaced by others ligands (e.g. resorcinol) and cannot be kept as part of the receptor.[24] Hence, the protocol used in all the docking calculations for Hsp90 includes a pharmacophore definition of two hydrogen bonds with Asp93:OD2 and one of the water molecules (included in all the runs as non displaceable), as previously defined in [15]. For undocking, the water molecule is added explicitly to the initial structure. In case of MAP4K4, 2 receptor structures from the PDB were supplied by the organizers (4OBO and 4U44). The main difference between the conformations of the two crystals is a loop folding towards the hinge region in 4OBO, thus decreasing the size and the solvent exposure of the binding site. Due to those restrictions we decided to use 4U44 as reference for all MAP4K4 applications, which had a bigger and more accessible binding site. In order to guide docking, we performed a pharmacophore search(more details in the next section)using all crystal structures of MAP4K4-ligand complexes in the PDB. We then supplied all docking calculations with a pharmacophore defined by a hydrogen bond with Cys108:N, located in the hinge region.

Pharmacophore Search

To get a reliable pharmacophore definition for the MAP4K4 system, a set of known protein-ligand 3D structures was necessary. We selected all MAP4K4 protein-ligand complexes from the PDB (4OBO, 4OBP, 4OBQ, 4RVT, 4U40, 4U41, 4U42, 4U43, 4U44, 4U45, 4ZK5 and 5DI1) and aligned them to the reference 4U44.The “Pharmacophore Search” tool of MOE was run and a hydrogen bond with Cys108:N in the hinge region was selected as pharmacophore. It was fulfilled by all 12 ligands in the PDB subset. Moreover, it was consistent with other protein-ligand interactions in the kinases family.[25, 26]

Molecular Docking

For all molecular docking simulations we used rDock[10-12], a fast and reliable docking program that we released as open source several years ago. To run rDock, only a correctly prepared 3D structure of the receptor and a definition of the binding site are needed. In this work, we defined the cavity using the crystallized ligand found in both PDB structures for Hsp90 and MAP4K4, 4YKY and 4U44 respectively. Some rDock rbcavity parameters were decreased with respect to the default values in order to optimize the binding site definition: *radius* (changed from 10.0 to 6.0), which defines the region around the reference ligand that will be used to define the docking binding site and *max_cavities* (from 99 to 1), as we only want to run docking in one cavity. The pharmacophoric restraints were defined as mandatory and all the ligands unable to fulfill the definition were discarded. For the docking protocol, no modifications were made to the standard as previously published [15]: 50 individual docking runs per ligand, which is considered exhaustive sampling, in order to ensure that the lowest-energy binding mode is found.

Receptor Preparation

The 3D structure of the receptor has to be provided to rDock with standard Tripos MOL2 format and atom types [28]. However, as rDock relies on the user-supplied structure, we need to provide it with correct protonation states and charges, as well as correct orientations of flexible side chains (rDock only considers as flexible atoms of the receptor the hydrogen atoms of terminal OH and NH3+groups within 3Å of the binding site cavity).The “Structure Preparation” tool from MOE [29] was used to protonate at pH 7.0 and correct all the issues found for Hsp90 and MAP4K4 receptors, such as chain breaks, missing loops or disulfide bonds, incorrect residue labelling or alternate conformations. The prepared structures were then saved in mol2 format and used as input for rDock.

Ligand Structure

As all ligands provided by the organization were in 2D format, Ligprep from Schrödinger [30] was used to calculate the 3D structure with correct topology, bond orders and geometry of bonds, angles, dihedrals and rings. The ionisable groups were protonated at pH=7 with a tolerance of +/- 1.All ligands were saved in MOL SDF format and used as input for docking.

Dynamic Undocking

We used Dynamic Undocking, or DUck, as a complementary tool to molecular docking in order to improve the overall performance of docking-based virtual screening.[21]DUck is a methodology developed in our group based on Steered Molecular Dynamics (SMD). The interaction of the ligand and the receptor with the key interaction point (specified when defining the cavity and protocols for docking) is monitored with SMD. In particular, DUck simulations consist on unbiased molecular dynamics (MD) simulations of the complex and repeated SMD simulations launched at 1 ns intervals of the MD to simulate the rupture of the ligand-receptor interaction and measure the force needed to achieve a state where the interaction has just been broken or, as we named it, a Quasi-Bound state. The work profiles obtained from the SMD simulations are processed to obtain the work to achieve the Quasi-Bound state (W_{QB}), which will be used to score and rank the ligands. Moreover, in order to increase throughput and reduce the influence of peripheral interactions and focus on the desired interaction, we use a model receptor that includes only a small part of the protein of interest. This portion is created around the defined key interaction point and preserves its local environment, simplifying also the dissociation pathway and avoiding artifactual results (more details about DUck can be found in reference [21] and http://www.ub.edu/bl/undocking/).For Hsp90 and MAP4K4, the following protocol was set: protein models were created containing the residues with any atom within6Å around the key interaction points (as detailed in *Selection of the Cavity* section) and manually refined to include other important residues for the binding site environment(Figure S1; Table S2). The best-scored docking poses for each ligand were subjected to an in-house script that automatically parameterized each ligand and prepared the necessary files for running the MD and SMD simulations of DUck. Each protein-ligand complex system was placed in a cuboid box with a minimum distance between each atom and the edge of the box of 12Å in every dimension and solvated with TIP3P water molecules and Na+ or Cl- ions were added to the solvation box depending on the charge of each of the protein-ligand complexes in order to ensure the electroneutrality of the simulated systems..Due to the artificiality of the protein models, MD simulations were run with harmonic restraints (1

kcal/mol·Å^2) in all heavy atoms of the receptor to prevent big conformational changes. In order to preserve key hydrogen bond interaction during the equilibration part of the simulations, distances beyond 3Å are penalized(parabolic restraint with k=1 kcal/mol·Å^2 between 3Å and 4Å; linear restraint with k=10 kcal/mol·Å beyond 4Å). All unbiased MD steps were run using a Langevin thermostat with the cutoff for non-bonded interactions set to 9Å and the collision frequency to 4 ps^{-1}.The equilibration consisted in 1000 cycles of minimization, gradual warming from 100K to 300K for 400 ps in the NVT ensemble and equilibration of the system for 1 ns in the NPT ensemble. At intervals of 1 ns (starting right after the equilibration), two SMD runs are executed from the same restart file (at 300K and 325K, as described in reference [21]) for 500 ps. During this time, the distance of the key hydrogen bond is steered from 2.5 Å to 5.0 Å with a spring constant of 50 kcal/mol·Å^2. More unbiased MD steps (1 ns each) were run to create more starting points for SMD runs to repeat the process as much as desired. All simulations were run with AMBER 14 [31] using in-house NVIDIA GeForce TITAN X GPUs or at the Barcelona Supercomputing Center using NVIDIA Tesla M2090 GPUs. AMBER force field 99SB was used for the protein and parm@Frosst[32] for the ligands.

Binding Mode Prediction

For all of the ligands where a binding mode was to be predicted, the protocol was the following: 1- Run docking as described in the “Molecular Docking” section above. 2- From the docking results, select a set of poses with a RMSD between them higher than 1 Å using the *sdrmsd* script from rDock package. 3- Run DUck to calculate the W_{QB} for all the sets of selected poses per ligand. 4- Select the pose with the highest W_{QB} as the correct binding mode and 5- visually inspect the results to check the selected poses fulfilled the defined interaction and the receptor conformation (more details in the following sections).

Ligand Ranking

A few differences from the protocol for binding mode prediction were introduced in case of ligand ranking: 1- Run docking as described in the “Molecular Docking” section above. 2- From the docking results, select the top scored pose for each ligand. 3- Run DUck to calculate the W_{QB} for the selected poses.4-For each of the ligands in the sets, the similarity to all known PDB ligands with measured affinity for the corresponding receptor(Hsp90 or MAP4K4)was calculated and taken into account to check the rankings and possible docking errors.5- Docking score and W_{QB} from DUck were normalized for each of the sets. All ligands were ranked according to the sum of the two corresponding normalized scores. In the cases where docking was not

able to find a good binding mode (i.e. the key interaction was not fulfilled), the similarity of each ligand with respect to other ligands in the challenge set and other ligands in PDB was used to assign a corrected ranking. Finally, a final step of visual analysis was carried on to check all ligands and re-rank some of them taking into account our previous experience.

Results & Discussion

Following our primary hypothesis, we designed a docking protocol that would reinforce the importance of the most important binding hot spot. This was done through the introduction of pharmacophoric restraints that forced the presence of hydrogen bonding groups at specific locations (Figure 1). The protein conformation was chosen to be as general as possible, thus for MAP4K4 we selected 4U44 as it has a bigger cavity than other structures available. For Hsp90, the biggest cavities present a ligand-induced hydrophobic sub-pocket (the PU3 cavity), but the associated protein conformation (helical) is energetically penalized and tends to downgrade the docking results[19]. For this reason, we chose a non-helical conformation (4YKY) taking care that the binding site was not blocked by any side-chain.

Binding Mode Prediction

We ran rDock to generate 50 poses per ligand. Poses with restraint penalties higher than 1 kJ/mol (indicating that the pharmacophore is not fulfilled) were discarded. After that, we selected a diverse set of the remaining poses, sorted by docking score to be re-evaluated by Dynamic Undocking (DUck). On average, 10 poses per ligand were selected for next step. DUck measures the work needed to break a given hydrogen bond (W_{QB}). We have found that true ligands in their correct binding mode, form hydrogen bonds that are much harder to break than decoys [21]. Here we employ this method to compare various binding modes of the same ligand. In the majority of cases, the binding pose with the best docking score also presented the highest W_{QB} value and was proposed as the correct solution. But often DUck provides a much more clear distinction between poses, removing uncertainty from the decision. This is illustrated with the Hsp90 ligand 40, which presented two alternative binding modes (Figure 2). In the first binding mode, the ligand interacts with Asp93 through the resorcinol, whereas the cyclic urea plays this role in the second binding mode. Though their docking scores are relatively similar (-23.4 and -18.9 kJ/mol, respectively), the hydrogen bond formed by the second binding mode is extremely labile (W_{QB} = 0.5 kcal/mol), which makes this binding mode very unlikely. By contrast, the first binding mode presented a very strong hydrogen bond (W_{QB} = 17.7 kcal/mol) and was selected with full confidence. For Hsp90, in several cases a lower ranking pose was selected based on the DUck calculation(Table S1). This is shown in Figure 3, where the Hsp90 ligand 73 presents a relatively similar binding mode with two different orientations. The first one (green) is the preferred one by docking

(score = -20.3 kJ/mol), whereas the second one (pink) is heavily penalized due to a steric clash of the 1-chloro-3-nitrobenzene moiety (score = 1.3 kJ/mol). Dynamic undocking indicated that the latter binding mode was actually preferred (W_{QB}= 11.6 kcal/mol vs. 10.9 kcal/mol), which prompted us to seek a protein conformation where the second binding mode would fit without clashing. In this particular case, the ligand binds to helical conformation (e.g. 2WI6) where a hydrophobic pocket (the PU3 pocket) emerges.[20]

The results submitted to the stage 1 of the D3R Grand Challenge 2015 are summarized in Table 1 and Table 2 for Hsp90 and MAP4K4, respectively. The accuracy of binding mode prediction is generally measured in terms of root mean squared deviation (RMSD) from the crystallographic pose. It is also common to convert this value to a binary decision (correct/incorrect) based on a fixed threshold (usu. 2.0 Å). This is a debated topic, and several alternative solutions have been put forward [18,19].

In practice, the best measure may depend on the particular problem that one is facing. For instance, a prediction that captures the main interactions is valid when dealing with a new chemotype, but inadequate at the lead optimization stage. Since our lab focuses on the hit identification stages of drug discovery, we are particularly interested in predicting the position of the central scaffold, i.e. the part of the ligand that forms the main interactions and defines the vectors of growth in the hit to lead stage. Thus, we have complemented the objective RMSD measure with a subjective binary classification telling if the prediction is sufficiently accurate to be used in the hit progression. In terms of RMSD, our average results were 1.6 ± 0.9 Å for Hsp90 (8th position among the participants of the D3R Grand Challenge 2015) and 3.7 ± 2.8 Å for MAP4K4 (3rd position). On the former set, we predict all but one ligand within 2.0 Å of RMSD. The only exception is ligand 44 (RMSD = 3.0 Å), but even then the position of the scaffold is correct and the deviation is due to the different orientation of a part of the ligand that does not engage in interactions with the protein(Figure S2).

The MAP4K4 results are much worse, but we still fared better than most participants, which highlight the difficulty of this set. Using the 2.0 Å RMSD cutoff, we only predicted 11 ligands correctly (37%). In our subjective assessment, we predicted the position of the scaffold correctly for 18 ligands (60%). The reason behind the poor performance is almost exclusively due to the flexibility of the protein. As this is a key issue in molecular docking, it will be discussed in detail. On the positive side, our protocol was still capable of predicting the main interaction correctly for a majority of ligands. Worthy of note, the structure of ligand 32 was originally inverted (Figure 4). Docking, but particularly Dynamic Undocking, argued strongly against this binding mode. After consultation with the crystallographers our predicted binding mode was accepted as

the proper binding mode. This is a reminder of the necessary dialogue between crystallographers and modellers, particularly where various binding modes are consistent with the observed electron density (e.g. due to tautomerism).[35]

Protein flexibility: the greatest docking challenge?

Reviewing the cause of the cases in MAP4K4where we failed in making a good prediction, we found that using a single receptor conformation was by far the most important factor. There is a large body of literature indicating the importance of protein flexibility[21-23],but back in 2005 we demonstrated that using multiple protein conformations could actually downgrade the results, particularly in virtual screening applications[19]. Since then, other authors have suggested that judicious selection of two or three structures can produce a small but systematic improvement over the best single structure.[39–41]However, as we did not have any previous knowledge on this system, we adopted the simple approach of using the biggest cavity (4U44), hoping that it would be valid for a larger proportion of ligands[42].

Once the experimental structures were disclosed, we observed that a large proportion of the ligands actually bind to a conformation where the cavity is partly occluded by the side-chain of Tyr36 in the P-loop(Figure 5). In order to measure the impact of these effects, we ran the exact same experiments using as receptor structure 4OBO (Tyr36-IN), which has this alternative conformation. As shown in Table 2, most of the recalculated poses have an RMSD lower than the one we submitted to the D3R Grand Challenge 2015. Taking the best RMSD of the different binding modes, we obtain an average RMSD improvement of 1.1 Å (2.6 vs. 3.7 Å) with 18ligands (60%) below the 2.0 Å threshold and 23 ligands (77%) with a correctly placed scaffold. While the results are still imperfect, one must consider that three structures are still insufficient to represent the whole array of conformational possibilities. In fact, we deem that there are only 2 ligands (7%) for which the failure cannot be attributed to the conformation of the protein: Ligands 4 and 17 do not form a hydrogen bond with the backbone of Cys108 (the hinge region) and are thus incompatible with our docking and dynamic undocking protocol. On the other hand, if the relative energies of the conformational states are not properly considered, using multiple structures may cause more problems than it solves.[43] In our opinion, except for direct experimental observation of the conformational states,[44] empirical knowledge gained from detailed analysis of multiple crystallographic structures is – at present – the only practical solution to this problem.

This is indeed the case for Hsp90, a system that we have studied thoroughly. Here, we were able to predict not only the structure of the ligands, but also which conformation would the protein adopt upon ligand binding. This aspect was not evaluated in the D3R Grand Challenge 2015. Considering the importance of this issue, we suggest that it should be included as a measurement of success in future editions. As shown in Table 3, the RMSD of the residues lining the binding site was below 0.4 Å in all cases, and the change in backbone conformation induced by ligand 73 could be predicted based on the DUck calculations (vide supra).

Virtual Screening

For Stage 2 of the D3R Grand Challenge 2015, we were asked to predict the affinities or affinity rankings for 180 ligands in Hsp90 and 18 ligands in MAP4K4 systems. The tools developed and used in our group are geared towards virtual screening, where we aim to identify true ligands from huge libraries of chemical compounds. As such, our predictions are fast and qualitative and not well suited to predict binding affinities, instead our goal was to produce a ranked list enriched with potent ligands in the top positions. For this reason, we only discuss the results in terms of virtual screening performance: area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve and Enrichment Factors (EF). This type of analysis could not be performed on the MAP4K4 set because 15 out of the 18 ligands were considered as active (IC_{50}< 1 µM) and the other three were in a close range (1.74, 2.25 and 10 µM). The Hsp90 set presented more dispersion: 40.6% of ligands (73 out of 180) had an IC_{50} lower than 1µM and are considered active, the remaining are considered inactive even though 21.7% (39 out of 180) have an IC_{50} between 1 and 10 µM. The fact that the inactive set contains molecules that are, a) true binders and, b) structurally very similar to the active ones makes this a very unusual and challenging test set. We encourage the organizers to include more standard virtual screening test sets in future editions of the challenge.

Our ranking protocol was based on an initial docking stage followed by DUck simulations of the top scoring pose. We combined the scores obtained from docking and DUck and, following visual inspection to check all the ligands and the corresponding rankings, the final position of 49 ligands (27%) in the ranked list was manually modified. Visual inspection introduces a subjective step that is difficult to control, but is essential in real applications to correct some of the limitations of docking. In our case, we used it mostly to rescue

compounds that were predicted as inactive because they had an incorrect binding mode (e.g. ligands binding to the helical conformation that could not fit in the docking cavity). Considering the qualitative nature of our approach, the ROC curve (Figure 6) demonstrates a very good performance, as do the corresponding enrichment factors (Table 4). To assess the effect of consensus ranking and visual inspection, we also plotted the ROC curve that would be obtained after the first stage (docking) and without the visual inspection (Figure S3). The AUC was much better for the combined ranking (0.71 vs. 0.55) and the enrichments were also higher for the combined ranking. This waste best performance across participants in this metric. Unexpectedly, we also ranked well in terms of Spearman correlation (0.39). This was surprising because both Docking and Dynamic Undocking are designed to discriminate between active and inactive compounds, rather than to obtain a quantitative assessment of their (relative) binding free energies. In part, this reflects our knowledge about this particular system, where we can anticipate from previous experience the conformational changes that take place in the protein and the ligand features that contribute to binding affinity. However, this correlation should not be considered a success, as it is likely insufficient to drive drug design. Instead it indicates that ranking ligands using structure-based methods is particularly challenging. In fact, many ligands in the test set have analogues with published binding affinity and we anticipate that a purely ligand-based strategy might have provided very good results. We suggest that the performance of one such knowledge-based approach would be useful as a benchmark of the performance of all participants in the contest..

Conclusions

Through the participation in the D3R Grand Challenge 2015, we have been able to validate the methods developed and used in our lab. We must emphasize that our main focus is virtual screening, an application that has not been considered explicitly in the challenge. Binding mode prediction is a first essential step for any subsequent prediction, so we had a particular interest on this part of the challenge. Binding affinity prediction (or ligand ranking) is much more demanding than virtual screening, and we participated in this part of the challenge somewhat reluctantly, expecting a clear underperformance compared to free energy methods.

We used a combination of qualitative techniques that, together, have worked much better than any of them separately. Namely, we used rDock for molecular docking with pharmacophoric restraints and DUck, a new technique based on molecular dynamics. For Stage 1, we were able to correctly predict how the ligands bind, particularly the position of the central scaffold forming the main interactions with the protein: for Hsp90 5 out of 6 ligands had an RMSD lower than 2 Å and 100% of the scaffolds were correctly predicted; for MAP4K411 out of 30 ligands had an RMSD lower than 2 Å and 60% of the scaffolds were correctly predicted. These figures would have increased to 18 out of 30 ligands and 77% of the scaffolds if one single additional conformation (Tyr46-IN) would have taken into account. Retrospectively, we performed additional experiments to understand the failures, finding that protein flexibility was the major factor limiting the quality of the results. Predicting protein conformations is feasible, but increasing the number of conformation generally leads to decreased docking performance [19]and even when few conformations are considered, their relative energies must be considered to avoid artifacts.[44]This is a tall order that we have by-passed by employing previous knowledge about the system, which enabled us to predict the most likely receptor conformation for each Hsp90 ligand purely based on chemical structure. The fact that we did not have this information for MAP4K4 explains the difference in performance between both systems. It should be possible to extract this type of knowledge automatically from existing crystal structures deposited in the PDB, but we are not aware of any tool capable of doing this task. Forcing certain interactions during the docking process is equally important because it corrects some of the limitations of the scoring functions. Fortunately, in this case, the main pharmacophoric points can be extracted easily and automatically with existing tools. In the absence of known ligands, binding hot spots can be identified from molecular simulations.[45]

In Stage 2, for Hsp90 we performed much better than expected considering the qualitative nature of our methods. The results were biased by our previous knowledge on this system, which had an important effect on

the final performance, but this reflects the typical situation in drug discovery, where expert users combine tools and previous knowledge whenever possible. Our relative success highlights the challenges that free energy methods are still facing, but also indicates that there is a lot of potential in combining relatively simple structure-based tools with knowledge-based approaches. No doubt, machine learning will play an increasingly important role in the future, driven both by the growing body of public data [29, 30]and major advances in the field.[31,32]

Finally, we have several suggestions to improve future editions of the challenge. Namely, the prediction of protein conformation as a measure of success in binding mode prediction, the inclusion of a virtual screening prediction set and the introduction of an automated ligand-based approach as a baseline for measuring success of ligand ranking applications. We consider that all these aspects may improve what is already an extremely useful and necessary exercise.

References

1. Barril X, Javier Luque F (2012) Molecular simulation methods in drug discovery: A prospective outlook. J Comput Aided Mol Des 26:81–86. doi: 10.1007/s10822-011-9506-1
2. Bajorath J (2015) Computer-aided drug discovery. F1000Research. doi: 10.12688/f1000research.6653.1
3. Sliwoski G, Kothiwale S, Meiler J, Lowe EWJ (2014) Computational Methods in Drug Discovery. Pharmacol Rev 61:67–75. doi: 10.1016/j.vascn.2010.02.005
4. Moult J, Fidelis K, Kryshtafovych A, et al (2014) Critical assessment of methods of protein structure prediction (CASP)--round x. Proteins 82 Suppl 2:1–6. doi: 10.1002/prot.24452
5. Janin J (2005) Assessing predictions of protein-protein interaction: the CAPRI experiment. Protein Sci 14:278–83. doi: 10.1110/ps.041081905
6. Muddana HS, Fenley AT, Mobley DL, Gilson MK (2014) The SAMPL4 host-guest blind prediction challenge: An overview. J Comput Aided Mol Des 28:305–317. doi: 10.1007/s10822-014-9735-1
7. Ferreira L, dos Santos R, Oliva G, Andricopulo A (2015) Molecular Docking and Structure-Based Drug Design Strategies. Molecules. doi: 10.3390/molecules200713384
8. Michel J, Essex JW (2010) Prediction of protein-ligand binding affinity by free energy simulations: Assumptions, pitfalls and expectations. J Comput Aided Mol Des 24:639–658. doi: 10.1007/s10822-010-9363-3
9. Steinbrecher TB, Dahlgren M, Cappel D, et al (2015) Accurate Binding Free Energy Predictions in Fragment Optimization. J Chem Inf Model 55:2411–2420. doi: 10.1021/acs.jcim.5b00538
10. Wang L, Wu Y, Deng Y, et al (2015) Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. J Am Chem Soc 150212152016002. doi: 10.1021/ja512751q
11. Chodera JD, Mobley DL, Shirts MR, et al (2011) Alchemical free energy methods for drug discovery: progress and challenges. Curr Opin Struct Biol 21:150–60. doi: 10.1016/j.sbi.2011.01.011
12. Christ CD, Fox T (2014) Accuracy assessment and automation of free energy calculations for drug design. J Chem Inf Model 54:108–20. doi: 10.1021/ci4004199
13. Shoichet BK (2004) Virtual screening of chemical libraries. Nature 432:862–5. doi: 10.1038/nature03197

14. Mobley DL, Graves AP, Chodera JD, et al (2007) Predicting absolute ligand binding free energies to a simple model site. J Mol Biol 371:1118–34. doi: 10.1016/j.jmb.2007.06.002

15. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, et al (2014) rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Comput Biol 10:e1003571. doi: 10.1371/journal.pcbi.1003571

16. Joseph-McCarthy D, Thomas BE, Belmarsh M, et al (2003) Pharmacophore-based molecular docking to account for ligand flexibility. Proteins 51:172–88. doi: 10.1002/prot.10266

17. Hindle S a, Rarey M, Buning C, Lengaue T (2002) Flexible docking under pharmacophore type constraints. J Comput Aided Mol Des 16:129–49.

18. Good AC, Cheney DL, Sitkoff DF, et al (2003) Analysis and optimization of structure-based virtual screening protocols. 2. Examination of docked ligand orientation sampling methodology: mapping a pharmacophore for success. J Mol Graph Model 22:31–40. doi: 10.1016/S1093-3263(03)00124-4

19. Barril X, Morley SD (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J Med Chem 48:4432–43. doi: 10.1021/jm048972v

20. Wright L, Barril X, Dymock B, et al (2004) Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms. Chem Biol 11:775–85. doi: 10.1016/j.chembiol.2004.03.033

21. Ruiz-Carmona S, et al. (2016) Dynamic Undocking and the Quasi-Bound State as Tools for Drug Design. Accepted

22. McGovern SL, Shoichet BK (2003) Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. J Med Chem 46:2895–907. doi: 10.1021/jm0300330

23. Verdonk ML, Mortenson PN, Hall RJ, et al (2008) Protein-ligand docking against non-native protein conformers. J Chem Inf Model 48:2214–25. doi: 10.1021/ci8002254

24. Barril X, Hubbard RE, Morley SD (2004) Virtual screening in structure-based drug discovery. Mini Rev Med Chem 4:779–91.

25. Bavi R, Kumar R, Choi L, Woo Lee K (2016) Exploration of Novel Inhibitors for Bruton's Tyrosine Kinase by 3D QSAR Modeling and Molecular Dynamics Simulation. PLoS One 11:e0147190. doi: 10.1371/journal.pone.0147190

26. Quesada-Romero L, Mena-Ulecia K, Tiznado W, Caballero J (2014) Insights into the interactions between maleimide derivates and GSK3β combining molecular docking and QSAR. PLoS One 9:e102212. doi: 10.1371/journal.pone.0102212

27. Morley SD, Afshar M (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. J Comput Aided Mol Des 18:189–208.

28. Clark M, Cramer RD, Van Opdenbosch N (1989) Validation of the general purpose tripos 5.2 force field. J Comput Chem 10:982–1012. doi: 10.1002/jcc.540100804

29. (2016) Molecular Operating Environment (MOE), 2013.08.

30. (2014) Schrödinger Release 2014-1: LigPrep, version 2.9.

31. Case DA, Babin V, Berryman JT, et al (2014) Amber 14.

32. Bayly CI, McKay D, Truchon J-F (2011) An Informal AMBER Small Molecule Force Field: parm@Frosst.

33. Kroemer RT (2003) Molecular modelling probes: docking and scoring. Biochem Soc Trans 31:980–4. doi: 10.1042/

34. Yusuf D, Davis AM, Kleywegt GJ, Schmitt S (2008) An alternative method for the evaluation of docking performance: RSR vs RMSD. J Chem Inf Model 48:1411–1422. doi: 10.1021/ci800084x

35. Warren GL, Do TD, Kelley BP, et al (2012) Essential considerations for using protein-ligand structures in drug discovery. Drug Discov Today 17:1270–1281. doi: 10.1016/j.drudis.2012.06.011

36. Cozzini P, Kellogg GE, Spyrakis F, et al (2009) Traget Flexibility: An Emerging Consideration in Drug Discovery and Design. J Med Chem 51:804–828. doi: 10.1021/jm800562d.Target

37. Spyrakis F, BidonChanal A, Barril X, Luque FJ (2011) Protein flexibility and ligand recognition: challenges for molecular modeling. Curr Top Med Chem 11:192–210. doi: 10.2174/156802611794863571

38. Barril X, Fradera X (2006) Incorporating protein flexibility into docking and structure-based drug design. Expert Opin Drug Discov 1:335–349. doi: 10.1517/17460441.1.4.335

39. Cheng LS, Amaro RE, Xu D, et al (2008) Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. J Med Chem 51:3878–94. doi: 10.1021/jm8001197

40. Abagyan R, Rueda M, Bottegoni G (2010) Recipes for the selection of experimental protein conformations for virtual screening. J Chem Inf Model 50:186–193. doi: 10.1021/ci9003943

41. Campbell AJ, Lamb ML, Joseph-McCarthy D (2014) Ensemble-based docking using biased molecular dynamics. J Chem Inf Model 54:2127–38. doi: 10.1021/ci400729j

42. Birch L, Murray CW, Hartshorn MJ, et al (2002) Sensitivity of molecular docking to induced fit

effects in influenza virus neuraminidase. J Comput Aided Mol Des 16:855–869. doi: 10.1023/A:1023844626572

43. Barril X (2014) Ligand discovery: Docking points. Nat Chem 6:560–561. doi: 10.1038/nchem.1986

44. Fischer M, Coleman RG, Fraser JS, Shoichet BK (2014) Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. Nat Chem 6:575–583. doi: 10.1038/nchem.1954

45. Álvarez-García D, Barril X (2014) Molecular Simulations with Solvent Competition Quantify Water Displaceability and Provide Accurate Interaction Maps of Protein Binding Sites. J. Med. Chem.

46. Bento AP, Gaulton A, Hersey A, et al (2014) The ChEMBL bioactivity database: An update. Nucleic Acids Res 42:1083–1090. doi: 10.1093/nar/gkt1031

47. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242. doi: 10.1093/nar/28.1.235

48. Lavecchia A (2015) Machine-learning approaches in drug discovery: Methods and applications. Drug Discov Today 20:318–331. doi: 10.1016/j.drudis.2014.10.012

49. Wale N (2011) Machine learning in drug discovery and development. Drug Dev Res 72:112–119. doi: 10.1002/ddr.20407

Tables

Table 1. **Summary of the results for the 6 ligands in the Hsp90 system Stage 1.**		
Ligand ID	**RMSD (Å)**	**Scaffold OK**
40	1.71	Yes
44	3.00	Yes
73	0.61	Yes
164	1.40	Yes
175	1.94	Yes
179	0.70	Yes
Summary	**RMSD (Å)**	**Scaffold OK**
Average	1.56	6/6 (100%)

Table 2.
Summary of the results for the 30 ligands in the MAP4K4 system Stage 1, and simulation of Stage 1 results for MAP4K4 taking into account additional conformations of Tyr36 (ligands with bad prediction only).

Lig ID	RMSD (Å)	RMSD Flexible (Å)	Scaffold OK	Lig ID	RMSD (Å)	RMSD Flexible (Å)	Scaffold OK
1	2.26	-	Yes	**17**	8.44	6.94 [b]	No / No
2	3.59	10.44[a]	No / No	**18**	1.99	-	Yes
3	1.02	-	Yes	**19**	1.85	-	Yes
4	7.16	6.14 [b]	No / No	**20**	1.29	-	Yes
5	8.90	1.13 [a]	No / Yes	**21**	0.96	-	Yes
6	2.50	-	Yes	**22**	1.47	-	Yes
7	1.03	-	Yes	**23**	2.66	1.46 [a]	Yes / Yes
8	1.43	-	Yes	**25**	3.40	3.13 [a]	Yes / Yes
9	2.49	3.55 [a]	No / No	**26**	6.77	6.93 [a]	No / No
11	0.68	-	Yes	**27**	1.29	-	Yes
12	11.06	1.42 [a]	No / Yes	**28**	1.68	-	Yes
13	5.71	1.06 [a]	No / Yes	**29**	6.34	6.58 [a]	No / No
14	4.95	1.50 [a]	No / Yes	**30**	3.83	0.52 [a]	Yes / Yes
15	2.14	-	Yes	**31**	6.64	7.25 [a]	No / No
16	5.69	4.83 [a]	No / Yes	**32**	3.09 [c]	1.32 [ac]	Yes / Yes
Summary		**Original**	**Flexible**				
Average RMSD (Å)		3.74	2.57				
Scaffold OK		18/30 (60%)	23/30 (77%)				

a) P-loop and Tyr36faced inwards to the cavity b) No hydrogen bond made with Cys108 in the hinge region c) Measured with respect to the corrected crystallographic pose

Table 3.
RMSD (Å) between binding site residues[a] of submitted Hsp90 receptor structures and crystal structure. For each of the crystal structures, the submitted PDB receptor structure is highlighted in italics.

PDB Code Submitted	**Crystal Structure**					
	40	**44**	**73**	**164**	**175**	**179**
2CCU	*0.29*	*0.30*	1.19	*0.27*	*0.25*	*0.37*
2WI6	1.05	1.05	*0.34*	1.05	1.07	1.03

[a] List of residues defining the binding site: LEU48, ASN51, SER52, ALA55, ASP93, ILE96, GLY97, MET98, ASN106, LEU107, GLY108, PHE138, TYR139, VAL150, THR152, THR184 and VAL186.

Table 4.
Summary of statistics for Hsp90 system Stage 2 results.

Ranking	**AUC[a]**	**Enrichment[b]**		
		1 %	**10 %**	**20 %**
Docking Score	0.55	1.23	0.68	1.37
Combination	0.71	2.47	1.91	1.99

[a]Max.value for AUC = 1.00 [b]Max.enrichment possible = 2.47

Figure Captions

Fig. 1 A) Hsp90 receptor definition. Asp93 and two surrounding water molecules (shown in sticks) define the key interaction element. The pharmacophoric points (transparent blue spheres) force the presence of a H-bond donor next to Asp93:OD2 and a H-bond acceptor next to the interstitial water molecule. B) MAP4K4 receptor definition. The hinge region is a characteristic binding hot spot of protein kinases. A pharmacophoric restraint forced the presence of a hydrogen bond acceptor next to Cys108:N (transparent blue sphere)

Fig. 2 Two binding modes proposed by docking for ligand 40 in the Hsp90 set (green and pink sticks; RMSD = 5.5 Å). Dashed lines represent the hydrogen bond between each ligand and Asp93. The crystal structure of ligand 40 is represented in white sticks for comparison

Fig. 3 Two different binding modes (RMSD: 2.5A) for ligand 73 in the Hsp90 set proposed by docking represented in green and pink sticks. The green one is the preferred conformation according to docking, whereas the pink one has a really bad score due to a clash penalization. With DUck, we could detect that the correct binding mode was the pink one. The crystal structure of ligand 73 is represented in white sticks for comparison, the RMSD with respect to the pink binding mode is 0.61

Fig. 4 Structure of MAP4K4 ligand MAP32 in the disclosed crystallographic structure (A) and the alternative mode we proposed (B). Note the different tautomers with inverted methyl and hydroxyl groups, where in the crystallographic pose there is a clash between the methyl group and Glu106, we found a well structured hydrogen bond between the hydroxyl group and Glu106

Fig. 5 Comparison between the two MAP4K4 supplied starting structures 4OBO (blue) and 4U44 (orange). In the former structure the side-chain of Tyr36 in the P-Loop is facing inwards, reducing the cavity space available for ligand binding

Fig. 6 ROC Curves of the 180 ligands in the Hsp90 Stage 2 Set. Ligands with an IC50 higher than 1 µM were considered as active. A)Ranking according to rDock docking scores (AUC=0.55). B)Consensus ranking as submitted to the challenge(AUC=0.71)

Figures

Fig. 1

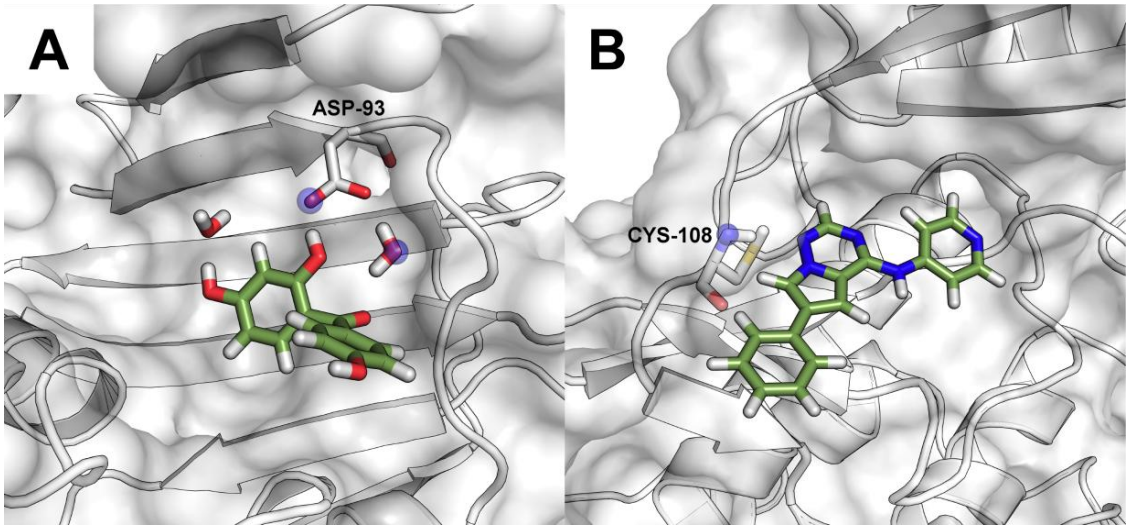


Fig. 2

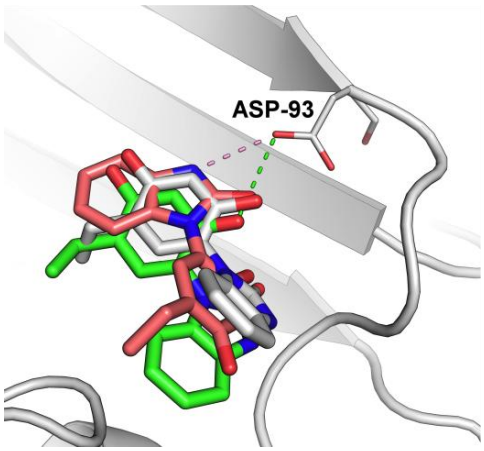


Fig. 3

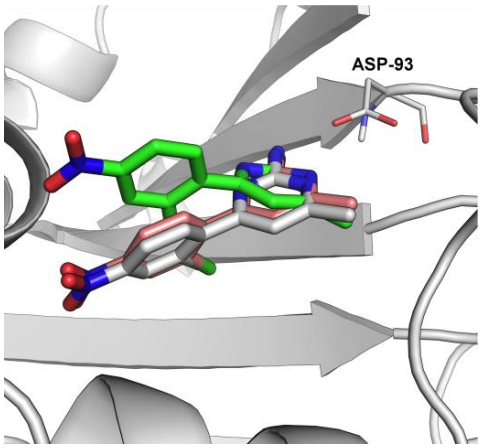


Fig. 4

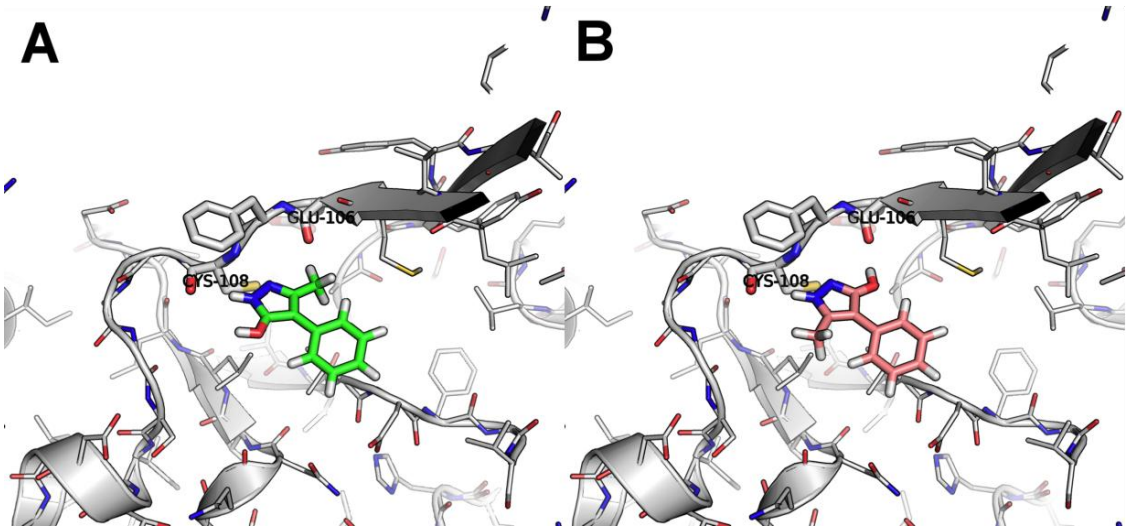


Fig. 5

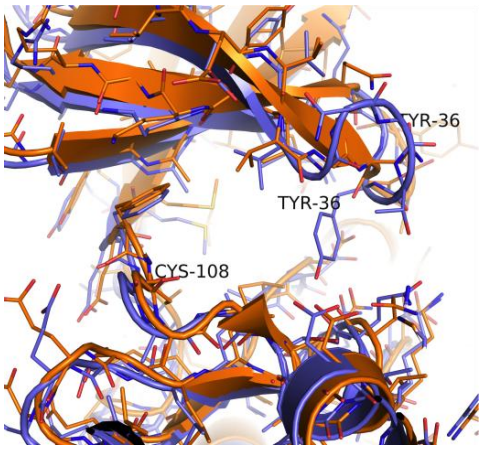
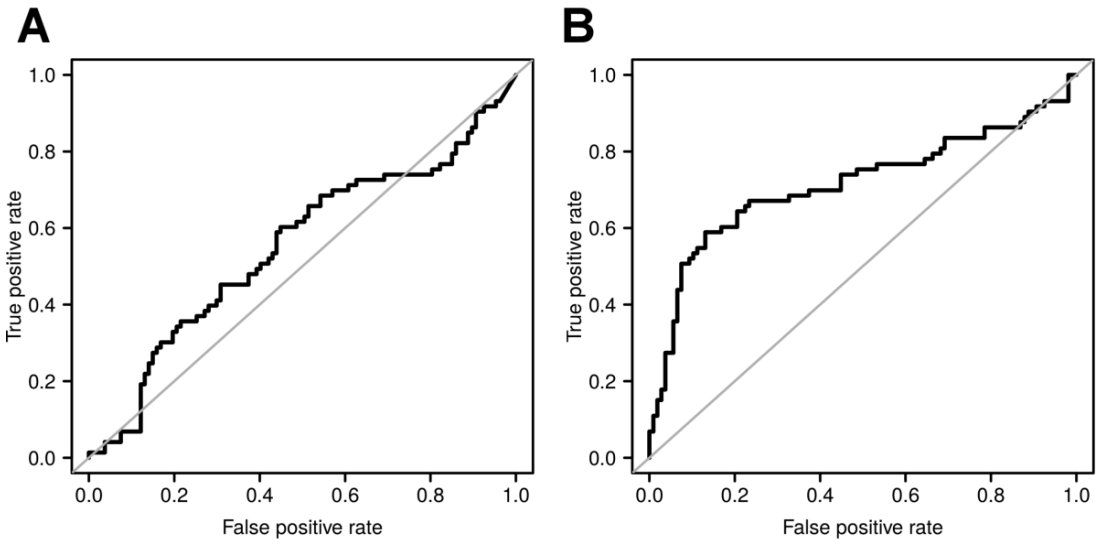


Fig. 6



Supplementary Material

Docking-undocking combination applied to the D3R Grand Challenge 2015

Sergio Ruiz-Carmona[1], Xavier Barril[1,2,*]

[1]Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain.

[2]Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

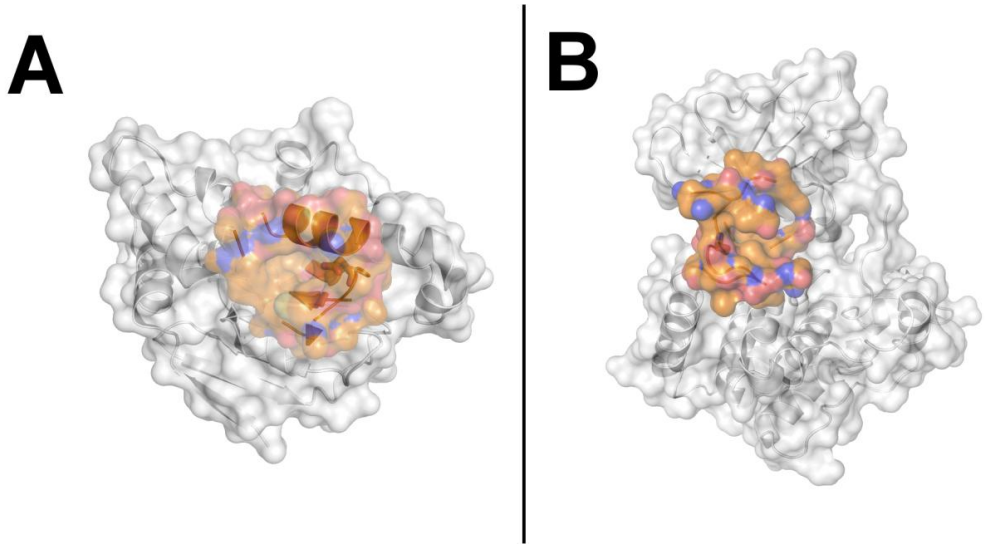


Fig.S1 A) Hsp90 protein model for Dynamic Undocking (orange) and full protein (gray). Asp93 (sticks) acts as the key interaction point. B) MAP4K4 protein model for Dynamic Undocking (orange) and full protein (gray). Cys108 (sticks) in the hinge region acts as the key interaction point. Residues forming the model binding site are listed in Table S2.

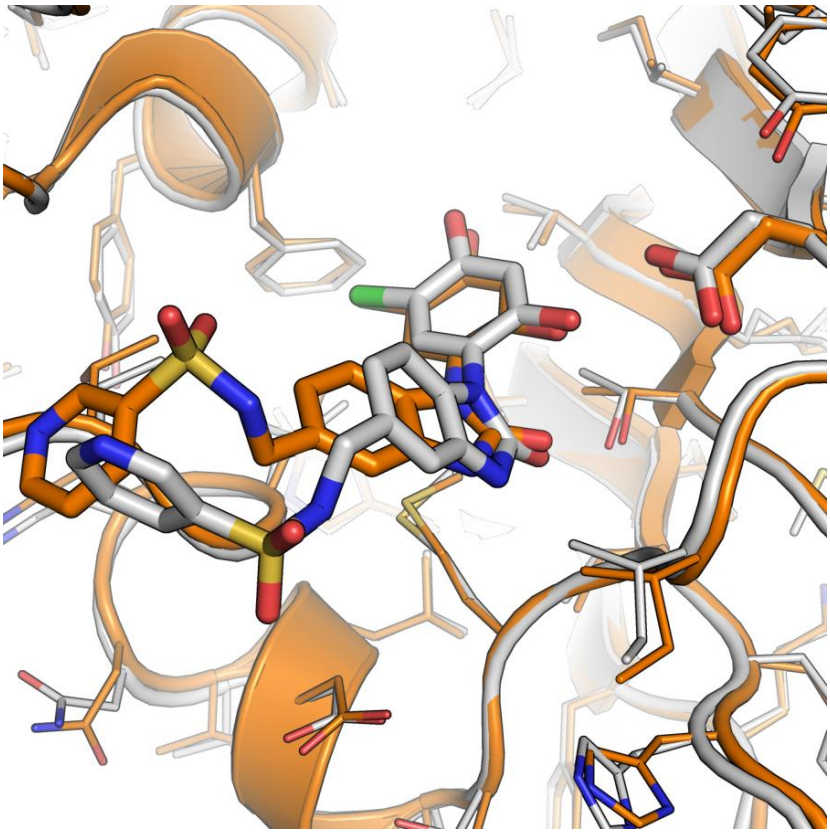


Fig. S2 Depiction of ligand 44 from the Hsp90 set to illustrate the differences between the x-ray structure (white) and the predicted binding mode (orange). As explained in the main text, the scaffold forming the main interactions is placed correctly, but the different orientation of the rest of the ligand increases the RMSD to 3.0Å.

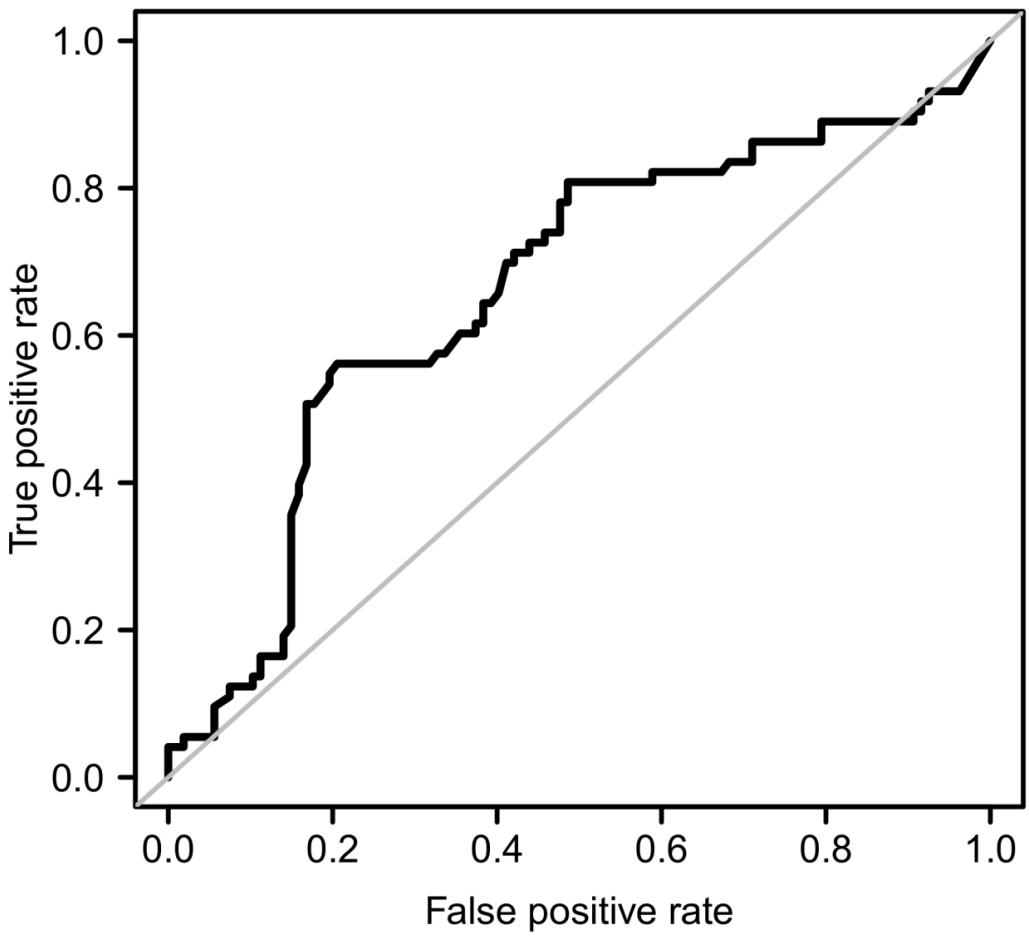


Fig. S3 ROC Curve of the 180 ligands in the Hsp90 Stage 2 Set ranked according to the consensus score considering rDock scores and W_{QB} values from DUck. Ligands with an IC50 higher than 1 µM were considered as active. The corresponding statistic values are AUC=0.66, EF1%=2.47, EF10%=1.23 and EF20%=1.37.

Table S1. Rank position of the Hsp90 ligands according to the rDock score and DUck W_{QB}. The fourth column indicates which pose was selected.

Ligand ID	rDock Ranking	DUck Ranking	Binding Mode Selected[a]
40	1	1	Yes
175	1	2	
	8	1	Yes
164	1	4	
	3	1	Yes
73	1	2	
	9	1	Yes
179	1	3	
	6	1	Yes
44	1	1	Yes

[a]When the best-scoring docking pose did not correspond to the best DUck pose, the DUck pose was prioritized.

Table S2. Receptor definition used in Dynamic Undocking simulations. Water and residue numbers were taken from the corresponding PDB file.

System	**Reference Atom**	**PDB Code (Chain)**	**Protein residues included as receptor**	**Water Molecules**
Hsp90	ASP93:OD2	4YKY (A)	GLU47 LEU48 ILE49 SER50 ASN51 SER52 SER53 ASP54 ALA55 LEU56 ASP57 LYS58 ILE78 ILE91 VAL92 ASP93 THR94 GLY95 ILE96 GLY97 MET98 GLY137 PHE138 VAL150 ILE151 THR152 GLY183 THR184 LYS185 VAL186	419 444 460 511
MAP4K4	CYS108:N	4U44 (A)	VAL31 LYS41 LEU50 ALA51 ALA52 ILE53 LYS54 ALA83 THR84 MET105 GLU106 PHE107 CYS108 GLY109 ALA110 GLY111 SER112 GLN157 VAL159 LEU160 LEU161 THR162 LYS168 VAL170 ASP171	-

Chapter 4

Results summary

4.1 rDock Molecular Docking

rDock is a molecular docking program developed at Vernalis which evolved from RiboDock [30], originally designed for docking ligands against Nucleic Acids. The program was redesigned and modified in order to be also used against Proteins and the whole docking motor was freshly implemented. It can be used for High Throughput Virtual Screening and binding mode prediction and allows the user to incorporate additional restraints and information to bias docking.

However, a lack of a deep validation and a position of rDock in the actual landscape of available docking software was needed. In this aspect, Glide (commercial) and Vina (open source) were selected to represent two currently used docking programs. The two following benchmarking experiments were planned:

- Binding Mode Prediction with ASTEX diverse set, for proteins, and RiboDock and DOCK6 sets for RNA.
- Virtual Screening Performance with DUD/DUD-E sets.

4.1.1 Preparation and Set Up

In both benchmarking experiments, the receptors were prepared the same way: The receptor structure files in DUD and ASTEX sets were processed using the Preparation Wizard tool from Maestro whereas the RNA structures were prepared using MOE. The docking cavities were defined in rDock using the "Reference Ligand" method and the coordinates were applied for Glide and Vina to ensure the least dissimilar cavities between each program.

The ligands in DUD and RNA sets were converted to smiles format and processed with Schrödinger's LigPrep software to create accurate, energy minimized 3D molecular structures and to expand tautomeric and ionization states, ring conformations and stereoisomers. For the ASTEX set, the ligands had already been manually prepared, thus processing was not needed.

The Molecular Docking protocols were defined to be the most similar as possible between each program. The exhaustiveness of all programs was set higher than default to remove uncertainty due to low sampling.

4.1.2 Virtual Screening

The DUD set, which consisted in 39 protein complexes with crystal structures, was run for Virtual Screening assessment. Each system contained an average of about 100 known active ligands and 36 per active ligand. The decoys have similar physicochemical properties to the active ligands but have different structures and shapes. The higher the active ligands are ranked with respect to the decoys, the better the enrichment. ROC curves were generated using ROCR [31] package for R with several statistical values, such as AUC and Enrichment Factors, being also calculated. On average, the best program was Glide for all the calculated metrics, with rDock performing as the average of the three tested programs (Table 4.1).

Table 4.1: Average of Virtual Screening Performance results for DUD set.

Program	AUC	logAUC	EFmax	EF 1%	EF 20%
rDock	0.69	0.26	98.7	11.4	2.5
Glide	0.78	0.37	334.6	22.6	3.2
Vina	0.66	0.24	124.3	8.9	2.2

4.1.3 Binding Mode Prediction

For the ASTEX and RNA sets, the RMSD of each predicted binding mode with respect to the crystallized ligand was calculated using Open Babel toolkit [32]. Random sets of 100 ligands were selected from all the resulting binding modes (if more than 100 ligands were available) and the percentage of the top-scored binding mode with an RMSD below 2Å for ASTEX set was calculated. For the RNA set, as the structures are more challenging than proteins (less closed cavities, less hydrophobic, featureless) and the ligands are larger and more flexible (7.8 ± 4.3 rotatable bonds vs. 5.16 ± 3.1 for the Astex set), the cut-off criterion is an RMSD below 2.5Å with respect to the crystal structure. Moreover, in order to assess whether the failures for each program were due to low sampling or bad scoring, the RMSD of all resulting binding modes (even with bad docking scores) was calculated.

These results are summarized in table 4.2, where we can see that the performance of rDock with respect to the binding mode prediction for the ASTEX set is pretty similar to the best program tested (Vina), whereas for the RNA set, rDock makes the best performance.

Table 4.2: Summary of Binding Mode Prediction results for ASTEX and RNA benchmark sets.

	% Correct in top 1 ASTEX set (% At least 1 correct pose)	**% Correct in top 1 RNA set**
rDock	76 ± 3 (99 ± 0.2)	54 ± 3
Glide	67.6 (83.8)	17.8
Vina	81.2 ± 2 (97 ± 0.5)	29 ± 2

4.1.4 Biased Docking

One of the main limitations in molecular docking is the quality of the scoring functions. It is therefore usual to introduce empirical bias, which can improve the quality of the results and also reduce the search space, thus improving performance. rDock allows the user to introduce empirical bias to guide docking in different ways: tethered template docking or pharmacophoric restraints.

Tethered template docking Tethered template docking can be used to enforce partial binding modes obtained from crystal structures of related molecules or constituent fragments, whereas the rest remains free (Figure 4.1). The template is defined by a reference bound ligand structure and a SMARTS query string defining the substructure to be tethered. rDocks's sdtether utility prealigns molecules with matching substructures with the reference substructure coordinates prior to docking and non-matching molecules are rejected. Molecules that have more than one substructure match with the query are replicated within the library of compounds to be docked, and each replicate prealigned and docked individually, thus ensuring that all possible substructure alignments are examined. For greater sampling efficiency, tethering in rDock is enforced absolutely during pose generation by restricting the randomisation and mutation functions for the tethered degrees of freedom, rather than through the use of an external penalty function.

Pharmacophoric restraints This feature ensures that pharmacophores (9 types are recognized: neutral hydrogen bond acceptor, neutral hydrogen bond donor, hydrophobic, hydrophobic aliphatic, hydrophobic aromatic, negatively charged, positively charged, and any heavy atom) are satisfied by all generated poses. Each pharmacophore restraint is defined by a combination of feature type and position, specified as a tolerance sphere with coordi-

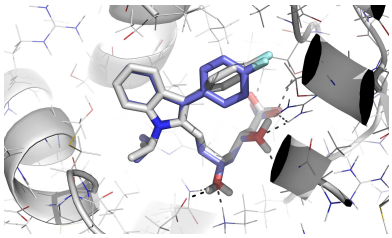


Figure 4.1: Example of tethered template docking results: while the indole moiety remains fixed, the rest of the molecule can be optimized. Blue and white sticks represent two docking solutions of a sample molecule with the indole group defined as tethered.

nate (x,y,z), and radius (r). Ligands that have insufficient quantities of the defined restraint feature types are removed prior to docking.

To illustrate this effect, we compared the outcome of a Glide and rDock Virtual Screening protocol on the DUD system Hsp90 by adding empirical bias: 2 interstitial water molecules and two pharmacophoric points were added. Two hydrogen-bonds between ASP93 and one of the waters with the ligand were defined as the pharmacophoric restraints. rDock and Glide were run with the same parameters as in the same DUD system without any pharmacophoric restraint (Vina cannot use pharmacophoric restraints) and the results were processed the same way as previously in DUD set.

As shown in Table 4.3, the metrics improved significantly and both programs became very similar in their outcome:

Table 4.3: Virtual Screening Performance results for Hsp90 with and without pharmacophoric restraints.

Program	AUC	logAUC	EFmax	EF 1%	EF 20%
rDock	0.63	0.20	3.9	0.0	1.5
Glide	0.77	0.28	7.4	0.0	2.1
Vina	0.55	0.16	1.4	0.0	0.75
rDock-Guided	0.92	0.46	36.9	12.3	4.3
Glide-Guided	0.90	0.46	17.4	6.9	4.6

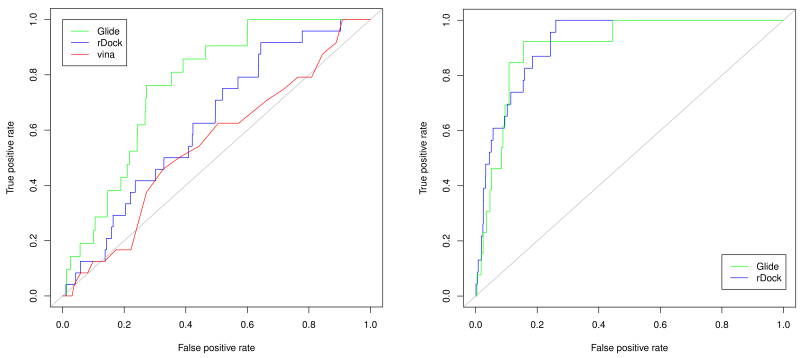


Figure 4.2: Comparison of ROC Curves for Hsp90 system using DUD set for unbiased docking (left) and docking guided using pharmacophoric restraints (right).

4.1.5 Off the record: Scoring Functions Improvement

Last but not least, in order to improve scoring functions, we tried to develop several atomic weighting factors to enhance docking performance. rDock was modified to read an extra

column in the input mol2 file with atomic weights, which changed the polar and Van der Waals contributions in the scoring stage of the docking protocol. Moreover, we also wanted to test whether the hotspots obtained running MDmix [23] could be automatically applied to rDock as "optional" pharmacophoric restraints to guide docking. The different hypotheses were tried:

- Rigid atoms should have more weight than flexible ones (from MD simulation B-factors).
- ABPAs should have more weight (calculated with fpocket)[29].
- Run MDmix and automatically extract hotspots and use them as a new type of pharmacophoric restraints to guide docking.

As a first test, we applied the approach of running MDmix to extract hotspots and automatically use them as a new type of pharmacophoric restraints in Hsp90. We ran MDmix simulations and derived up to 10 hot spots (HS) located in the binding site which were then supplied to the standard scoring function to evaluate its effect. We also studied whether the inclusion of just the HS with the lowest (and the 2 lowest ones) interaction energy already improved the results. Additionally, we compared the results with a manually-derived pharmacophore from existing ligands in the PDB (the one used in section 4.1.4). The corresponding results are found in Table 4.4 and Figure 4.3.

Table 4.4: Calculated statistic values for Hsp90 virtual screening with the different configurations detailed in the main text.

Settings	**AUC**	**EF 1%**	**EF 20%**
Unbiased	0.63	0	1
PDB-guided	0.91	0	4.3
MDmix-guided 1 HS	0.85	7.9	3.3
MDmix-guided 2 HS	0.88	19.6	3.9
MDmix-guided all HS	0.79	15.7	2.5

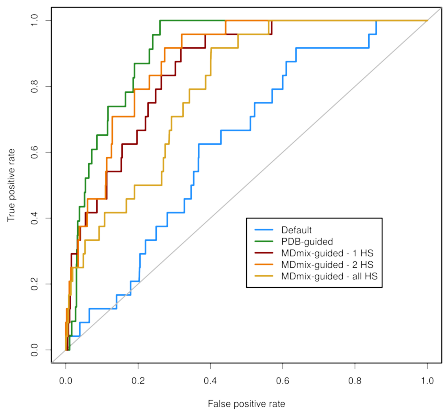


Figure 4.3: Comparison of ROC Curves for Hsp90 systems with the different settings as specified in the legend. Associated statistic values can be found in Table 4.4.

However, although we obtained promising results in this area, they have not yet been published as more experiments are needed. A collaboration with Marcelo Martin and Adrian Turjanski's group in Argentina has been established to investigate further and a publication with these results is planned for the next months.

4.2 Dynamic Undocking and the Quasi-Bound State

There is an urgent need for new methods that increase the efficacy and efficiency of rational drug design. Structure-based approaches are powerful, but they usually provide crude estimates of the binding affinity between small molecules and their target, because obtaining quantitative predictions is extremely challenging even for the most sophisticated (and computationally expensive) approaches. We postulate that, on top of presenting favourable binding free energies, true ligands should be recognizably by their structural stability, defined as the ability to maintain a precise and stable binding mode, which implies strong resistance to even minor structural deviations.

4.2.1 Background and theory

We developed a fast computational method to assess the cost of breaking key native contacts (hydrogen bonds) and demonstrated, for the first time, that structural stability is a natural property of ligands that can be exploited in virtual screening applications. This approach benefits from inexpensive calculations that have been implemented in an automated way, which makes our method amenable to high-throughput applications. Furthermore, it is an excellent complement to existing techniques, particularly molecular docking. The method, termed as "Dynamic Undocking" (DUck), calculates the work necessary to reach a quasi-bound state (W_{QB}), where the ligand has just broken the most important hydrogen bonds with the receptor (a 2.5Å displacement, from 2.5Å to 5Å), allowing us to differentiate between active and inactive ligands (Figure 4.4). DUck is complementary to day-to-day techniques such as molecular docking and it is intended to foster drug design efforts in the lead optimization stage by improving the efficiency of the in silico assessment of protein-ligand binding affinity.

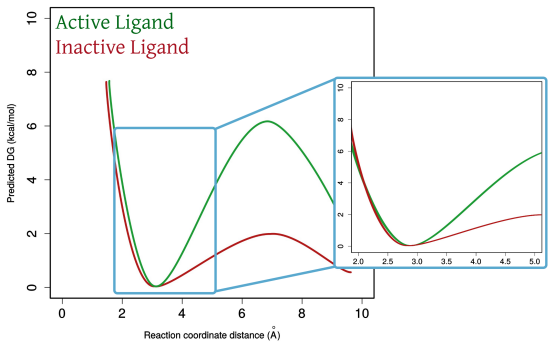


Figure 4.4: Schema of the background idea. Focusing on the first part of the disociation (highlighted plot), we are able to differenciate between active (green) and inactive (red) ligands.

In principle, the Jarzynski's equality (JE) [33] could be used to calculate Free Energy (FE) differences from non-equilibrium work distributions along a reaction coordinate, as it relates the work performed during a non-equilibrium process to the FE difference between two equilibrium states. However, in practice, as we have no guarantee that the reaction coordinate is correct, we select the lowest maximum point in the work profiles from all simulations as the W_{QB}.

Though W_{QB} is irrelevant from a thermodynamic perspective, we found that true ligands present higher W_{QB} values because their binding mode corresponds to a deep and narrow minimum in the free energy landscape.

4.2.2 Implementation

One of the main points of Dynamic Undocking is to use a reduced portion of the system instead of the whole protein including only the minimal subset of the protein necessary to preserve the local environment around the hydrogen bond that is being monitored. This transformation minimizes the influence of peripheral interactions, thus simplifying the dissociation pathway and facilitating convergence. As an added bonus, it speeds up the calculations by a factor of 5.

However, it has some drawbacks: it is not possible to recover the full dissociation profile, but most of the work needed to break a complex is used in the initial stages, where native hydrogen bonds are broken. Therefore, the goal is not to calculate the FE difference between the bound and unbound states, but to identify those ligands with higher resistance to dissociate. On the other hand, it has some important advantages: thanks to the reduction of the system size, the ligand dissociation path gets simplified which makes the choice of reaction coordinate trivial and, as the calculation speeds up accordingly to the system size reduction, it makes possible to screen hundreds to thousands of compounds. Finally, one requisite for the system to give good results is that most of the activity of the ligand must be explained by a key anchoring point, a hydrogen bond.

Molecular Docking

As a starting point, DUck needs a 3D structure in the binding site of each of the ligands generated by molecular docking (rDock and Glide in our particular case). For running docking, the receptor structures are prepared using MOE and the ligands with Schrodinger's Ligprep, as also detailed in section 4.1. In all cases, pharmacophoric restraints were used to ensure that the key interaction point was matched by every molecule in the dataset. rDock and Glide were run with the default parameters for standard docking and the best-scoring solution was accepted as the putative binding mode and used as input for DUck. Ligands that did not fulfill the pharmacophore were identified by the restraint penalty and eliminated from the dataset. Hence, all ligands that were subjected to DUck simulations fulfilled the defined interaction.

Steered Molecular Dynamics

Dynamic Undocking *per se* is a particular type of Steered Molecular Dynamics (SMD) simulations, where we force the rupture of an intermolecular hydrogen bond formed between a pre-defined interaction point in the receptor and a complementary atom in the ligand. To prepare the model receptor for the DUck simulations, all residues with at least one atom within 6Å of the atom of reference in a 3D structure of the protein of interest are selected and visually inspected to refine the selection if additional residues are considered necessary to maintain the local environment.

Once the model receptor is prepared and the set of ligands are properly oriented, a MOE SVL script developed in-house automatically calculates charges and assigns parameters to all the ligands, prepares all the input files and generates the topology and coordinate files for each ligand-receptor complex. Minimization and equilibration steps are run for each complex after which SMD simulations can be run to calculate the W_{QB}. Afterwards, unbiased MD simulations can be run to generate diverse starting points for additional SMD runs.

In the video available when scanning the QR code in Figure 4.5, you will see a visual summary of how DUck works and how the SMD pull the ligand out from the receptor until the hydrogen bond defined is broken.

Finally, work profiles outputted by the different SMD replicas are processed to calculate the W_{QB} values.

Figure 4.5: QR code for the Dynamic Undocking video summary.

MMPB/GBSA

For CDK2, MMGBSA and MMPBSA calculations using AMBER12 software were also performed and compared against the rest of methods. Each ligand was simulated for 5 ns with the full size receptor using the same MD configuration as in DUck simulations. For each one, a total of 25 snapshots separated by 200 ps were used and the free energies were averaged over the ensemble of conformations.

4.2.3 Validation Experiments

To validate our approach, retrospective and prospective validation experiments were carried on.

One one hand, we performed retrospective virtual screening against important therapeutic targets (CDK2, Trypsin and Adenosine A2A Receptor) based on the the DUD-E benchmark set. For an extra validation, BRD4 was used to study whether the experimental binding affinity and W_{QB} were correlated. As detailed in section above, different docking programs and other Free Energy methods were compared in order to demonstrate that DUck was orthogonal to the different existing methodologies.

On the other hand, we also run prospective in silico fragment screening against the ATP binding site of Hsp90. In collaboration with Vernalis, a UK-based company, different experiments were executed to confirm the fragment hits found by DUck (NMR, SPR and X-ray crystallography).

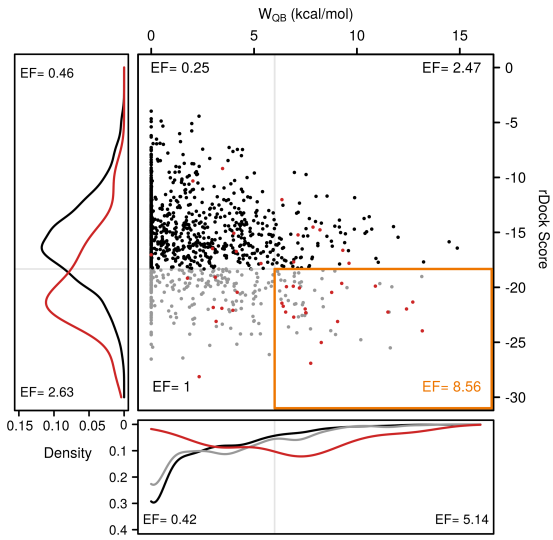


Figure 4.6: Representation of W_{QB} (X-axis) vs rDock score (Y-axis). Active ligands are shown in red and decoys in black or gray (for best 25% docking poses). The enrichment of each sector is also shown, highlighting the sector with the best enrichment factor. The lines correspond to the density functions and are coloured according to the set of the main plot they represent.

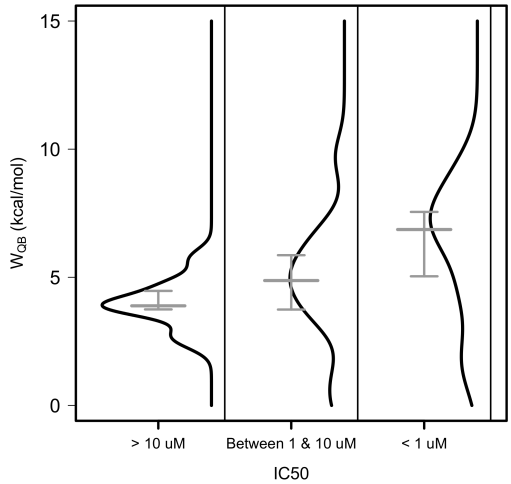


Figure 4.7: Distribution of W_{QB} values as a function of binding affinity (IC50), for the BRD4 set. Compounds with the same binding affinity present a wide distribution of W_{QB} values, but there is a tendency towards higher values for more potent compounds. Most notably, very low W_{QB} values are rare for potent ligands

Datasets

We designed datasets focusing on fragment-sized ligands because they present more scaffold diversity, make fewer peripheral interactions that could mask the main interactions and because Fragment-Based Drug Discovery (FBDD) approaches are increasingly important as hit identification strategy. For CDK2, a diverse subset from all ligands with molecular weight below 300 Da and known binding affinity (IC50) present in the PDB was selected. A pool of 30 decoys per active fragment was obtained with the DUD-E decoy generator, which puts together a set of putatively inactive molecules with physicochemical properties very similar to active ones (as in a regular DUD or DUD-e set). For BRD4, as we wanted to study the correlation between experimental binding affinity and W_{QB}, only the ligands with known binding mode and measured IC50 or KD were considered. The crystal structure of each ligand-protein complex was obtained from PDB and used as input for subsequent calculations. In the case of AA2AR, as there were few structures in the PDB, the active fragments were taken from the DUD-E benchmark set. The rest of the procedure is the same as described for CDK2. For Trypsin, we found that few ligands had a low molecular weight so we did not filter by size. Instead, a random subset of 2000 active ligands and decoys was selected from DUD-E. In the case of Hsp90, all candidate molecules come from a unified collection generated in-house from the commercial libraries of five preferred vendors (Specs, Enamine, Life Chemicals, Princeton Biomoleculars and Asinex). In this case we set an upper limit of 250 Da, obtaining 280000 candidate fragments.

Retrospective Validation

For the retrospective validation of CDK2, AA2AR and Trypsin, docking was run to generate all the starting binding modes, forcing all ligands to fulfill the defined key hydrogen bond (Leu83 for CDK2, Asn253 for AA2AR and Asp189 for Trypsin). Afterwards, all docking results were subjected to Dynamic Undocking, from where we could obtain the corresponding W_{QB} for each compound. The active fragments defined as strong binders (IC50 $< 1\ \mu M$) presented a $W_{QB} > 6$ kcal/mol, whereas more than 50% of weak binders (IC50 $> 1\ \mu M$) had a $W_{QB} < 6$kcal/mol. However, the distribution of W_{QB} values is completely different for decoys, with 61% and 49% of molecules presenting an initial dissociation cost below 2

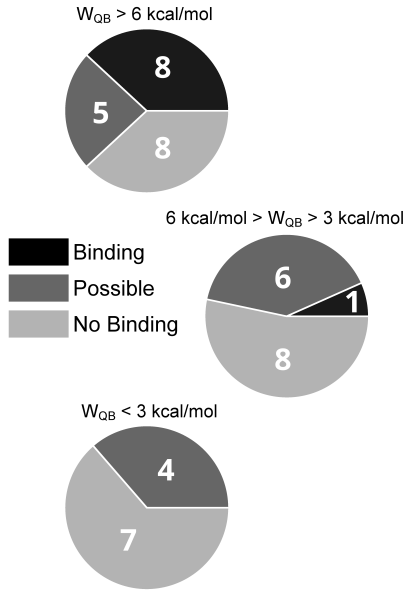


Figure 4.8: Pie charts showing the hit rates for the set of compounds with high W_{QB} -"strong" set- (top), medium W_{QB} -"medium" set- (middle) and low W_{QB} -"weak" set- (bottom). The area in black corresponds to bona fide hits, dark gray represents compounds that give a positive signal in 1 or 2 NMR experiments, pale gray corresponds to inactive compounds. Labels indicate the number of compounds of each class.

kcal/mol and 1 kcal/mol, respectively (Figure 4.6 as an example for CDK2).

For BRD4, the correlation between binding affinity (IC50) and W_{QB} was calculated. Despite not finding a good correlation, a qualitative trend was present (Figure 4.7).

Prospective Validation

Posteriorly, we applied the method prospectively for the identification of novel Hsp90 ligands. The collection of 280000 purchasable fragment-sized molecules was docked to the ATP binding site of Hsp90 with an optimized protocol, where the key hydrogen bond with Asp93 was enforced. The top scoring poses were sujected to DUck simulations and classified in three categories according to their Structural Stability: weak ($W_{QB} < 3$; N=44; 32%), medium ($3 < W_{QB} < 6$; N=67; 48%) and strong ($W_{QB} > 6$; N=28; 20%). Some of them were tested at Vernalis, finding 9 positive, 8 of them corresponding to the "strong" category (Figure 4.8), which corresponded to a 38% accuracy. Furthermore, they represented distinct chemotypes, thus demonstrating that Dynamic Undocking brings out a major improvement in VS processes and validating use of docking/undocking as a valid source of fragment hits, which can be used as a complement or alternative to general fragment screening libraries.

4.3 D3R Grand Challenge 2015

After developing and applying successfully Dynamic Undocking in section 4.2, we tested our approach by taking part in the D3R Grand Challenge 2015.

4.3.1 The Challenge

New Computer-Aided Drug Discovery (CADD) methods are constantly under development and, depending on the specific target or stage in the drug discovery process, a wide spectrum of options is available to the scientific community. Independent big-scale experiments are extremely useful to test the different methods, try them out under different circumstances and validate them for a specific goal. For instance, there have been experiments to help the development of protein structure modelling software, the prediction of protein protein interactions or the subtle characterization of small molecules properties.

The Datasets

In this direction, the D3R Grand Challenge 2015 provided an independent exercise to assess and validate ligand-protein docking algorithms and their scoring protocols. Two different datasets (Hsp90 and MAP4K4) were provided as blinded unpublished sets with high quality crystal structures and activity data for testing and improving CADD tools.

On one hand, the binding site of Hsp90 has been targeted by several drug discovery programs related with cancer over the past 20 years and consequently there is a lot of data in the PDB and literature. Nevertheless, due to the high flexibility of the receptor binding site and the -sometimes underestimated- importance of water-mediated interactions, it still persists as a challenge in drug design. The HSP90 dataset was formed by 8 crystal structures with a resolution of less than 2.0Å and binding data for 180 compounds.

On the other hand, MAP4K4 is a kinase member of the STE20 family and it is involved in metabolism, cancer and inflammation. In this case, the additional challenge was related with the flexibility of the P-loop from kinases. The corresponding dataset was formed by 30 compounds with crystal structures with a resolution of less than 2.5Å and binding data obtained by SPR for 18 of them.

The Experiments

The challenge was divided in two different stages for each of the datasets. In the first stage, the SMILES string of several ligands and different protein-ligand co-crystal structures were provided and we had to submit a prediction of the binding pose of these ligands for each protein. In the second stage, the correct structure of the ligands in the previous stage was disclosed and a bigger set of ligands was provided. The output in this stage was to predict the affinities or rankings for all the ligands.

Possible strategies

Docking, scoring and free energy methods have been widely applied in drug discovery campaigns, as they provide an excellent assistance particularly in early stages of the development of new drugs. Docking and scoring can be used for virtually assay thousands to millions of drug-like molecules in a relatively short amount of time, speeding up the finding of promising candidates and dramatically decreasing the cost in comparison with the experimental alternative. However, there are several limitations that we need to be aware of: the flexibility of drug-like ligands is currently taken into account by most docking programs but the receptor flexibility is still an unresolved issue under investigation[ref], most scoring functions

are developed based on specific training sets and they will be limited by the characteristics of these sets and, besides aiming on predicting real affinities, docking scores might be interpreted qualitatively rather than being a reliable quantitative measure.

4.3.2 Docking-Undocking Combination

In our particular approach, the main tools we used were Docking and Dynamic Undocking, which was recently developed in our group, as presented in this thesis.

Specifically, we used rDock software complemented with pharmacophoric restraints as a first approach. First of all, we identified the key interactions that the compounds should fulfill upon binding to both of the proteins. In the case of Hsp90, the interaction was a hydrogen bond with one of two oxygen atoms of Asp93 sidechain and, in case of MAP4K4, a hydrogen bond with the nitrogen atom of Cys108 in the hinge region (residues from Met105 to Gly111 acting as a hinge between the N-terminal and C-terminal domains).

Then, to address the protein flexibility, we took a knowledge-based approach. For Hsp90, we selected the most common conformation for running docking and revised its quality depending on the structure of each ligand. In contrast, MAP4K4 is a completely unfamiliar protein to us and we selected a conformation based on our previous knowledge on other kinases. Hence, this different degree of previous knowledge for each system was also a good point we discussed after finishing the D3R Grand Challenge 2015. Finally, to overcome the qualitative measures of docking, we also applied Dynamic Undocking, developed during this thesis. DUck gave us an orthogonal measure of the protein-ligand interaction and allowed us to detect not only false positives but also false negatives from docking results. As ligands were re-evaluated taking into account a different and complementary property with respect to docking, we identified good binding poses which were *apriori* discarded due to bad docking scores, overall increasing the quality of our calculations. This also helped us in predicting with high accuracy how the ligands bound to the corresponding receptor, increasing our confidence that the best binding mode predicted was the correct one.

As a summary, we performed docking-based virtual screening biased with pharmacophoric restraints and selected receptor configurations according to our degree of previous knowledge. Afterwards, we applied Dynamic Undocking to complement docking calculations and get more confident predictions rather than limiting to docking.

Binding Mode Prediction

For both of the datasets, we ran rDock to generate different binding modes for each ligand. After that, we selected a diverse subset and subjected them to DUck simulations. Taking into account both the docking score and the DUck W_{QB}, the best docking pose was selected as the most likely binding mode. The accuracy of binding mode prediction is generally measured in terms of root mean squared deviation (RMSD) from the crystallographic pose. It is also common to convert this value to a binary decision (correct/incorrect) based on a fixed threshold (usually 2.0Å). Our average results were 1.6 ± 0.9 Å for Hsp90 and 3.7 ± 2.8 Å for MAP4K4, which placed us in the 8th and 3rd position, respectively.

Ligand Ranking

For the next task, we were asked to predict the affinities or affinity rankings for 180 ligands in Hsp90 and 18 ligands in MAP4K4 systems. The tools developed and used in our group are geared towards virtual screening, where we aim to identify true ligands from huge libraries of chemical compounds. As such, our predictions are fast and qualitative and not well suited to predict binding affinities, instead our goal was to produce a ranked list enriched with

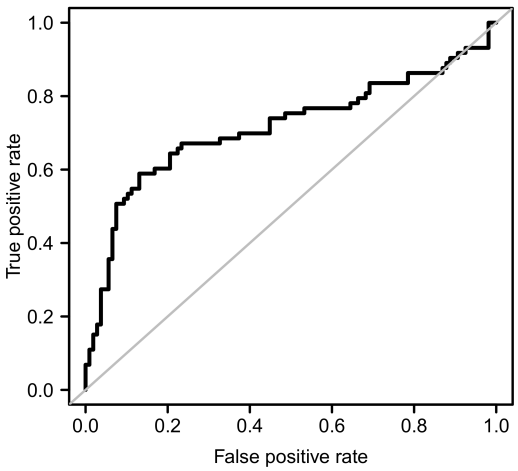


Figure 4.9: ROC Curves of the 180 ligands in the Hsp90 Ligand Ranking Set. Ligands with an IC50 higher than 1 μM were considered as active (AUC=0.71).

potent ligands in the top positions. For this reason, we only presented the results in terms of virtual screening performance: area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve and Enrichment Factors (EF). However, this type of analysis could not be performed on the MAP4K4 set because 15 out of the 18 ligands were considered as active.

Our ranking protocol was based on an initial docking stage followed by DUck simulations of the top scoring pose. We combined the scores obtained from docking and DUck and, following visual inspection, the final position of some ligands in the ranked list was manually modified.

The ROC curve (Figure 4.9) demonstrated a very good performance, as did the corresponding AUC of 0.71, which was the best performance among all participants in this metric.

Chapter 5

Discussion

In the previous chapters the papers published during this thesis were presented and the work carried on for each of them was summarized. In this chapter I will briefly discuss all the main findings and challenges that were arisen.

Molecular Docking has been used for many years and there are still several limitations that we need to be aware of: the flexibility of drug-like ligands is currently taken into account by most docking programs but the receptor flexibility is still an unresolved issue under investigation [34, 35, 36], most scoring functions are developed based on specific training sets and they are limited by the characteristics of these sets and, besides aiming on predicting real affinities, docking scores might be interpreted qualitatively rather than being a reliable quantitative measure.

Nevertheless, when comparing the different software available [37, 16, 38], we can assess their performance for different approaches in Drug Discovery. Something that is often ignored is the scalability of a given program. In the case of Glide, as it is token-based a regular user will be limited by the number of available tokens. Whereas for Vina and rDock, both can be straightforwardly parallelized to increase the speed by orders of magnitude, thus allowing small research groups or companies with limited budget, to run High Throughput Virtual Screening. One of the main issues discussed in Paper 1 is the effect of guiding docking by introducing external bias. As it is demonstrated, the results improve significantly with respect to unbiased docking. This is interesting from a wider point of view: if standard Docking-based Virtual Screening included this external bias, we could obtain better enrichments. This issue is also discussed in Paper 3, where we also used guided docking. In comparison with other groups that did not use guided docking, the results were much better.

As it was introduced in Chapter 1, challenging targets need novel methods that tackle unexplored properties of ligand-binding. However, a demonstration that these methods can be applied to less challenging targets is a necessary starting point. In the case of Dynamic Undocking, we proved that if could be effectively run in virtual screening mode for different systems, as shown in Figure 5.1.

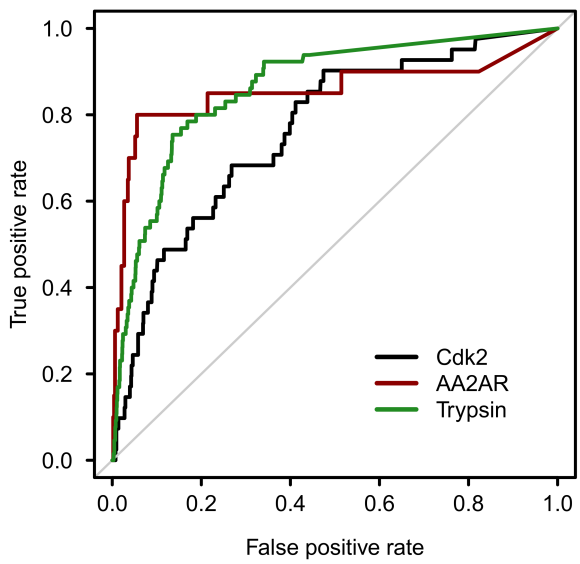


Figure 5.1: ROC Curves of the application of Dynamic Undocking for Virtual Screening of CDK2, AA2AR and Trypsin sets.

Not only that, but we also demonstrated its application as a complement to docking. Theoretically, both of them are measurements of completely different ligand properties: Docking measures overall complementarity and takes into account all possible interactions whereas

Dynamic Undocking focuses in a single interaction and calculates Structural Stability based solely on that (Figure 5.2). In principle, they should complement each other and this is actually what happens: True ligands display overall complementarity but also structural stability.

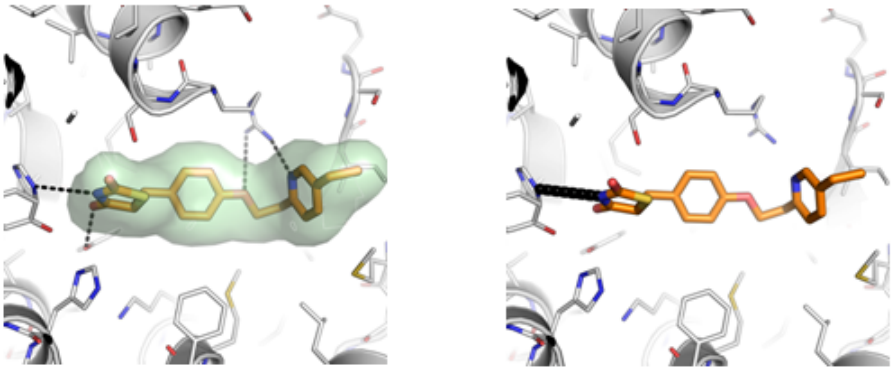


Figure 5.2: Example of the complementarity of the measurements in docking (left) and Dynamic Undocking (right).

All parameters defining DUck were validated and optimized according to our hypothesis. In order to focus on just one specific hydrogen bond, we use model receptors comprising only the protein region around this hydrogen bond. This way, we prevent the ligand to form unspecific interactions with other parts of the protein and we increase computational efficiency.

In the prospective validation with Hsp90, a hit-rate of 38% was achieved. If compared with a standard fragment screening approach for the same system [39], a hit-rate of around 4% was expected. Hence, our improvement was almost 10 fold.

Finally, we applied our approach in the D3R Grand Challenge 2015, which results have been summarized in the previous chapter.

Regarding the ranking of ligands using our approach, we need to be aware that the test set was very unusual and challenging, as the proportion of active ligands was really high (40%) and the inactive ligands were very similar to the active ones. Nevertheless, such subtle changes in structure were correctly captured by our approach and the combination of docking and Dynamic Undocking worked extremely well. Moreover, we also ranked well in terms of correlation of rankings and real measured affinities. This was surprising to us because both docking and Dynamic Undocking are designed to discriminate between active and inactive compounds, rather than to obtain a quantitative assessment of their binding free energies. In part, this reflects our knowledge about this particular system, discussed in more detail in the paper.

On the other hand, regarding the accuracy of binding mode prediction, we discussed how good are the general measurements of "correct or incorrect" in terms of RMSD with respect to the crystallographic pose [40, 41].

In practice, we think that the best measure may depend on the particular problem that one is facing. For instance, a prediction that captures the main interactions is valid when dealing with a new chemotype, but inadequate at the lead optimization stage. Since we are particularly interested in the position of the central scaffold, we took into account if the prediction is sufficiently accurate to be used in the hit progression. On Hsp90 set, we predict all but one ligand within 2.0 Å of RMSD but, looking at the central scaffold, we correctly predicted all of them [42]. In contrast, for MAP4K4 we only predicted 11 ligands correctly

(37%) using the RMSD criterium. Using the scaffold assesment, we predicted the correct position for 18 ligands (60%).

We highlighted that the reason behind the poor performance was almost exclusively due to the flexibility of the protein. There is a large body of literature discussing the importance of protein flexibility [41, 43], and newer studies have suggested that multiple-receptor docking can improve the results over a single structure [44, 45, 46].

Chapter 6

Conclusions

Global Conclusions

- Challenging targets demand better computational methods. rDock and "Dynamic Undocking" have been both the main projects in my research, whereas the application of these methods in real challenging systems has been one of the main motivations of this thesis.

- Drug candidate assays in pharmaceutical industry are very time and resources demanding. Taking advantage of a combination of computational and experimental approaches, we have been able to find active hits with a feasible amount of time and resources invested.

- We have demonstrated that instead of directly testing millions of commercially available compounds, with the consequent financial efforts, it is possible to use Virtual Screening, in our specific case rDock and "Dynamic Undocking", and supplementary chemo-informatics methods for selecting and buying only tens of compounds and perform experiments to characterize their effect.

- Orthogonal methods can be complemented and exploited to a higher success ratio. Both in the cases of "Dynamic Undocking" validation with Hsp90 and in the D3R Grand Challenge 2015 this has been studied.

Specific Conclusions

- Molecular Docking has important limitations and being aware of them and how to overcome them is key. We always run Molecular Docking with external information in order to guide it, and we have demonstrated that it is much more reliable.

- rDock has been thoroughly tested and validated with different systems. For Binding Mode Prediction, it worked almost like the best of the tested programs. In the case of Virtual Screening, it performed as average, with the advantage of high scalability. However, when supplemented with external information, it was ranked the best.

- Pharmacophoric restraints or tethered template docking can prove crucial. Choosing a docking program with such features, such as rDock, is essential for success.

- I have run rDock in every project I have been involved in this thesis. The code is maintained as open source and the interest of the scientific community is growing steadily.

- "Dynamic Undocking" allows the follow-up of a ligand unbinding (partially) event at timescales accessible to Molecular Dynamics. Since the first application of Steered Molecular Dynamics [47], it has gradually become a tool for protein-ligand recognition and interaction [48].

- The Quasi-Bound state is a state where the main interaction between ligand and receptor has just been broken and, though it is thermodynamically irrelevant, the work needed to reach it W_{QB} can be related to the Structural Stability and used for ligand ranking and Virtual Screening.

- The election of W_{QB} instead of Jarzynski Equality to calculate the work associated to the Structural Stability resides on its sensitivity to the choice of the reaction coordinate if the number of replicas is not high enough (reliable results can be obtained starting at 2 or 8 SMD simulations). Moreover, it can be calculated straightforwardly.

- Hydrogen-bonding groups in the active site are privileged structures to fix the ligand in place, particularly when they act as key binding hot spots and can form water-shielded hydrogen bonds [29].
- "Dynamic Undocking" is orthogonal to Molecular Docking. Although it is suitable as a Virtual Screening method itself, it is ideal as a post-docking filter.
- Being a novel approach, a thorough validation of all the variables, such as portion size, Steered Molecular Dynamics parameters or W_{QB} has been carried out.
- Molecular Docking with rDock and "Dynamic Undocking" were combined to participate in D3R Grand Challenge 2015, with excellent results for both Binding Mode Prediction and Ligand Ranking.
- The role of previous knowledge in a given system can result crucial for its outcome. Special remark for receptor flexibility and previous molecules that interact.

Chapter 7

Bibliography

[1] C. a. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Advanced Drug Delivery Reviews*, vol. 64, pp. 4–17, Dec. 2012.

[2] C. a. Lipinski, "Lead- and drug-like compounds: the rule-of-five revolution.," *Drug discovery today. Technologies*, vol. 1, pp. 337–41, Dec. 2004.

[3] F. Pammolli, L. Magazzini, and M. Riccaboni, "The productivity crisis in pharmaceutical R&D.," *Nature reviews. Drug discovery*, vol. 10, pp. 428–38, June 2011.

[4] A. Anderson, "The process of structure-based drug design," *Chemistry & Biology*, vol. 10, pp. 787–797, 2003.

[5] T. L. Blundell, B. L. Sibanda, R. W. Montalvão, S. Brewerton, V. Chelliah, C. L. Worth, N. J. Harmer, O. Davies, and D. Burke, "Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 361, pp. 413–23, Mar. 2006.

[6] C. N. Cavasotto and S. S. Phatak, "Homology modeling in drug discovery: current trends and applications.," *Drug discovery today*, vol. 14, pp. 676–83, July 2009.

[7] Ricardo and Macarron, "Critical review of the role of HTS in drug discovery," *Drug Discovery Today*, vol. 11, no. 7-8, pp. 277–279, 2006.

[8] T. A. Halgren, "Identifying and characterizing binding sites and assessing druggability," *Journal of Chemical Information and Modeling*, vol. 49, pp. 377–389, 2009.

[9] P. Schmidtke and X. Barril, "Understanding and predicting druggability. A high-throughput method for detection of drug binding sites.," *Journal of medicinal chemistry*, vol. 53, pp. 5858–67, Aug. 2010.

[10] S. Henrich, O. M. H. Salo-Ahen, B. Huang, F. F. Rippmann, G. Cruciani, and R. C. Wade, "Computational approaches to identifying and characterizing protein binding sites for ligand design.," *Journal of molecular recognition : JMR*, vol. 23, no. 2, pp. 209–19, 2010.

[11] N. Huang, B. K. Shoichet, and J. J. Irwin, "Benchmarking Sets for Molecular Docking," *Journal of Medicinal Chemistry*, vol. 49, pp. 6789–6801, Nov. 2006.

[12] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet, "Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking.," *Journal of medicinal chemistry*, vol. 55, pp. 6582–94, July 2012.

[13] P. D. Lyne, "Structure-based virtual screening: an overview.," *Drug discovery today*, vol. 7, pp. 1047–55, Oct. 2002.

[14] M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson, and C. W. Murray, "Diverse , High-Quality Test Set for the Validation of Protein-Ligand Docking Performance," *Journal of medicinal chemistry*, vol. 50, pp. 726–741, 2007.

[15] R. a. Friesner, J. L. Banks, R. B. Murphy, T. a. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin, "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.," *Journal of medicinal chemistry*, vol. 47, pp. 1739–49, Mar. 2004.

[16] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading," *Journal of computational chemistry*, vol. 31, no. 2, pp. 455–461, 2011.

[17] G. Schneider, "Virtual screening: an endless staircase?," *Nature reviews. Drug discovery*, vol. 9, pp. 273–6, Apr. 2010.

[18] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications.," *Nature reviews. Drug discovery*, vol. 3, pp. 935–49, Nov. 2004.

[19] E. Liepinsh and G. Otting, "Organic solvents identify specific ligand binding sites on protein surfaces," *Nature Biotechnology*, vol. 15, no. 3, pp. 264–268, 1997.

[20] C. Mattos, C. R. Bellamacina, E. Peisach, A. Pereira, D. Vitkup, G. A. Petsko, and D. Ringe, "Multiple solvent crystal structures: probing binding sites, plasticity and hydration.," *Journal of molecular biology*, vol. 357, no. 5, pp. 1471–82, 2006.

[21] O. Guvench and A. D. MacKerell, "Computational fragment-based binding site identification by ligand competitive saturation.," *PLoS computational biology*, vol. 5, p. e1000435, July 2009.

[22] J. Seco, F. J. Luque, and X. Barril, "Binding site detection and druggability index from first principles.," *Journal of medicinal chemistry*, vol. 52, pp. 2363–71, Apr. 2009.

[23] D. Álvarez García and X. Barril, "Molecular Simulations with Solvent Competition Quantify Water Displaceability and Provide Accurate Interaction Maps of Protein Binding Sites," *Journal of Medicinal Chemistry*, no. 57, 2014.

[24] K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande, "Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements.," *Journal of chemical theory and computation*, vol. 8, pp. 1409–1414, Apr. 2012.

[25] R. Abel, T. Young, R. Farid, B. J. Berne, and R. a. Friesner, "Role of the active-site solvent in the thermodynamics of factor Xa ligand binding.," *Journal of the American Chemical Society*, vol. 130, pp. 2817–31, Mar. 2008.

[26] T. Lazaridis, "Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory," *The Journal of Physical Chemistry B*, vol. 102, pp. 3531–3541, Apr. 1998.

[27] K. Haider and D. J. Huggins, "Combining Solvent Thermodynamic Profiles with Functionality Maps of the Hsp90 Binding Site to Predict the Displacement of Water Molecules.," *Journal of chemical information and modeling*, Oct. 2013.

[28] F. Colizzi, R. Perozzo, L. Scapozza, M. Recanatini, and A. Cavalli, "Single-molecule pulling simulations can discern active from inactive enzyme inhibitors.," *Journal of the American Chemical Society*, vol. 132, pp. 7361–71, June 2010.

[29] P. Schmidtke, F. J. Luque, J. B. Murray, and X. Barril, "Shielded hydrogen bonds as structural determinants of binding kinetics: application in drug design.," *Journal of the American Chemical Society*, vol. 133, pp. 18903–10, Nov. 2011.

[30] S. D. Morley and M. Afshar, "Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock.," *Journal of computer-aided molecular design*, vol. 18, pp. 189–208, Mar. 2004.

[31] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer, "ROCR: visualizing classifier performance in R.," *Bioinformatics (Oxford, England)*, vol. 21, pp. 3940–1, Oct. 2005.

[32] N. M. O'Boyle, M. Banck, C. a. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel: An open chemical toolbox.," *Journal of cheminformatics*, vol. 3, p. 33, Jan. 2011.

[33] C. Jarzynski, "Nonequilibrium Equality for Free Energy Differences," *Physical Review Letters*, vol. 78, pp. 2690–2693, Apr. 1997.

[34] S. L. McGovern and B. K. Shoichet, "Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes.," *Journal of medicinal chemistry*, vol. 46, pp. 2895–907, July 2003.

[35] M. L. Verdonk, P. N. Mortenson, R. J. Hall, M. J. Hartshorn, and C. W. Murray, "Protein-ligand docking against non-native protein conformers.," *Journal of chemical information and modeling*, vol. 48, pp. 2214–25, Nov. 2008.

[36] X. Barril, R. E. Hubbard, and S. D. Morley, "Virtual Screening in Structure-Based Drug Discovery," *Mini Reviews in Medicinal Chemistry*, vol. 4, no. 7, pp. 779–791, 2004.

[37] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, a. B. Garmendia-Doval, S. Juhos, P. Schmidtke, X. Barril, R. E. Hubbard, and S. D. Morley, "rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids.," *PLoS computational biology*, vol. 10, p. e1003571, Apr. 2014.

[38] T. a. Halgren, R. B. Murphy, R. a. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, and J. L. Banks, "Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening.," *Journal of medicinal chemistry*, vol. 47, pp. 1750–9, Mar. 2004.

[39] I.-J. Chen and R. E. Hubbard, "Lessons for fragment library design: analysis of output from multiple screening campaigns.," *Journal of computer-aided molecular design*, pp. 603–620, June 2009.

[40] A. C. Good and D. L. Cheney, "Analysis and optimization of structure-based virtual screening protocols (1): Exploration of ligand conformational sampling techniques," *Journal of Molecular Graphics and Modelling*, vol. 22, no. 1, pp. 23–30, 2003.

[41] X. Barril and S. D. Morley, "Unveiling the full potential of flexible receptor docking using multiple crystallographic structures.," *Journal of medicinal chemistry*, vol. 48, pp. 4432–43, June 2005.

[42] L. Wright, X. Barril, B. Dymock, L. Sheridan, A. Surgenor, M. Beswick, M. Drysdale, A. Collier, A. Massey, N. Davies, A. Fink, C. Fromont, W. Aherne, K. Boxall, S. Sharp, P. Workman, and R. E. Hubbard, "Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms," *Chemistry and Biology*, vol. 11, no. 6, pp. 775–785, 2004.

[43] C. F. Wright, K. Lindorff-Larsen, L. G. Randles, and J. Clarke, "Parallel protein-unfolding pathways revealed and mapped.," *Nature structural biology*, vol. 10, pp. 658–62, Aug. 2003.

[44] L. S. Cheng, R. E. Amaro, D. Xu, W. W. Li, P. W. Arzberger, and J. A. McCammon, "Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase," *Journal of Medicinal Chemistry*, vol. 51, no. 13, pp. 3878–3894, 2008.

[45] R. Abagyan, M. Rueda, and G. Bottegoni, "Recipes for the selection of experimental protein conformations for virtual screening," *Journal of Chemical Information and Modeling*, vol. 50, no. 1, pp. 186–193, 2010.

[46] A. J. Campbell, M. L. Lamb, and D. Joseph-McCarthy, "Ensemble-based docking using biased molecular dynamics," *Journal of Chemical Information and Modeling*, vol. 54, no. 7, pp. 2127–2138, 2014.

[47] B. Isralewitz, S. Izrailev, and K. Schulten, "Binding pathway of retinal to bacterio-opsin: a prediction by molecular dynamics simulations.," *Biophysical journal*, vol. 73, pp. 2972–9, Dec. 1997.

[48] M. Sotomayor and K. Schulten, "Single-molecule experiments in vitro and in silico.," *Science (New York, N.Y.)*, vol. 316, pp. 1144–8, May 2007.

Chapter 8

Appendix

During this thesis I also participated in other projects and collaborations. As a special remark, here you can find one of the projects that are closely related with this thesis. Despite not being published yet, it is in preparation and will be submitted in the following months.

8.1 Rational design of Protein-Protein Interface binders with "Glue" Activity

In this project, we applied all the methodologies available in our lab at its corresponding time in order to rationally design ligands that are able not only to bind in a Protein-Protein Interface (PPI), but also to increase the complex affinity, acting as a "glue" molecule.

More specifically, what we proposed is a novel alternative mechanism of action consisting in stabilizing specific protein complexes using drug-like molecules, which should bind at the dimer interface. As long as these drug-like molecules present much higher affinity for the complex than for any of the monomers, they will act as 'glue', over-stabilizing the complex.

In particular, after running fpocket to the whole PDB and MDmix to a selected subset, a complex formed by *E.Coli* proteins CheA and CheY was been selected for validating the rational design of PPI stabilizers by a combination of both computational and experimental approaches.

Molecular docking, chemoinformatic tools, molecular dynamics simulations, X-ray chrystallography, SPR and MST are the different techniques applied to discover several drug-like molecules acting with the desired effect and, as proof-of-concept, we were able to demonstrated our hypothesis.

Rational Design of Protein-Protein Interfacial Binders with "Glue" Activity: *E.Coli* CheA-CheY Complex as Proof-of-Concept

Sergio Ruiz-Carmona[†,1], **Montserrat Pujadas**[†,1], **Daniel Alvarez-Garcia**[1], **Peter Schmidtke**[1], **Jesus Seco**[1] **and Xavier Barril**[1,2]

In Preparation

1 Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Facultat de Farmacia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain, **2** Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.
[†] These authors contributed equally to this work.

Rational Design of Protein-Protein Interfacial Binders with "Glue" Activity: *E.Coli* CheA-CheY Complex as Proof-of-Concept

Sergio Ruiz-Carmona[†,1], Montserrat Pujadas[†,1], Daniel Alvarez-Garcia[1], Peter Schmidtke[1], Jesus Seco[1] and Xavier Barril [1,2,*]

[1] Institut de Biomedicina de la Universitat de Barcelona (IBUB) and Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain.

[2] Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

[†] These authors contributed equally to this work.

* Send correspondence to: xbarril@ub.edu

Abstract

Paradoxically, as the financial resources invested in drug discovery and our knowledge are exponentially increasing, the number of new chemical entities released to the market and the number of novel targets remain constant year after year. Hence, it is very necessary to expand the "druggable genome" and, therefore, increase the productivity of the pharmaceutical industry. Here we present protein-protein interfaces as a real option to deal with this issue.

We have used the combination of computational and experimental approaches to find potentially active compounds that bind to protein-protein interfaces. Molecular docking program rDock has been used to perform Virtual Screening and to select a small subset of possible interface binders. These compounds have then been tested experimentally using Biacore technology and three hits were obtained. Two of them were inhibiting the formation of the complex and one of them was increasing the affinity of both proteins, in other words, it was acting as glue.

There is not any deliverables in this publication. All the data used in any moment are publicly available online.

Introduction

It is known that the productivity of the pharmaceutical industry, as the financial resources invested are growing and the number of new chemical entities remain constant, is decreasing year after year. This seems to be explained by different reasons, such as unpredicted toxicity or lack of demonstrated efficacy [1], for instance. In relation with this issue, it is surprising that the number of new pharmacological targets per year has remained constant for the last 20 years [2,3].

Hence, expanding the "druggable genome", which is defined as the subset of genes in the human genome that express proteins suitable for binding drug-like molecules, would be a key step in order to increase the productivity of the pharmaceutical industry and the discovery of new chemical entities. On the other hand, the currently "one target-one disease" model is being challenged by a more complex point of view thanks to systems biology and functional proteomics, which study the interrelation between components of biological systems to obtain a full view of cells, organisms or pathways. According to this new paradigm, the goal of drug therapy has shifted from acting in individual components towards altering the equilibrium of a whole system [4].

In accordance with the two last ideas, the development of protein-protein interaction inhibitors has received considerable attention but, as the interacting surface is usually buried and polar, inhibitors are hard to find and rarely valid as drug candidates [5]. As this kind of targets is not being currently exploited [6], this is a very challenging alternative for rational selection of targets and drug candidates. Accordingly, we propose a novel alternative mechanism of action consisting in stabilizing specific complexes using drug-like molecules which could bind at the interface between two monomers. As long as these drug-like molecules present much higher affinity for the complex than for any of the individual monomers, they will act as "glue", over-stabilizing the complex.

All the structures in the PDB have been scanned and tens of druggable binding sites located at the interface of protein complexes were found. After studying several of these interfacial systems using

our own methods[7, 8], one protein-protein complex was chosen based on its properties and its feasible validation. The chosen complex was formed by E.Coli chemotaxis proteins CheA and CheY, two proteins involved in bacterial motility. CheA has an interaction domain for CheY (PDB structure 1a0o) which involves different phosphorilation affinities depending on the environment that alter motility patterns [9]. CheA has a molecular weight of more than 70.000 Da, whereas its binding site for CheY and CheY itself have similar molecular weights smaller than 20.000 Da. Hence, in order to facilitate all the work, we decided to use the described binding domain of CheA for CheY instead of the CheA full-length. This was a good system because it was easy to produce and purify the proteins needed for assaying the activity of the selected compounds *in vitro*, taking advantage of SPR technology, and *in vivo* by means of a motility assay.

We plan to validate this novel alternative by a combination of both computational and experimental approaches. First by selecting tens of commercially available compounds using Virtual Screening and then by testing them using experimental methods.

Methods

Molecular Docking using rDock software

Molecular docking is a computational technique widely used to predict the binding mode of a complex formed by one or more molecules [10. 11], in our case, one ligand binding to a protein-protein complex interface. In high-throughput screening, millions of compounds are experimentally tested for activity using miniaturized assays [12]. In an analogous manner, Virtual Screening techniques allow to detect potentially active compounds from a huge list of candidates and test them experimentally [13].

Here, we have used rDock as the software to perform the molecular docking. It provides a package of software to facilitate the preliminary tasks and posterior analysis steps of docking results [14].

However, the actual implementation of the method is slightly different from the work presented by Morley and Afshar and, as it is in preparation for publication, here we have summarized the main issues.

The master scoring functions that rDock implements is equation 1. It is formed by a weighted sum of intermolecular (equation 2), ligand intramolecular (equation 3), site intramolecular (equation 4) and

restraint (equation 5) terms.

$$\mathbf{S^{total}} = S^{inter} + S^{intra} + S^{site} + S^{restraint} \quad (1)$$

S^{inter}, S^{intra}, S^{site} represent the cavity-ligand interaction score, the relative score of the ligand conformation and the relative score of the active site, respectively. The three equations (below these lines), are formed by a set of constituent potentials.

$$\begin{aligned}\mathbf{S^{inter}} = {} & W^{inter}_{vdw} \cdot S^{inter}_{vdw} + W^{inter}_{polar} \cdot S^{inter}_{polar} + \\ & + W^{inter}_{repul} \cdot S^{inter}_{repul} + W^{inter}_{arom} \cdot S^{inter}_{arom} + \\ & + W_{solv} \cdot S_{solv} + W_{dot} \cdot N_{rot} + W_{const}\end{aligned} \quad (2)$$

$$\begin{aligned}\mathbf{S^{intra}} = {} & W^{intra}_{vdw} \cdot S^{intra}_{vdw} + W^{intra}_{polar} \cdot S^{intra}_{polar} + \\ & + W^{intra}_{repul} \cdot S^{intra}_{repul} + W^{intra}_{dihedral} \cdot S^{intra}_{dihedral}\end{aligned} \quad (3)$$

$$\begin{aligned}\mathbf{S^{site}} = {} & W^{site}_{vdw} \cdot S^{site}_{vdw} + W^{intra}_{polar} \cdot S^{site}_{polar} + \\ & + W^{site}_{repul} \cdot S^{site}_{repul} + W^{site}_{dihedral} \cdot S^{site}_{dihedral}\end{aligned} \quad (4)$$

$S^{restraint}$ stands for the score of several functions involving the default and user-defined restraints that can be used to alter the docking runs in several useful ways, for instance, highly penalizing those poses of the ligand that are trying to go away from the defined docking cavity.

$$\begin{aligned}\mathbf{S^{restraint}} = {} & W_{cavity} \cdot S_{cavity} + W_{nmr} \cdot S_{nmr} + \\ & + W_{tether} \cdot S_{tether} + W_{ph\,4} \cdot S_{ph\,4}\end{aligned} \quad (5)$$

In our Virtual Screening, we have defined the docking cavity where the previous druggability studies found at the interface of both CheA and CheY. We also added two pharmacophoric restraints that were

also found in previous druggability studies, one hydrogen bond acceptor involving the NH group of CheY TYR10, and one hydrophobic region, stacking with CheY ILE12. The system is represented in figure 4. The receptor flexibility was accepted within 3.0 Å of the docking cavity, which means that terminal OH and NH3+ groups will be treated as flexible.

The compounds used to perform the Virtual Screening were exported from our own database of commercially available compounds. This database has almost 5 million non-redundant compounds from 13 different vendors. All the compounds have been prepared using ligprep software from Schrödinger, which converts from two-dimensional to three-dimensional structural data. The compounds generated have been limited to have, at most, 8 stereoisomers and 6 tautomers. Moreover, when more than one ionization state is present (in a proportion higher than 10%) in a pH range between 6.0 and 8.0, both structures are kept.

We exported a set of about 1.200.000 compounds, which accomplished several conditions: the number of violations of Lipinski's and Veber's rules and the number of reactive atoms was 0, the number of conformational states was less than 4 and in a proportion higher than 0.1 and the number of hydrogen-bond donor and acceptor was higher than 1.

Finally, we performed the molecular docking using a High Throughput Virtual Screening protocol for rDock, which indicated how many steps to run for each of the ligands docked and the energy cutoff for selecting the best poses. We have defined this protocol to run three steps: first, make 5 runs and check if the S^{inter} value (we have selected this term of the scoring function as the reference one because we are comparing the interaction between thousands of ligands and the protein-protein interface) is smaller than -15. Second, if any pose passes the first filter, then run 15 more steps and check again if S^{inter} is smaller than -20. And, the last step, if any pose has passed the last two filters, run up to 50 steps and write the resulting poses that have not moved away from the docking cavity to the output.

Surface Plasmon Resonance

To validate the computational experiment results, an experimental confirmation was needed. The chosen technique to do so was Surface Plasmon Resonance (SPR). SPR technique is a method that, using optical measures, can determine in real time the interaction of two or more molecules by

comparing the refractive index of material adsorbed on a metal, usually gold. The material adsorbed has to be forming very thin layers. Then, we can straightforwardly transform this response into interaction affinities [15]. Applying this, we can assess the interaction between small molecules and proteins [16]. In our case, it was slightly different, as we did not work with one protein but with the complex formed by two proteins.

We used Biacore technology with CM5 chips. All the buffers used were HBSN, that contains 10mM of HEPES and 150 mM of NaCl in water, with 1% of DMSO at pH 7.4. The temperature was always set to 25 Celsius degrees. Both proteins included a His-tag in order to make the production and purification stages easier.

For the covalent immobilization on the chip, the protein chosen was CheY, as it gave us better response in first tests. To immobilize it, we ran a protocol consisting in three stages: first, 7 minutes of a EDC/NHC 1:1 solution for activating the surface; second, Chemotaxin Y during as many time as needed to reach about 120 of RU fixed; and third, 7 minutes of Ethanolamine solution for deactivating the excess of reactive groups (these activating and deactivating solutions were provided within Biacore consumables). All these steps were performed using a flow rate of 10 μL/min.

For compound screening, we used the previously selected compounds to determine a reproducible and reliable experimental protocol. It consisted in adding a mixture of CheA and one compound, which allowed us to determine the effect of each compound by comparing its response with a reference with only CheA. All the interaction steps were run using a flow rate of 30 μL/min during 60 seconds. Two regeneration stages were needed after running every interaction step. The first one with SDS 0.25% during 60 seconds and the second one with NaOH 0.1 M during 30 seconds (also using a flow rate of 30 μL/min).

Results

<u>Virtual Screening</u>

Once the docking job was finished, we merged all the output results and get a total of more than 64.000 docking poses with the S^{inter} score smaller than -20 (the minimum and mean scores were -30.69 and -21.28, respectively).

164 Then, we applied several filtering steps in order to decrease the number of compounds and facilitate a
165 final manual inspection to select the most likely ones to be actives. We filtered out the poses with
166 $S^{restraint}$ higher than 0.5 and selected the pose with smallest S^{inter} of each ligand. As we had used a
167 pharmacophoric restraint involving one hydrogen-bond in the active site, we decided to divide all the
168 resulting ligands in different subclasses depending on the chemical group which was fulfilling the
169 restraint. Afterwards, the resulting groups were clustered by fingerprint similarity and a last manual
170 inspection was performed to make a selection the most interesting ones that spanned as many
171 chemical groups as possible. The evolution of the whole filtering process is represented in table 1.
172 The selection of the compounds was performed with two main objectives: first, we selected the
173 structural patterns that were present in the binding pocket and accomplished the defined restraints in
174 order to obtain a set as diverse as possible, but also maintaining the representativeness of the selected
175 compounds. Then, we performed a clustering filter and manually inspected all the centroids.
176 This way, we obtained a set of more than 50 compounds that was diverse and was formed by
177 molecules that represented almost all the output compounds selected in the Virtual Screening process.
178
179 Experimental Validation
180 *Kinetics of CheA and CheY complex*. The immobilized amount of CheY in this experiment was about 120
181 RU. The two molecular weights of the immobilized protein (CheY) and binding protein (CheA) are
182 15.000 and 18.000 Da, respectivelly. The theoretical Rmax is about 145 RU whereas the predicted
183 one is about 130 RU. Figure 1 depicts the response during contact time with a solution of CheA from
184 0 to 50 μM. Moreover, we predicted the kD for the protein-protein complex, being about 3.7 μM. In
185 figure 2, the fitted curve for kD prediction is represented.
186
187 *Compounds screening*. We tested all the compounds by adding a concentration of 100 μM of each one
188 with CheA at a concentration of 1.25 μM.
189 We found two compounds that seemed to inhibit the interaction of the complex (compounds I1 and
190 I2). Adding these compounds one by one in combination with CheA, made the response decrease to
191 almost zero, as can be seen in figure 2. Moreover, we also found one compound that was increasing the

affinity for the interaction of both proteins (compound G1). In figure 2, it is represented the increasing of the response when combining this compound and CheA.

The binding mode of these three compounds when binding at the protein-protein interface was predicted in the molecular docking stage. In figure 4, the two-dimensional structures of each compound and the predicted binding mode of the stabilizing compound are represented.

Discussion

Using computational methods, we reduced the number of compounds to be assayed from more than one million to less than one hundred. This has been done, first, by performing Virtual Screening and, second, by several filtering stages based on the diversity of the chemical groups that fulfil the hydrogen bond pharmacophoric restraint. The final selection, both clustering by fingerprint and manual steps, was performed with the intention of maintaining the proportion of each of the classes selected. As seen in table 1, the compounds selected and tested posteriorly spanned all the interesting chemical patterns. This way, we ensured that diversity was such that we could obtain as many active hits classes as possible, facilitating further structure-based drug design studies.

In the experimental approach, two main tasks were performed. First of all, we characterized the interaction between the two proteins. The response increases as the concentration of CheA also increases, reaching a saturation maximum at about 145 RU, when the concentration of CheA is 50 μM. The kD predicted is about 3.7 μM, in the same order than previous studies [17]. Second, and most important, we performed the screening of the compounds selected in the computational approach. 68 compounds were tested, from which 3 of them were producing significant effects to the complex interaction. 2 out of this 3 compounds were decreasing the affinity, whereas the other one was increasing it. We also run the experiments without CheA in order to ensure that the compounds were interacting at the interface. As shown in figures 2 and 2, they had no effect by themselves. Surprisingly, we found two compounds that were inhibiting the interaction of the complex instead of the expected effect, the stabilization. In previous studies has been shown that small modifications in ligands may produce the opposite effect [18]. Hence, we are planning to repeat these tests with analogous compounds trying to enhance complex stabilization.

Moreover, we also tried to check the binding of all the compounds with individual proteins. As the fixed protein to the chip was CheY, we could not assess the effect of the compound when binding to CheA. Hence, we performed experiments adding only one compound at each step to the fixed CheY and we did not see any significant effect, although it could be due to low sensitivity of SPR technology with such differences in molecular weight. Thus, more experiments in this direction are needed to explain these changes in the expected activity.

In the case of the stabilizing compound G1, we also going to run more tests with analogous compounds to try to stablish a structure-activity relationship. In figure 4, it is possible to see the fixed structures of each compound for this structure-activity relationship studies.

We also made several dose-response tests with the three compounds, I1, I2 and G1. In all cases, the affinity of the complex returned back to the values without any compound. On the other hand, when increasing the concentration of G1, the response achieved a maximum corresponding to the predicted Rmax (data not shown). For example, in the case of figure 2, the RU observed when having both the CheA and the compound were about 140, while the expected RU for that specific concentration of CheA was about 60. This was in concordance with the experiment for determining the kD, where the predicted Rmax was about 130 RUs.

Finally, the predicted binding mode energy for I1, I2 and G1 was -23.44, -20.14 and -23.17, respectively.

Conclusions

Drug candidate assays in pharmaceutical industry are very time and resources demanding. Here, we have taken advantage of a combination of computational and experimental approaches to perform a feasible way of finding active hits. We demonstrated that instead of directly testing millions of commercially available compounds, with the consequent financial efforts, it is possible to use Virtual Screening and chemo-informatics methods for selecting and buying tens of compounds and perform experimental tests to characterize their effect.

Despite the difficulty of finding protein-protein interface inhibitors [19], we found three protein-protein interface binders from the total 68 compounds tested. Two compounds decreased

the affinity of the complex and one increased it.

Finally, and most important, with this proof-of-concept we have also confirmed protein-protein interfaces as possible novel drug targets. Rational design of these targets could become a real option in present and future of drug-discovery, from binding pocket detection and druggability prediction to drug-target interaction and active hits selection. Thus, as our knowledge about protein-protein complexes and crystal structures is increasing, we believe that it would indeed be possible to increase productivity of drug-design industry.

As stated in some recent articles [6], protein-protein interfaces are not currently being exploited in drug-design. We are planning to also explore interfaces between domains in multi-domains proteins, between different protein subunits in future studies.

As a concluding remark, it should be noticed that further validation studies are planned to be carried out. Molecular dynamics simulations will be performed with the stabilizing compound to see if the predicted binding mode is indeed correct and *in vitro* experiments for obtaining the crystal structure of the complex with the compound will also be done.

Acknowledgements

We thank Marta Taulés for her help and suggestions in Biacore experiments. We thank the Spanish Ministerio de Ciencia e Innovación and Universitat de Barcelona for financial support.

References

[1] G. A. Fitzgerald, "Anticipating change in drug development: the emerging era of translational medicine and therapeutics," *Nat Rev Drug Discov*, vol. 4, no. 10, pp. 815–818, 2005.

[2] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins, "How many drug targets are there?," *Nat Rev Drug Discov*, vol. 5, no. 12, pp. 993–996, 2006.

[3] A. L. Hopkins and C. R. Groom, "The druggable genome," *Nat Rev Drug Discov*, vol. 1, no. 9, pp.

275 727–730, 2002.
276
277 [4] L. Hood and R. M. Perlmutter, "The impact of systems approaches on biological problems in drug
278 discovery," *Nat Biotech*, vol. 22, no. 10, pp. 1215–1217, 2004.
279
280 [5] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Hot spots. A review of the protein-protein
281 interface determinant amino-acid residues," *Proteins*, vol. 68, no. December 2006, pp. 803–812,
282 2007.
283
284 [6] D. C. Swinney and J. Anthony, "How were new medicines discovered?," *Nature reviews. Drug*
285 *discovery*, vol. 10, pp. 507–19, July 2011.
286
287 [7] P. Schmidtke, A. Bidon-Chanal, F. J. Luque, and X. Barril, "MDpocket: open-source cavity detection
288 and characterization on molecular dynamics trajectories.," *Bioinformatics (Oxford, England)*, vol. 27,
289 pp. 3276–85, Dec. 2011.
290
291 [8] J. Seco, F. J. Luque, and X. Barril, "Binding site detection and druggability index from first
292 principles.," *Journal of medicinal chemistry*, vol. 52, pp. 2363–71, Apr. 2009.
293
294 [9] H. Thakor, S. Nicholas, I. M. Porter, N. Hand, and R. C. Stewart, "Identification of an Anchor
295 Residue for CheA-CheY Interactions in the Chemotaxis System of Escherichia coli," *Journal of*
296 *Bacteriology*, vol. 193, pp. 3894–3903, Aug. 2011.
297
298 [10] M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson, and
299 C. W. Murray, "Diverse , High-Quality Test Set for the Validation of Protein-Ligand Docking
300 Performance," *Journal of medicinal chemistry*, vol. 50, pp. 726–741, 2007.
301
302 [11] S. Cosconati, S. Forli, A. L. Perryman, R. Harris, S. David, A. J. Olson, C. Farmaceutica, and S. D.

Napoli, "Virtual Screening with AutoDock : Theory and Practice," *Expert Opin Drug Discov*, vol. 5, no. 6, pp. 597–607, 2011.

[12] Ricardo and Macarron, "Critical review of the role of HTS in drug discovery," *Drug Discovery Today*, vol. 11, no. 7-8, pp. 277–279, 2006.

[13] B. K. Shoichet, S. L. McGovern, B. Wei, and J. J. Irwin, "Lead discovery using molecular docking.," *Current opinion in chemical biology*, vol. 6, pp. 439–46, Aug. 2002.

[14] S. D. Morley and M. Afshar, "Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock.," *Journal of computer-aided molecular design*, vol. 18, pp. 189–208, Mar. 2004.

[15] P. P. Pattnaik, "Surface plasmon resonance: applications in understanding receptor-ligand interaction," *Applied Biochemistry and Biotechnology*, vol. 126, no. 2, pp. 79–92, 2005

[16] S. Perspicace, D. Banner, J. Benz, F. Müller, D. Schlatter, and W. Huber, "Fragment-Based Screening Using Surface Plasmon Resonance Technology," *Journal of Biomolecular Screening*, vol. 14, pp. 337–349, Apr. 2009.

[17] R. C. Stewart and R. Van Bruggen, "Association and Dissociation Kinetics for CheY Interacting with the P2 Domain of CheA," *Journal of Molecular Biology*, vol. 336, pp. 287–301, Feb. 2004.

[18] S. Wisen, E. B. Bertelsen, A. D. Thompson, S. Patury, and P. Ung, "Binding of a Small Molecule at a Protein-Protein Interface Regulates the Chaperone Activity of Hsp70-Hsp40," vol. 5, no. 6, pp. 611–622, 2010.

[19] P. Block, N. Weskamp, A. Wolf, and G. Klebe, "Strategies to Search and Design Stabilizers of

Protein-Protein Interactions: A Feasibility Study," vol. 186, no. September 2006, pp. 170–186, 2007.

.

342 **Tables**

343

344 **Table 1**. Summary of post-docking filtering stages.

Smarts pattern	Structural Descriptor	Compounds	Compounds after clustering	Selected compounds
[SX4]([OX1])([OX1])	Sulfonyl	1134	155	12
s1c2ncncc2cc1	Thienopyrimidine	46	9	3
o1nccc1	Isoxazole	773	89	7
c1cccnc1	Pyridine	603	213	2
n12N=CC=Cc1nnc2	Pyrimidine-Triazole	300	6	2
[#7, #8]	Hydrogen bond acceptor	3430	1501	9
[OX2H]	Hydroxyl	487	204	4
C(=O)N	Amides	697	345	13
-	Unclassified*	4500	-	-

345 *The compounds marked as unclassified were not matching any smarts pattern or were outside
346 pharmacophoric restraint region. They were manually inspected, but none of them seemed interesting
347 to us.

348

Figure Captions

Figure 1. The different curves represent the response units versus time for the interaction between fixed CheY and variable CheA in the following concentrations: 0, 0.1, 0.6, 1.25, 2.5, 5 and 50 μM.

Figure 2. Calculation of the kD for the complex. In black, the fitted response and, in gray, the value predicted as the kD.

Figure 3. Response of the two inhibitors compounds. I1 and I2 are represented in blue and red, when combinated with CheA, and in light blue and light red, when added without the protein. In black, CheA alone and, in gray, a reference with running buffer (a). Increase of the response produced by the combination of the compound G1 with CheA. The response of the protein with G1 is represented in green and, in light green, the compound without the protein. CheA alone is represented in black and the reference with running buffer is represented in gray (b).

Figure 4. Two-dimensional representation of the three compounds. In the left and in the center, the two inhibitors. On the right, the stabilizing compound. The structure depicted in color red is the common structure that we will remain fixed for following structure-activity relationship studies (top). View of the predicted binding mode of G1 compound in the protein-protein interface. CheA and CheY are represened in orange and yellow, respectively. The two restraints are also displayed: the blue stick is the NH group of TYR10 which forms the hydrogen bond, and the turquoise sphere represents the hydrophobic region (bottom).

Figures

Figure 1.

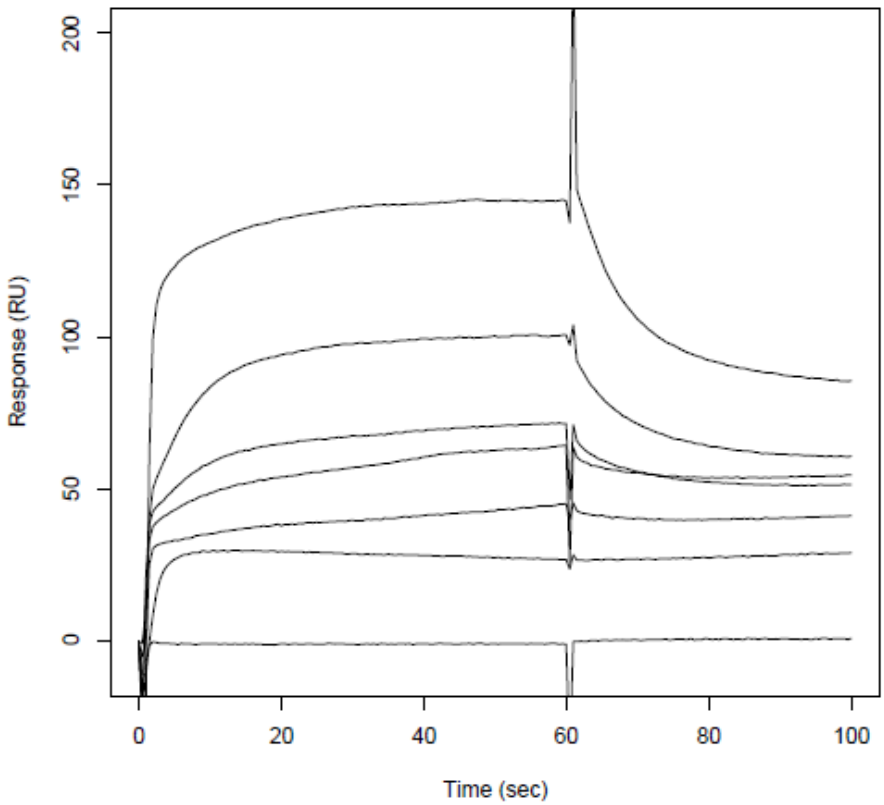


Figure 2.

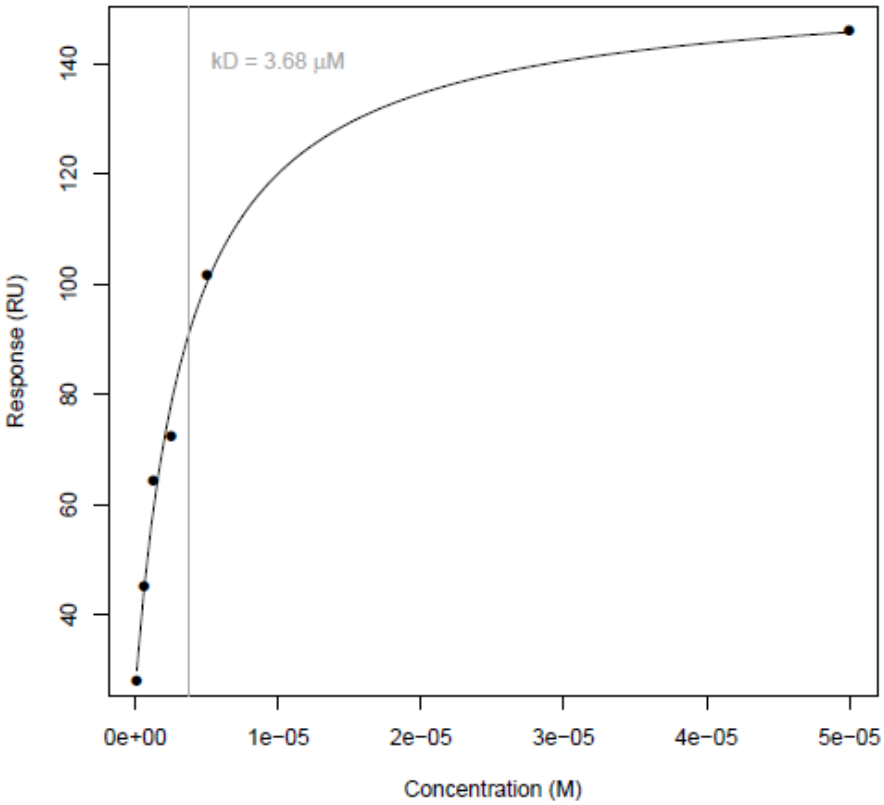
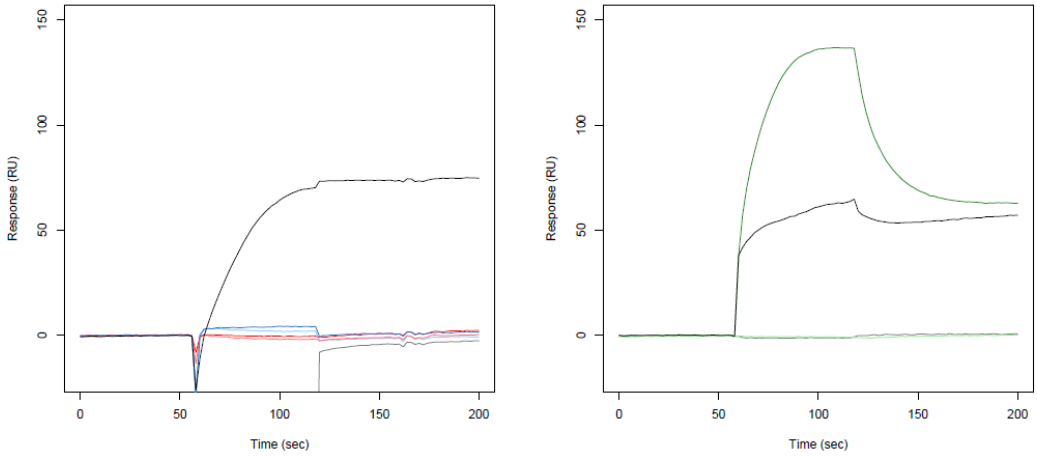

kD = 3.68 µM
Response (RU)
40
60
80
100
120
140
0e+00
1e−05
2e−05
3e−05
4e−05
5e−05
Concentration (M)

Figure 3.


Response (RU)
150
100
50
0
0
50
100
150
200
Time (sec)
Response (RU)
150
100
50
0
0
50
100
150
200
Time (sec)

(a) (b)

Figure 4.