Università degli Studi di Modena e Reggio Emilia

Dipartimento di Scienze della Vita

Scuola di Dottorato in Scienze e Tecnologie dei Prodotti per la Salute

XXVI ciclo

Rational design of multi-target ligands as potential drugs through computational methods

Dott. Andrew Anighoro

Relatore:

Prof. Giulio Rastelli

Direttore della Scuola e Coordinatore del Dottorato:

Prof.ssa Annalisa Tait

Table of Contents

1. Introduction

During my Doctorate degree course I was involved in an international collaboration network including the research groups of: my tutor, Prof. Giulio Rastelli (University of Modena and Reggio Emilia, Italy); Prof. Jürgen Bajorath (Rheinische Friedrich-Wilhelms-Universität Bonn, Germany); Dr. Leonard Neckers (National Cancer Institute, USA). My thesis project was mainly focused on the set up and application of a computational polypharmacology protocol targeting the interactome of heat shock protein 90 (Hsp90). The main objectives of which were *i)* the identification of promising target combinations for dual-inhibitor design and *ii)* the discovery of small molecules active against the selected targets. I pursued these goals by mining public molecular databases and performing a combined ligand- and structure-based virtual screening. I conducted the computational part of the work. In particular, I performed the ligand-based phase of the project as a visiting PhD student at the Bonn-Aachen Institute for Information Technology (B-IT), under the supervision of Prof. Jürgen Bajorath. I carried out the structure-based virtual screening at the University of Modena and Reggio Emilia. The selected candidate dual-inhibitors were experimentally tested by Dr. Leonard Neckers.

In addition to this, I performed further studies on computational drug design methods through the validation of docking and post-docking processing techniques on G-protein coupled receptors (GPCRs), with a focus on possible multi-target applications. I applied the same techniques in a prospective virtual screening campaign aiming at the discovery of allosteric inhibitors of cyclin-dependent kinase 2 (CDK2), a validated anticancer target.

Herein I describe the rational of these studies and the results obtained.

1.1 The process of drug discovery and computational methods

Pharmaceutical research aims at identifying new therapeutics to treat diseases affecting contemporary society. In general, the process of drug discovery for a given pathology consists of the following phases:

- Target selection;
- Lead discovery;
- Lead optimization;
- Preclinical tests;
- Clinical trials;

The target selection phase consists in the identification and validation of a biomolecular structure (e.g. an enzyme or a receptor) whose activity might be modulated in order to affect the pathology mechanism. The lead discovery phase consists in the identification of a small molecules displaying even a low level of activity (micromolar) on the selected target. Subsequently, the lead compounds are optimized through structural modifications in order to improve their activity (nanomolar). Preclinical trials are performed on animals to address safety concerns of the drug candidate. Finally, clinical trials (phase I, phase II, and phase III) are conducted on human to thoroughly assess safety and efficacy of the drug. A drug fulfilling efficacy and safety criteria may be approved for commercialization by a competent agency, such as the Food and Drug Agency (FDA, http://www.fda.gov/) in the USA or the European Medicines Agency (EMEA, http://www.ema.europa.eu/ema/) in Europe. The time laps between the concept and commercialization of a new drug is on average 14 years and the costs might exceed 1.0 billion USD. Moreover, the chance of successful market launch for each drug entering phase I clinical trial is as low as 5% [1]. The high costs and attrition rates associated to drug discovery makes it a highly risky business.

In order to speed up and cut the costs of drug discovery, several computational methods including bioinformatics analysis, virtual screening, QSAR models, ADME predictions, have been developed. In particular, chemoinformatics and molecular modeling methodologies can be applied in early stages of drug discovery for the identification of new hit compounds. Virtual screening (VS) is a validated method that permits a timely fashion ranking of databases containing millions of compounds according to their likelihood to be active on the studied target. Depending on whether the VS methodologies take into

account structural data of the target or not, they can be classified as structure- or ligand-based. Compared to a traditional high throughput screening, VS has the advantage that only a reduced number of compounds will be experimentally tested. This permits to save time and economical resources devoted to *in vitro* experiments. In this work both ligand- and structure-based methodologies have been applied in combination for the identification of novel active compounds against a single target or a combination of targets.

1.1.2 Structure-based virtual screening

Structure-based virtual screening (SBVS) aims at ranking a database of compounds according to a score representing the binding affinity of each compound with the target [2]. This procedure can be performed when a three-dimensional structure of the target is available from X-ray crystallography, NMR studies or homology modeling. The main tool in the context of SBVS is molecular docking [3]. Docking programs try to predict a putative binding mode for each compound of the database. The predicted binding modes are ranked by means of a scoring function that estimates binding affinity. There are many docking programs implementing different algorithms and scoring functions [4]. Molecular docking is a validated method, able to discover new active compounds. Nevertheless, it generally has limitations in the accuracy of binding affinity calculation and may lack a proper consideration of the flexibility of the target structure. Binding affinity scores are calculated implementing relevant approximations in order to decrease the computational time. These approximations may generate scores that do not correlate with experimental data. Moreover, the binding of a ligand may involve conformational changes of the receptor binding site (induced-fit). The use of a static X-ray structure of the protein may not provide a binding site conformation suitable to accommodate all the potential active compounds. A way to partly account for target flexibility is to use multiple target conformations (ensemble docking) [5]. Ideally, the different structures should include different classes of co-crystallized ligands in order to augment the chance of providing suitable binding site conformations. In general, it is widely accepted that docking output may benefit from post-docking processing. Molecular dynamics simulations provide a tool to assess more thoroughly the predicted binding modes. The application of more accurate scoring functions can improve the performance of docking programs. In this work, docking programs AutoDock4 [6] and Glide [7-8] were used to predict binding modes of large databases and target focused libraries of compounds. A post-docking processing protocol,

BEAR [9-10], was applied to refine docking complexes and estimate the energy of interaction via Molecular Mechanics-Poisson-Boltzman Surface Area (MM-PBSA) algorithm. BEAR was developed in the Molecular Modelling & Drug Design Lab at UNIMORE, where I performed my PhD studies. Molecular dynamics simulations in explicit solvent were also performed to validate binding modes and re-score docking complexes.

1.1.3 Ligand-based virtual screening

In contrast to SBVS, ligand-based virtual screening (LBVS) does not need structural information of the target. LBVS methodologies exploit the information encoded in the structure of known active compounds to discover novel chemotypes with a similar biological activity [11]. This approach is particularly useful when a crystal structure or a reasonable quality homology model of the target are not available. Famous ligand-based methodologies are similarity searching (SS), pharmacophore modeling, and QSAR. Recently, machine learning and data mining methodologies have also been successfully applied [12]. In this work, similarity searching (SS) and support vector machines (SVM) were applied to build libraries of compounds focused on specific target combinations. SS is based on the principle that similar compounds have similar biological activity. This principle is supported by numerous observations. Nevertheless, it is known that even small changes in chemical structures of a ligand may cause drastic changes in biological activity [13]. SVM requires a set of active and one of inactive compounds to construct a hyperplane in a high-dimensional reference space to classify data. In order to perform SS and SVM calculations a descriptor of molecular structure is needed. Different kinds of fingerprints have been designed at this purpose. In particular, in this work a dictionary-based fingerprint (*i.e.* MACCS structural keys) [14], a circular fingerprint (*i.e.* ECFP4) [15], and a 2D 3-point pharmacophore fingerprint (i.e. GpiDAPH3) [16] were used for SS and SVM. It is reported that the combination of different fingerprints may improve the performance of virtual screening.

Ligand-based calculations typically require less computational time with respect to docking. Here ligand-based calculations were applied to assess similarity between known active compounds of different targets, to build target focused libraries of compounds, and to expand novel active compounds identified.

1.2 Polypharmacology

Drug discovery has been strongly focused on the research of compounds active against a specific target, suitable for the development of highly potent and selective drugs. This paradigm is based on a gene centric view, linking the activity of a specific gene product to the observed pathological phenotype. According to this concept, a drug able to modulate in the proper way the activity of a deregulated protein should elicit a biological response reverting the pathological phenotype. In case of an augmented activity or over-expression of a particular enzyme, a selective inhibition could be enough to elicit a desirable therapeutic effect. This approach is also driven by the fact that non-selective (i.e. promiscuous) drugs with a wider and sometimes unpredicted spectrum of biological activities could eventually lead to adverse reactions, causing safety issues.

In the last decade, several studies have been published arguing that selective inhibition of a specific target can't explain the complex mechanism of action of some successful drugs. Moreover, a single target oriented drug design may not be suitable to develop new effective therapeutics for complex diseases. Indeed, the design of drugs able to interact with multiple targets (*i.e.* polypharmacology) is gaining major consideration in drug discovery [17-19]. Recent studies on polypharmacology suggest that the evolution of early biological organisms able to adapt to environmental changes might have favored the selection of biomolecular systems prone to chemical promiscuity [20]. A better understanding of the origin of polypharmacology is an important condition for improving our ability to rationally exploit it for therapeutic purposes. Multi-target drug activities may potentiate efficacy – either additively or synergistically - and be less sensitive to the insurgence of mutations that would result in drug resistance. Moreover, redundancy of biological networks should be considered when selecting target(s) to be modulated in a therapeutic regimen. Turning off a single switch may not be enough to shut down a cellular pathway. This concept is particularly relevant in the case of complex pathologies - such as cancer - that are often polygenic in nature and tend to involve the deregulation of complex and extended networks of proteins rather than a single one. In these conditions, the modulation of an optimal array of targets may provide a more efficient strategy to hit pathological mechanisms of complex diseases [21]. The modulation of multiple targets is often pursued in combination therapies by administrating two or more drugs. Polypharmacological drugs obtain the same objective by using a single drug molecule. It is generally recognized that several approved drugs elicit their therapeutic effect through

complex polypharmacology [22]. In most cases, it is important to note that such behavior was discovered only retrospectively [18] and that, in most cases, the complete set of current drug-target associations is still far from being completely characterized [23]. Therefore, next challenge will be how to a priori predict the desired polypharmacological profile.

The success of multi-target approaches will depend on several challenges, many of which still have to be faced. Rational design of multi-target active compounds is still in its infancy and surely needs substantial further work and methodological development. In this thesis an *ad hoc* polypharmacology protocol have been set up and applied to the case of study represented by the interactome of Hsp90. In particular, different computational methodologies have been combined aiming at the selection of the most promising target combinations and at the identification of candidate dual inhibitors. Moreover, a SBVS protocol was validated on G-protein coupled receptors, assessing its ability to recognize multi-target active compounds within a pool of molecular decoys.

1.3 Hsp90 interactome

Hsp90 is a molecular chaperone operating through ATP hydrolysis cycles. It assists the maturation, activation and stability of a wide range of client proteins belonging to the classes of kinases, transcription factors and others. Hsp90 interacts also with a range of co-chaperones involved in its activity. These proteins make up for an interactome of more than two hundred individuals [24]. Interestingly, many of them are validated drug targets for cancer therapy, involved in signal transduction and malignancy related pathways. Hsp90 itself is overexpressed in cancerous cells to support the activated forms of oncoproteins and buffer cellular stress promoting escape from apoptosis. Hsp90 is a target being pursued with increasing interest in cancer therapy, with 17 inhibitors currently undergoing clinical trials. Nevertheless, there are not yet approved drugs hitting this target. Epatotoxicity is a main concern related to Hsp90 inhibition. There are several types of Hsp90 inhibitors (Figure 1) such as the natural products geldanamycin, radicicol and their analogs. In addition, purines, resorcylic pyrazoles, isoxazoles, pyrimidines, aminopyridines, azoles and other chemotypes have been discovered over the years via high-throughput, fragment or VS [25]. The N-terminal domain of Hsp90 contains an ATP binding pocket which constitutes the primary site for inhibitor design. The Hsp90 interactome is an interesting area for the application of polypharmacology due to the pivotal role of Hsp90 and the presence of several proteins involved malignant phenotype pathways. Indeed, multi-target approaches are currently studied in clinical trials by means of combination therapies. Hsp90 inhibitors have been combined with taxanes, cisplatin, proteasome inhibitors, death receptor ligands, histone deacetylase inhibitors, and protein kinase inhibitors. The combination of 17-AAG, an Hsp90 inhibitor, with the erbB2 directed antibody trastuzumab [26] gave promising results in phase II clinical trials.

In this thesis, I present the set up of a computational protocol aiming to identify the most promising target combinations made up by Hsp90 and one protein of its interactome, and to select candidate dual-inhibitors. The selection of the target combinations was based on the analyses of the known actives deposited in public molecular databases. The information deposited in ChEMBL database (ChEMBLdb) [27] and in BindingDB (BDB) [28] was extensively analysed, extracting known inhibitors of Hsp90 and of its interacting proteins. The presence of known multi-target inhibitors and chemical similarity between known single target inhibitors have been chosen as a criteria to rank possible target combinations. Moreover, relevance for cancer therapy has been taken into account leading

to the selection of highly relevant proteins. LBVS including similarity searching and support vector machines have been performed to build focused compound libraries for the different target combinations selected. Subsequently, docking and post-docking processing methodologies have been applied to rank the libraries and select the most promising candidate compounds. This approach was repeated on different target combinations and with different conformations of the target proteins. The selected compounds have been purchased and sent for biological assays at NCI in order to test their activity towards the respective targets. Active compounds underwent hit expansion in the effort of optimizing multi-target activity profile.

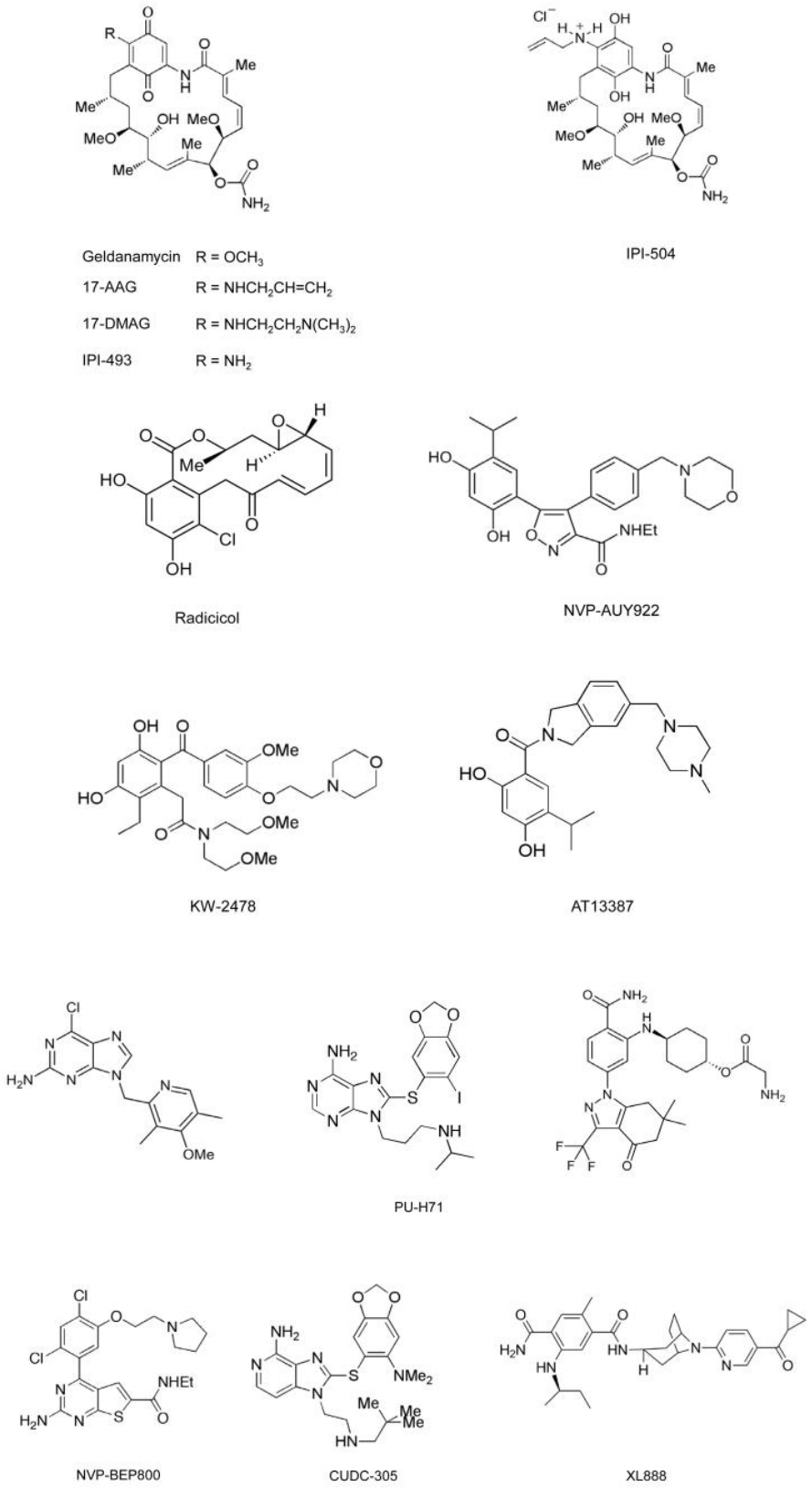

Figure 1. *Molecular structures of representative known Hsp90 inhibitors. Copyright Clinical Cancer Research [25]*

1.4 G-protein coupled receptors

G-protein coupled receptors (GPCRs) belong to a superfamily of signaling proteins that play a pivotal role in a number of physiological functions and in multiple diseases [29-30]. Most GPCRs can be classified in five families, namely Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2 and Secretin. They are characterized by a transmembrane region made up by seven alpha-helices. Between 30 and 40% of marketed drugs target GPCRs [31]. While more than 800 GPCRs are expressed in humans, current drugs that target these receptors are directed toward only a few GPCR members. In the near future, there is a strong belief that a growing number of hitherto unexplored receptors of this superfamily will be validated and exploited for pharmacological intervention. As a matter of fact, progresses in the structural biology of GPCRs are opening unprecedented opportunities for determining crystal structures of novel receptors, and the application of structure-based virtual screening (SBVS) tools to GPCR crystal structures will likely provide an increasing number of new chemical entities and hit candidates as more structures will be elucidated [32]. Therefore, there is considerable interest in validating current SBVS methods for optimal performance in GPCRs. SBVS is able to screen large libraries of small organic molecules against a biological target in a timely and cost-effective way. This approach relies on molecular docking, which is a validated and widely used in silico technique for the prediction of ligand binding modes and for the calculation of scores representative of the strength of the ligand-target association.

Here I report the results of a virtual screening and enrichment factor (EF) analysis that I performed on a set of GPCRs with known crystal structure. At the time this work was started, the crystal structures of five GPCRs were available in the Protein Data Bank (PDB), namely β2-adrenergic (β2), adenosine A_{2a} (A_{2a}), dopamine D_3 (D_3), histamine H_1 (H_1), and chemokine CXCR4 (CXCR4) receptors. More recently, several additional structures were solved, and others are reported as proximally to be published within the next few years [31]. In this work we analyzed β_2, A_{2a}, D_3, and H_1 receptors, four validated targets exploited in different therapeutic areas. β_2 is a target for asthma and pulmonary disease states; A_{2a} is targeted by drugs against asthma, Parkinson's disease, seizures, and different neurological disorders; D_3 is targeted by antipsychotics for schizophrenia; and H_1 is targeted by antihistamines.

Consistently with the study of McRobb et al. [33], for each receptor, a database of small molecules was built by seeding its respective known antagonists taken from the GLIDA

(GPCR-Ligand Database) database [34] into a larger set of drug-like decoys available from Schrödinger, made of 1000 compounds exhibiting drug-like properties [7-8]. For the analyzed receptors, the GLIDA database contained between 20 and 56 known antagonists. Even if a crystal structure was available, CXCR4 was excluded from our set because only four known antagonists of this receptor were available in GLIDA, and these were characterized by very complex structures such as high molecular weight peptides.

Virtual screening experiments were performed with AutoDock4 [6]. As discussed in section 1.1.2, scoring functions and binding mode predictions are necessarily approximated, and it is general opinion that docking results may need to be post-processed in order to achieve a better agreement with experimental activities. Recently, the research group where I worked during my doctorate studies has put forward a postdocking procedure, named BEAR (binding estimation after refinement), that automatically refines docking poses through molecular mechanics (MM) and molecular dynamics (MD) simulations and predicts binding affinities using more rigorous MM-PBSA and MM-GBSA free-energy based scoring functions [9]. Validation experiments performed so far showed that, in a number of cases, these two scoring functions were able to provide higher enrichment factors with respect to those obtained with other scoring functions implemented in docking methods, especially in the case of MM-PBSA, which usually yielded more accurate results compared to MM-GBSA [35-36]. However, the BEAR postdocking workflow was never tested on GPCRs before. A structural refinement of the docked complexes could be particularly useful in GPCRs, because flexibility and induced-fit are recognized as crucial for potency. Likewise, a more accurate description of ligand desolvation energies (like the one performed by MM-PBSA) could improve the predictions, considering that the vast majority of antagonists are biogenic amines protonated at physiological pH. For these reasons, the docking results obtained with AutoDock4 were postprocessed with these methods and the results were compared.

1.5 CDK2

Protein kinases have a crucial role in the modulation of cell pathways and are important targets for anticancer drug discovery [37]. As discussed above, polypharmacology is gaining popularity in drug discovery, this is generally true also in the case of kinase inhibitors. Nevertheless, in some cases selective inhibition of a specific kinase that has a critical role in the pathology might still be effective. Targeting the ATP binding site with the so called class I inhibitors has recently come up against selectivity problems. These issues can be considerably reduced by following an allosteric modulation approach. As a matter of fact, novel approaches that target allosteric sites distinct from the ATP site (class III inhibitors) are currently gaining a foothold in protein kinase drug discovery. Allosteric sites in protein kinase are less conserved throughout the human kinome and provide a possibility to design selective inhibitors. Moreover, these sites are accessible upon conformational changes that might result in higher residence time of allosteric inhibitors [38]. Type III inhibitors bind exclusively to allosteric pockets located in the proximity of the αC helix, which is a central mediator of allosteric activation/inactivation [39]. This helix is located in the N-lobe beside the active site and is usually swung outwards when the kinase is in an inactive conformation [40]. Being closely related to protein kinase inactivation, allosteric targeting via displacement of the conserved structural αC helix allows a direct and specific modulation mechanism. Allosteric modulation mechanisms involving this helix have been structurally elucidated in a variety of protein kinases, such as MEKs, PDK1, PKCζ, AKT1, and more recently CDK2 [39,41-42].

CDKs have been intensively investigated as anticancer targets, but none of the ATP-competitive CDKs inhibitors have yet been approved for clinical use, mainly because of selectivity issues [43]. In fact, the majority of the inhibitors tested so far were able to inhibit several CDKs with comparable IC50, but due to undesired toxicity, their use at effective doses was limited. Nevertheless, the relatively more selective (albeit not as selective as an allosteric inhibitor would be) CDK4/CDK6 inhibitor PD-0332991 was recently reported to triple the progression-free survival in combination with an aromatase inhibitor in patients with estrogen receptor–positive breast cancer, further supporting the notion that higher selectivity in CDK inhibitors may lead to clinically-relevant candidates [44].

CDK2 is an important member of the CDK family which plays an important role in

controlling the G1- to S-phase checkpoint and in DNA replication [45] and, therefore, represents an important pharmacological target for arresting or recovering control of the cell cycle in dividing cells [43,46]. In CDK2, the αC helix is involved in Cyclin recruiting to form an active CDK2/Cyclin complex. By inducing or stabilizing an outward orientation of the αC helix with an allosteric ligand, the CDK-Cyclin interface would be disrupted and kinase activation would be impaired [39]. Recently, crystal structures of CDK2 bound with the extrinsic fluorophore 8-anilino-1-naphtalene sulfonate (ANS) have been reported [47-48] In the crystals, two molecules of ANS bind a cavity formed by αC and the nearby strands α4 and α5 and induce a remarkable outward displacement of the αC helix, ultimately resulting in disruption of the recognition and binding site of Cyclin A. Ternary crystal structures of CDK2 in complex with ANS and type I inhibitors (JWS648, SU9516, and staurosporine) confirmed the truly allosteric nature of this fluorophore and resulted in CDK2 conformations nearly identical to that of the binary CDK2-ANS complex [47-48]. This is the first time that a completely allosteric ligand proved to be able to displace the αC helix and to inactivate a CDK kinase via an allosteric mechanism, prospecting a new strategy to design inhibitors with potentially improved selectivity. However, ANS is readily displaced from CDK2 upon Cyclin binding, because its affinity for CDK2 (Kd=37 μM) is significantly lower than that of Cyclin A [47]. Therefore, the design of compounds able to bind this newly discovered allosteric pocket with higher affinity and with better drug-like properties is highly desirable.

In contrast to the multi-target approach described in section 1.3, in this case allosteric inhibitors of a single target, namely CDK2, represent the objective of the virtual screening campaign. A database of commercially-available compounds (Asinex collection, ~600.000 compounds) was docked in the allosteric pocket of the CDK2-ANS binary complex. The screening was conducted by using a combination of docking with AutoDock4 [6] and our in-house post-docking tool BEAR [29]. The binding mode of the best ranking complexes was visually inspected. The most promising compounds where chosen for experimental assessment of their activity on CDK2 on both purified proteins and breast cancer cell lines. Moreover, hit expansion of active compounds was performed via similarity searching. The candidates were tested thanks to a collaboration established with the group of Dr Massimo Broggini (Istituto di Ricerche Farmacologiche Mario Negri, Milano).

2. Methods

2.1. Hsp90 interactome

2.1.1 Known active compounds collection and activity annotations analysis

Known compounds active on proteins belonging to the Hsp90 interactome were extracted from ChEMBLdb [27] (release 13, ~1,200,000 compounds) and BindingDB [28] (Accessed on April 2012). Activity annotations were analyzed by using KNIME (http://www.knime.org/).

2..1.2 Ligand-based virtual screening, ChEMBL database

MACCS fingerprints were calculated for the ChEMBLdb and for 418 Hsp90 actives. All fingerprints types used in this project were calculated with Molecular Operating Environment (MOE) [16]. Twelve rankings of ChEMBLdb were obtained via 1-nearest neighbor (1-NN) SS, taking as a reference the Hsp90 actives and each of the 11 subsets of multi-target compounds. Two additional rankings were obtained by performing SVM calculations [49] on the same database, using the Hsp90 actives and a set made up by the 25 multi-target compounds. SVM calculations need wider reference set compared to SS, hence the 25 multi-target compounds were pooled together for SVM instead of considering 11 subsets as it was done for SS. The activity annotations of the top 1000 molecules of each of the 14 rankings were analyzed, selecting compounds annotated as active against at least a protein of the Hsp90 interactome.

2.1.3 Ligand-based virtual screening, ZINC database

Thirteen target combinations were considered for building focused compound libraries by a LBVS of the ZINC [50] (release 12, clean and drug like subset, ~11,000,000 compounds). Reference sets were chosen from the same ChEMBL compounds selected in the previous phase. When up to 50 ChEMBL compounds were available for a target combination, all compounds were included in the reference set. When more than 50 ChEMBL compounds were available, a representative compound was chosen for each molecular scaffold to

avoid excessive redundancy. Core structures were obtained by removing all substituents, cyclic carbon structure were obtained by converting all atom types to carbon, and reduced cyclic carbon structures were obtained by reducing all bond order to one and considering all ring systems as equal between them.
MACCS and ECFP4 fingerprints were calculated for the database and the reference sets. SS, 1-NN and 10-NN, and SVM calculations were performed to rank the ZINC database, obtaining 6 rankings (three methodologies per two different fingerprints). Thirteen focused libraries, one per target combination, were obtained by merging the overlapping compounds in the top 2000 positions with the top 200 compounds in each of the six rankings.

2.1.4 Ligand-based virtual screening, selected vendor catalogs

A database of ~4,200,000 commercially-available compounds (ASINEX, Maybridge, LifeChemicals, ChemBridge, InterBioScreen, AMRI, aronis, Enamine, VitasM) was used in this phase.

Compound 4 and compound 6 were chosen for hit expansion. Three types of fingerprints were calculated (MACCS, ECFP4, GpiDAPH3) for the database and reference compounds (compound 4 and 6). Hence, three rankings of the database were obtained for each hit compound via SS. The two hit expansion libraries were obtained by merging the top 500 compounds from each ranking with the compounds shared by any two or three rankings (overlap) in the top 5000 positions.

Known dual-inhibitor expansion was conducted to build a new focused library. Three dual-inhibitors active on Hsp90 and erbB2 were extracted from the ChEMBLdb and used as reference set. 3-NN SS calculations were performed to obtain three rankings of the database (one per fingerprint used). A final library was obtained by merging the top 500 compounds from each ranking with the compounds shared by any two or three rankings (overlap) in the top 5000 positions.

New focused libraries were built for four target combinations. A set of known inhibitors was extracted from ChEMBLdb for Hsp90 (23) and each of the second targets (26 for serine/threonine-protein kinase B-raf (Braf), 33 for receptor tyrosine-protein kinase erbB-2

(erbB2), 37 for proto-oncogene tyrosine-protein kinase Src (Src), 54 for glucocorticoid receptor (GR) by selecting a representative for each cyclic carbon scaffold. Three different fingerprints (MACCS, ECFP4, GpiDAPH3) were calculated. SS and SVM calculations were performed. Hence, six rankings were obtained for each target combination (two methodologies per three fingerprints). New focused compound libraries were built by fusing the rankings obtained with the same methodology (SS or SVM) and the same fingerprint. Rank fusion was obtained by averaging the position occupied by each compound in two rankings.

2.1.5 Structure-based virtual screening, ChEMBL database

Five Hsp90 conformations were considered. X-ray crystal structures of Hsp90 in complex with Geldanamycin, PU3 and an isoxazole based inhibitor were extracted from the PDB (PDB IDs: 1OSF [51], 1UY6 [52], 2VCI [53]). Three water molecules bridging between the co-crystallized ligands and the protein were conserved in the protein structures (WAT 402, 407, 418 in 1OSF; HOH 2059, 2134, 2137 in 1UY6; HOH 2099, 2232, 2233 in 2VCI). An additional conformation was obtained by including 1UY6 structure with a fourth water molecule (WAT 2262). Moreover, a homology model of the protein including a portion of the C-term was built using 1UY6 as a template. Tautomers and protonation states of the compounds to be docked were calculated with Schrödinger Ligprep. Docking calculations were performed with AutoDock4 (number of genetic algorithm energy evaluations = 250,000, genetic algorithm population size = 150, genetic algorithm runs = 20, rmsd tolerance = 2), grids were calculated with AutoGrid4, centered on the mass centre of co-crystallized ligands and with dimensions suitable to contain the binding site (spacing 0.375, number of grid points = 45x50x45). AutoDock results were post-processed with BEAR. The protocol consists of three steps based on molecular mechanics (MM) and dynamics (MD) cycles. In particular, 2000 steps of MM energy minimization of the whole protein−ligand complex were performed, followed by 100 ps MD where the ligand was allowed to move, and a final reminimization of the entire complex. At the end, the binding free energy of each refined complex was computed by using the MM-PBSA method. Full details about energy minimization, MD, and MM-PBSA binding free energy predictions can be found in ref 9.

Best scoring compound were processed through MD simulations in explicit water solvent performed with AMBER module *sander* [54]. The complexes were solvated in an octagonal

box of TIP3P water molecules. The structures were then prepared for MD by heating the system at 300K and running equilibration. This was accomplished in the following steps:

- 5000 steps minimization of the whole complex;
- 10 ps of MD at constant volume with 1 kcal/mol restraint on the receptor and the ligand;
- 10 ps of MD at constant temperature and pressure with 1 kcal restraint on the receptor and the ligand;
- 10 ps at constant temperature pressure with 0.5 kcal/mol restraint 1ns of MD with constant pressure without restraints;
- 1 ns of MD at constant pressure without restraints.

Finally, 20 ps of production MD was ran at constant temperature and pressure and without restraints. Binding free energies were calculated over the 20 ns of production MD by averaging energy scores of 4000 snapshots taken every 5 ps.

2.1.6 Structure-based virtual screening, ZINC database

X-ray crystal structures were selected for each second target (3PP0 [55] for erbB2, 3SKC [56] and 3D4Q [57] for Braf, 2H8H [58] for Src, 1M2Z [59] for GR). Two different structures were considered for Braf as they displayed some significant differences in the binding mode of the two co-crystallized ligands (3SKC binding the amide linked pyrazolopyridine inhibitor BR2, 3D4Q binding the pyrazole-based inhibitor SM5). A conserved water molecule was considered for erbB2 and Src. This molecule establishes a hydrogen bond network between the ligand and residues in the ATP binding site (WAT 22 in 3PP0, HOH 543 in 2H8H). Four water molecules involved in a hydrogen bond network in the protein structure were considered in GR (HOH 1, 37, 112, 195). Glide was used as an additional docking program to increase the chances of obtaining correct binding modes. Glide was applied only on four Hsp90 conformations (1UY6 with only three water molecules was considered as no substantial differences were observed by applying Glide on 1UY6 with 4 water molecules). The default docking parameters were used (Glide Standard Performance protocol (SP), Scaling Factor applied to the ligand = 0.8, one docking conformation saved for each compound). The default parameters were also used for grid generation, centred on the co-crystallized ligand. AutoDock4 and BEAR parameters were the same as described above. A selection of complexes with MM-PBSA score ≤ -20 (BEAR

scores) was processed with more extended molecular dynamics (MD) simulations as described in section 2.1.5.

2.1.7 Structure-based virtual screening, selected vendor catalogs

A protocol comparable to the one described for the ZINC compounds was applied also for the new libraries derived from the database of selected vendors. A protein structure without any water molecule was consideredfor erbB2 and Src, hence two conformations were used in this round for all kinases. AutoDock4, Glide, BEAR, and MD simulations were performed with the same parameters described above.

2.1.7 Biological assays

All the compounds selected via virtual screening were assayed for inhibitory activity on Hsp90. SkBr3 cells were lysed in TNES lysis buffer containing protease inhibitors. Equal amounts of total protein were then treated with 1 μM, 3 μM, 10 μM, 30 μM, 100 μM concentration of candidate inhibitor (or untreated) for 1hr under rotation at 4C prior to geldanamycin (GA) bead precipitation for an additional hour. Samples were assayed by Western blot to visualize whether the compounds were able to compete with GA for Hsp90 binding.

ZINC compounds were tested also on the second targets, namely erbB2, Braf, SRC, and GR. Intact SkBr3 cells were incubated overnight with 1 μM, 3 μM, 10 μM, 30 μM, 100 μM concentrations of each compound, then lysed in TNES buffer and analyzed via Western blot for inhibition of target proteins.
All assays have been performed in Dr Leonard Necker’s lab at NCI.

2.2 Virtual screening on G-protein coupled receptors

The crystal structures used for docking were obtained from the Protein Data Bank (PDB), with PDB codes 2RH1 [60] for β_2, 3EML [61] and 4EIY [62] for A_{2a}, 3PBL [63] for D_3, and 3RZE [64] for H_1.

Except otherwise noted, all structures were prepared by removing all solvent molecules. In the case of H_1, a phosphate anion was retained in the receptor structure. This phosphate occupies an anion binding site at the entrance of the binding pocket that affects both the stability of the receptor and the binding of ligands such as doxepin, the co-crystallized antagonist. In the case of A_{2a}, it is known that the presence of water molecules in the binding site is important for predicting correct ligand binding geometries and that SBVS may gain in accuracy when these are retained in the receptor structure. Previous virtual screenings retained from 3 to 8 co-crystallized water molecules [33, 65]. Therefore, VS experiments on A_{2a} were performed using a variable number of water molecules, and the results are reported only for the best performing conditions (in terms of EF and % of correct binding modes).

In A_{2a} (3EML) and H_1 (3RZE) crystal structures, a few residues of the second extracellular loop (ECL2) were missing because of disorder. Since ECL2 is close to the binding site, these residues have been modelled as described below. In A_{2a}, the ECL2 loop was taken from the crystal structure of A_{2a} in complex with the agonist adenosine (PDB code 2YDO [66], 3Å resolution). The conformation of this loop turned out to be almost identical to the one observed in the higher resolution crystal structure of A_{2a} (PDB code 4EIY) in complex with the same antagonist 4-(2-[7-amino-2-(2-furyl)-[1,2,4]triazolo-[2,3-a] [1,3,5]triazin-5-ylamino]ethyl)-phenol (ZM241385) that was reported only very recently (resolution of 1.8Å and 2.6Å for 4EIY and 3EML, respectively), and in which ECL2 was well defined and complete. In H_1, since 3RZE was the only structure available in the PDB for this receptor, the seven missing residues in ECL2 were modelled using the Modweb and ModLoop web servers [67-69]. A starting conformation of the seven missing residues was generated using Modweb. Then, this portion of the loop was refined using ModLoop. This tool computes 300 loop conformations, and returns the one estimated to be energetically most favorable. Moreover, 4EIY features a highly conserved binding site for an ion (Na^+) that was not detected previously in 3EML. Although the sodium ion does not interact directly with the co-crystallized antagonist, the authors confirmed the allosteric

effects of Na^+ on ligand binding and thermal stability of A_{2a}. Therefore, the 4EIY structure was added to the present analyses in order to compare the results and check whether a higher resolution and the presence of the sodium ion may improve docking performance.

Hydrogen atoms were added to all the receptors using the Leap module of Amber 10 [54], then the structures were further prepared through an energy minimization conducted with the *sander* module of Amber. During structure preparation and energy minimization, the co-crystallized antagonists were retained in their binding pockets, then they were removed before docking.

Known antagonists were extracted from GLIDA database [34], while molecular decoys were taken from the Scrhödinger decoy set. Protonation states of the ligands were calculated with the Epik module of the Schrödinger suite [70-71]. Gasteiger charges were computed for known antagonists and decoys. Docking was performed with AutoDock4. Grids dimensions for each receptor were set considering the volume of the cavity and the molecular size of the known antagonists. The grids were centered on the center of mass of the cocrystallized antagonists. Docking and BEAR parameters are the same described in section 2.3.1. EF was calculated as the fraction of known active ligands identified in certain thresholds of the ranked database.

2.3 CDK2

2.3.1 Virtual screening on CDK2

The Asinex collection of commercially-available compounds was downloaded in Mol2 format from the ZINC database (“Usual” subset) (July 2011). This dataset contained a total of 600.722 compounds including protonation variants and tautomers at pH values from 5.75 to 8.25, corresponding to 545.609 unique ZINC IDs.

The crystallographic coordinates of CDK2 in complex with two molecules of 8-anilino-1-naphtalene sulfonate (ANS) used for docking were taken from the Protein Data Bank (PDB), PDB code 3PXF [72]. The two ANS and all water molecules were removed from the structure. Hydrogen atoms were added to the protein structure using the *Leap* module of Amber10. All Asp and Glu were negatively charged and all Lys and Arg were positively charged. This structure was used as starting point in both docking and post-docking calculations.

Docking was performed with AutoDock4 with the same parameters described in section 2.3.1. The original CM2 charges present in ZINC were retained for the post-docking analyses made with BEAR while the parameters were set as in section 2.3.1. Grid dimensions were set so as to include the two ANS molecules and a number of buried water molecules around them (W318, 327, 338, 374, 406, 473, 478, 496 in 3PXF) that could be displaced by the docked compounds. The number of resulting grid points were 62x46x44 (0.375 Å spacing). Re-docking of ANS with these parameters retrieved a binding mode that was almost identical with that of the co-crystallized ligand (RMSD of 0.529 Å). Of the two ANS molecules bound to the crystal, AutoDock placed the ligand in the inner (residue 305.A in 3PXF) and conceivably more favourable binding site, with a score of -7.4 kcal/mol.

The Asinex database was ranked according to the lowest-energy scoring docking solution.

The best scoring docking complexes obtained with AutoDock were post-processed with BEAR. Post-docking was performed on the top fraction of the ranked database obtained by applying an AutoDock energy score cut-off of -11 kcal/mol (~27.000 compounds, corresponding to ~5% of the initially docked database). The best AutoDock score was -15.6 kcal/mol. Each complex was submitted to the BEAR refinement and rescoring procedure. Post-docking results were ranked according to the MM-PBSA binding free energy scores, and top ranked compounds were visually inspected with Chimera 1.6.2

[73]. Before selection, the top 100 compounds were clustered into chemical classes using the Canvas similarity and clustering utility available in Maestro 9.2 [74], using default settings. Thirty-five compounds were finally selected, purchased, and submitted to biological evaluations.

The binding modes of the active compounds 1-7 discovered by virtual screening were confirmed by performing a more exhaustive sampling and clustering of docked orientations compared to the original conditions applied for virtual screening (100 runs of Lamarckian genetic algorithm instead of 10). Since the ZINC database did not contain all possible stereoisomers of compounds 1-7, complete sets of stereoisomers were re-generated using the LigPrep utility available in Maestro (Schrödinger 2011 suite). For each enantiomer, 100 docking runs were performed with AutoDock and the resulting orientations were clustered using a default RMS tolerance of 2 Å. For each enantiomer, clusters of orientations were ranked according to *i)* the lowest energy score predicted by AutoDock, and *ii)* the population of each cluster, *i.e.* the number of orientations that populated each cluster. Then, representative orientations of the five best-scoring clusters and of the five most-populated clusters were post-processed with BEAR. CM2 atomic charges present in the ZINC database were used for post-docking refinement. Finally, the overall best-scoring solutions according to MM-PBSA (BEAR) were chosen as representatives of the binding mode of each compound.

2.3.2 Hit expansion

With the aims of exploring the SAR of the identified hits and potentially discovering other biologically active hits, commercially available analogues of compound 4, *i.e.* the compound showing the most interesting activity in cancer cell-based assays, were searched in the Asinex database and then processed with the same docking and post-docking procedure described above. To prepare the focused library, compound 4 was taken as a reference for similarity searching calculations. Molecular ACCess System (MACCS) fingerprints [75] as implemented in OpenBabel 2.3.1 [76] were computed for the reference compound 4 and for the complete Asinex database. Then, similarity searching calculations made with OpenBabel were ranked according to Tanimoto coefficient (Tc) values computed for each database molecule *vs* the reference compound 4, and compounds with a Tc value of at least 0.7 were selected. This selection led to a focused library of 1108 unique compounds, that corresponded to 2217 database entries by

including multiple tautomers and protonation states. Of these, 681 were already processed in the primary virtual screening, while the remaining 1536 had been excluded owing to the AutoDock energy score cut-off of -11 kcal/mol imposed when selecting molecules for post-docking refinement and rescoring. The 2217 compounds were then submitted to docking and post-docking calculations using the same procedure described for the primary screening.

2.3.4 ADME predictions

ADME properties of the tested compounds were evaluated by using QikProp v.3.4 (Schrödinger suite) [77].

2.3.3 Biological assays

Recombinant CDK2 was purified from *E. coli* using a GST-CDK2 fusion construct. The purity of the protein was checked by SDS-PAGE electrophoresis and the amount determined using the Protean assay (BIORAD).
ANS displacement assay was performed as described by Martin *et al*. [78].
To discriminate between type I/II and type III ligands, competition assays in the presence of Staurosporine (2.5 μM), an ATP-competitive inhibitor, were performed as described in ref 12.
The inhibition of CDK2/Cyclin A kinase activity was determined using a commercially available luminescent-based method (ADP-Glo™ Kinase Assay, Promega) determining the ability of recombinant CDK2/Cycin A complex to phosphorylate one of its major substrate, Rb. The assay was performed following the manufacturer's instructions. Staurosporine was used as positive control.

Cell growth inhibitory activity was determined using the MTS test. MDA-MB231 and ZR75-1 breast cancer cells were used for these experiments.

Phosphorylation of CDK2 substrate Rb in cells was determined by Western blotting.

The tested compounds were purchased from Asinex. The vendor had verified compound purity by liquid chromatography-mass spectrometry (LC-MS) or nuclear magnetic

resonance (NMR) experiments. All compounds had at least 95% purity except **1** (92%) and **3** (93%). All chiral compounds were purchased and tested as racemic mixtures.

All biological tests have been performed in the group of Dr Massimo Broggini at Istituto di Ricerche Farmacologiche Mario Negri, Milano.

3. Results

3.1 Hsp90 interactome

3.1.1 Known active compounds collection

The information on Hsp90 inhibitors was analyzed by extracting compounds annotated as active against the chaperone from ChEMBLdb and BindingDB. This analyses resulted in a subsest of 1358 ChEMBL compounds and 500 BindingDB compounds. Due to a partial overlap between the two databases, a total of 1583 unique compounds were found. Notably, public molecular databases include activity annotations relative to assays performed and reported with different levels of reliability. Therefore, a *high confidence* subset of Hsp90 actives was built by including the annotations matching the following criteria:

- activity expressed as Ki or IC50;
- activity value not higher than 50,000 nM;
- when multiple annotations were available for the same compound, only compounds for which they did not differ for more than one order of magnitude were retained;
- ChEMBL assay confidence score equal to 9 (the maximum).

A *low confidence* dataset of Hsp90 actives was built by including all activity types and values for this target. As a result, the *high confidence* subset contained 418 compounds, while the wider *low confidence* subset contained all 1583 compounds. Visually inspecting the *high confidence* compounds, the major chemotypes of Hsp90 inhibitors were found. Hence, this subset was considered representative of the known Hsp90 inhibitors chemotypes and suitable for further analyses without the need to include *low confidence* data.

Table 1. *Hsp90 active compounds (low and high confidence activity annotations) found in ChEMBLdb and BindingDB.*

Activity annotation confidence	ChEMBLdb (Hsp90 actives)	BindingDB (Hsp90 actives)	Overlap (Hsp90 actives)	Total (Hsp90 actives)
Low	1358	500	275	1583
High	78	395	55	418

3.1.2 Second target annotation analyses

In databases such as ChEMBLdb and BindingDB, many compounds are annotated as active against multiple targets. Provided the set of 418 *high confidence* Hsp90 actives, their activity annotations towards other proteins (second targets) were extensively analyzed. In order to conduct a comprehensive search of possible second targets, no activity annotation confidence level filters were applied in this phase. Among the 418 compounds, 102 were found to have activity annotations towards 163 second targets. These compounds are characterized by a multi-target activity profile, being able to bind Hsp90 and at least another target. Interestingly, 26 of these compounds were annotated as active against 17 proteins that are listed as Hsp90 interactors (Table 2).

Table 2. *Number of multi-target compounds within Hsp90 actives and number of second targets hit .*

Number of multi-target compounds	Number of second targets	Second target class
102	163	Any
26	17	Hsp90 interactor

The data reported in Table 2 highlight the fact that the analyzed databases include compounds with the desired multi-target activity profile. Moreover, at least 17 Hsp90 interactors have a binding site that is able to accommodate ligands in common with Hsp90 (Table 3). From a structural point of view, each of these proteins represent a potential candidate to be paired with Hsp90 for the design of multi-target inhibitors. Considering their relevance as targets for anticancer drug discovery, 11 second targets were initially prioritized for further analyses (bold character in Table 3). When considering these

proteins, the number of multi-target active compounds shared with Hsp90 was 25. Some of the second targets shared only one active compound with Hsp90 while estrogen receptor β (ER-β) shared 12 compounds.

Table 3. *The seventeen Hsp90 interactors sharing active compounds with Hsp90. The relative number of active compounds is also reported. In bold characters are reported the prioritized Hsp90 interactors.*

Target name	Uniprot ID	Active compounds shared with Hsp90
Estrogen receptor β	Q92731	12
Estrogen receptor α	P03372	6
Polo-like kinase	P53350	4
Mitogen-activated protein kinase kinase kinase 5 (ASK1)	Q99683	3
Receptor protein-tyrosine kinase erbB2	P04626	3
Mitogen-activated protein kinase 6	Q16659	2
Vitamin D receptor	P11473	2
Serine/threonine-protein kinase PIM1	P11309	2
Microtubule-associated protein tau	P10636	2
Apoptosis regulator Bcl-2.	P10415	2
Tyrosine-protein kinase FES	P07332	2
Protein kinase C ε	Q02156	1
Heat shock factor protein 1	Q00613	1
DNA-dependent protein kinase	P78527	1
FK506 binding protein 12 (mTor)	P42345	1
Serine/threonine-protein kinase Braf	P15056	1
Tyrosine-protein kinase SRC	P12931	1

3.1.3 ligand-based virtual screening, ChEMBL database

ChEMBLdb (release 13) contains 1,296,266 compound records. Many of the deposited compounds have activity annotations against multiple targets. It is unlikely that all the

possible compounds have been (or perhaps will ever be) tested against all possible targets. Therefore, some compounds in ChEMBLdb might be active also on targets against which no assays are reported. These data sparseness was addressed by a ligand-based approach aiming at uncovering multi-target compounds missed in the previous step. Similarity searching (SS) and support vector machines (SVM) calculations were performed to rank the ChEMBL taking as reference the complete set of 418 Hsp90 actives and the subset of 25 multi-target actives (details in methods section). Among the top ranked compounds we searched for activity annotations against at least a protein of the Hsp90 interactome. According to their high ranking, these compounds should also have a good chance to include currently unknown Hsp90 actives. The number of compounds identified through this procedure is summarized in Tables 4 and 5.

Table 4. *Number of compounds with annotations against the Hsp90 interactome identified via SS and SVM.*

Reference set: 418 Hsp90 actives				
Activity annotation (Targets)	**SS (number of compounds)**	**SVM (number of compounds)**	**Overlap (number of compounds)**	**Total (number of compounds)**
Hsp90 OR Hsp90 interactor(s)	356	286	62	580
Hsp90	131	34	28	137
Hsp90 interactor	256	263	41	478
Hsp90 AND Hsp90 interactor(s)	31	11	7	35
Reference set: 25 multi-target actives				
Activity annotation (Targets)	**SS (number of compounds)**	**SVM (number of compounds)**	**Overlap (number of compounds)**	**Total (number of compounds)**
Hsp90 interactome	306	419	53	672
Hsp90	49	17	10	56
Hsp90 interactor	277	416	52	641
Hsp90 and Hsp90 interactors	20	14	9	25

Table 5. *Number of Hsp90 interactome second targets emerged via SS and SVM.*

Reference set: 418 Hsp90 actives			
SS (number of targets)	**SVM (number of targets)**	**Overlap (number of targets)**	**Total (number of targets)**
58	70	54	74
Reference set: 25 multi-target actives			
SS (number of targets)	**SVM (number of targets)**	**Overlap (number of targets)**	**Total (number of targets)**
58	69	52	75

The data in Table 4 and 5 suggest that the promiscuous nature of the subset of 25 multi-target compounds reflects in the identification of a higher rate of compounds with second target annotations. In the calculations performed taking the 418 Hsp90 actives as reference, a total of 580 unique compounds active against 74 targets within the Hsp90 interactome were identified.. On the other hand, taking the 25 multi-target compounds set as reference, 672 compounds active against 75 targets were identified. Moreover, the calculations performed taking the 25 multi-target active compounds as reference lead to a higher number of compounds active against Hsp90 interactors with respect to the 418 Hsp90 actives (641 *vs* 478). In particular, this behavior is observable in the SVM results where 416 compounds active against Hsp90 interactors were identified taking as reference the multi-target compounds against the 263 from the Hsp90 actives. Another interesting result is that a significant degree of overlap is found between the results obtained via SS and SVM. This aspect was further analyzed by building the histograms shown in Figures 2 and 3. The overlap noticed in Tables 4 and 5 is confirmed by a general trend of consensus between the relative number of active compounds retrieved for each target by SS and SVM. Higher number of compounds were identified for estrogen receptors α and β, microtubule-associated protein τ, and DNA polymerase η. To which extent the consensus between SS and SVM results is influenced by the methodologies themselves or by a possible bias in the database (simply more annotations might be available for some targets) could be assessed by further methodological studies that go beyond the objectives of the present work.

The previous list of 17 second targets sharing known active compounds with Hsp90 was expanded to 81 second targets by merging the results obtained with the two reference sets. Each target combination (Hsp90-interactor) was at this point associated to a set of

ChEMBL compounds active against at least the second target and with a reasonable probability of being active also against Hsp90. These compounds represent a valuable dataset for the search of currently unknown dual-inhibitors. Their activity should be experimentally confirmed only against Hsp90.

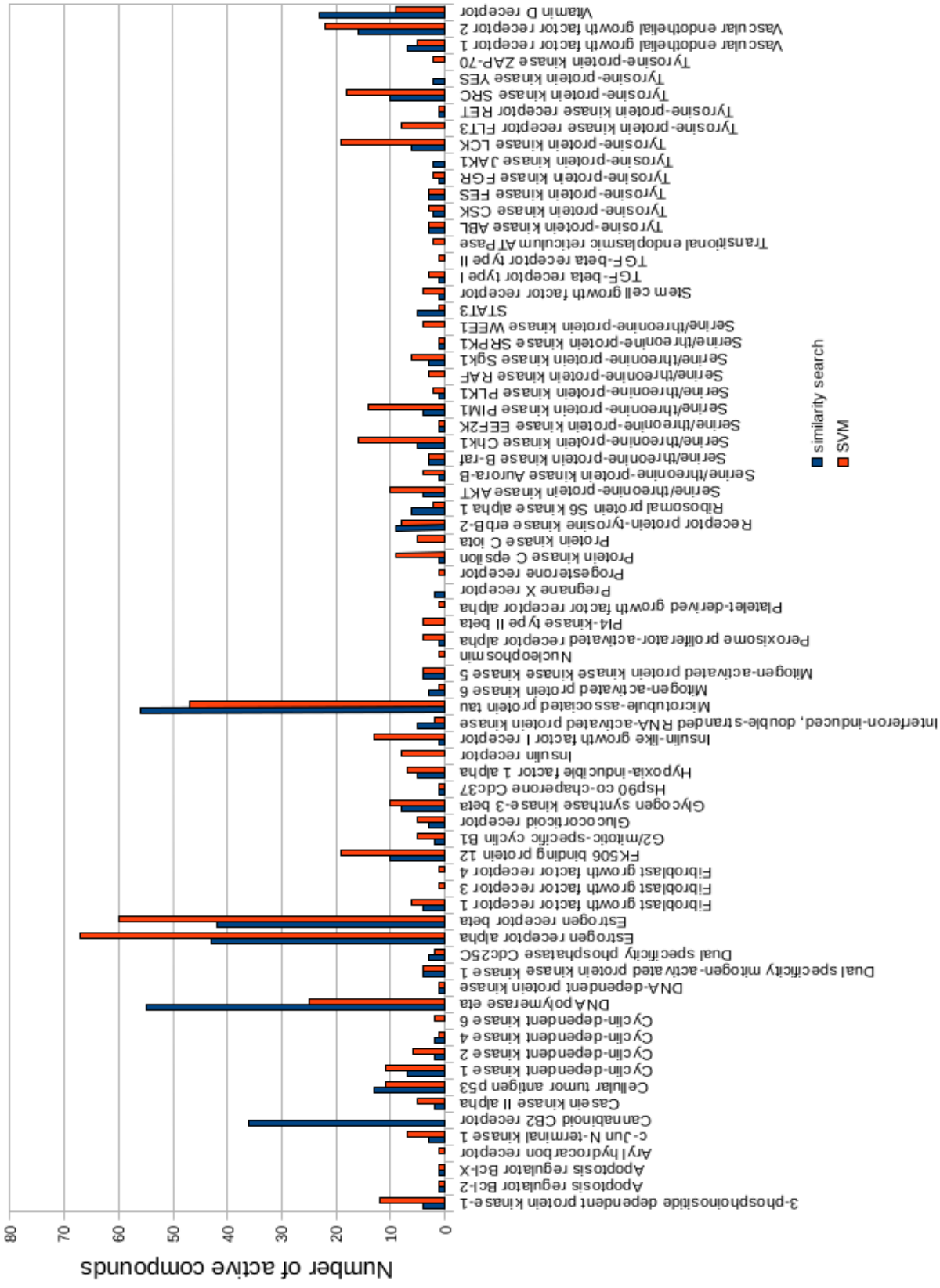

Figure 2. *Number of active compounds identified for each second target found by SS and SVM, using the 418 Hsp90 actives as reference compounds.*

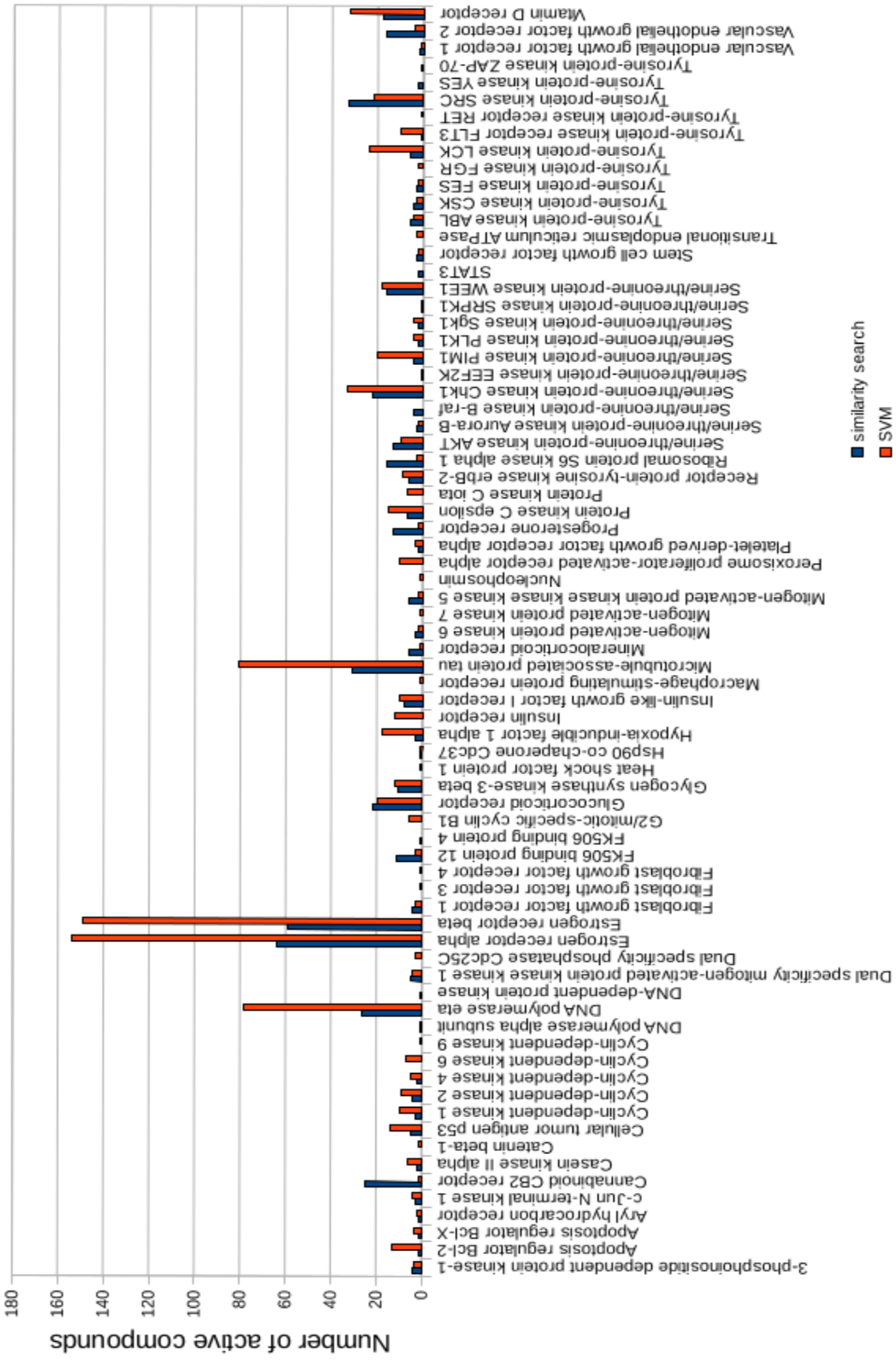

Figure 3. *Number of active compounds identified for each second target found by SS and SVM, using the 25 multi-target actives as reference compounds.*

3.1.4 Ligand-based virtual screening, ZINC database

The list of 81 second targets was ranked according to the number of ChEMBL compounds retrieved by the screening described above. Each of these targets together with Hsp90 represents a potential target combination for dual-inhibitor design. Among the top ranked combinations, 13 second targets were selected (Table 6) on the basis of the *i)* availability of an X-ray crystal structure in the PDB for the structure-based phase and *ii)* the relevance of the target in cancer drug discovery. For each of the 13 selected targets a ligand-based virtual screening was conducted against the ZINC database. This new round of calculations was performed to build commercial compound libraries focused on the target combinations selected. Reference compounds were selected within the ChEMBL compounds identified for each combination. The chemical diversity of the reference sets was evaluated by reducing the structures to molecular scaffolds (Table 6). The data obtained indicate a pronounced chemical diversity, denoting highly informative reference sets.

Table 6. T*he thirteen second targets selected for building focused ZINC compounds libraries. The number of ChEMBL compounds identified in the previous analysis is also reported. Chemical diversity of ChEMBL compounds is evidenced by the number of distinct molecular scaffolds to which they can be assigned.*

Second target	Uniprot ID	ChEMBL compounds set	Core Structures	Cyclic Carbon Structures	Reduced Cyclic Carbon Structures
Estrogen receptor β	Q92731	190	62	47	19
Estrogen receptor α	P03372	184	60	44	18
Tyrosine-protein kinase SRC	P12931	62	48	38	18
Serine/threonine-protein kinase Chk1	O14757	61	35	28	21
Vascular endothelial growth factor receptor 2 (KDR)	P35968	49	39	36	18

Vitamin D receptor	P11473	46	34	29	16
Glucocorticoid receptor	P04150	43	21	19	14
FK506 binding protein 12 (mTor)	P42345	41	21	17	12
Serine/threonine-protein kinase PIM1	P11309	30	17	13	6
Receptor protein-tyrosine kinase erbB-2, HER-2 Substrate	P04626	20	17	13	7
Apoptosis regulator Bcl-2	P10415	15	9	7	6
Mitogen-activated protein kinase kinase kinase 5 (ASK1)	Q99683	9	7	6	4
Serine/threonine-protein kinase B-raf	P15056	9	9	7	6

Similarity searching and SVM were performed (details in methods). These methodologies showed a certain degree of consensus in the ranking of ChEMBLdb. This behavior was further analyzed within the top positions of the ZINC rankings. Six rankings were obtained for each target combinations. The number of compounds shared within the top 1000 and top 2000 positions are annotated in Table 9.

Table 9. *Ranking consensus analyses and final focused compound libraries.*

Target	Consensus top1000	Consensus top2000	each top100 + cons. Top2000	each top200 + cons. Top2000
ASK1 (MAPKKK5)	13	21	356	765
Bcl2	13	25	456	851
Braf	0	0	426	833

Eralpha	11	22	440	839
Erbeta	2	10	445	868
erbB2	0	3	521	1037
mTor	10	17	421	848
SRC	0	1	493	980
VDR	13	25	467	947
Chk1	0	0	484	981
GR	116	211	401	675
KDR	1	1	507	1018
PIM1	126	238	384	619

In just 2 cases, namely Braf and Chk1, there was no consensus in the top 2000 positions, while between 1 and more than 200 shared compounds were identified in the remaining rankings. These data confirm an interesting level of consensus between SS and SVM ranking of molecular databases. For each target combination, the shared compounds in the top 2000 positions of the rankings were pooled together with the top 100 compounds positioned in each of the 6 rankings, leading to datasets including between 350 and 500 unique compounds. These libraries represent a first set of compounds that could be passed to the structure-based phase. Given the limited size of these libraries, they were expanded by adding the compounds listed in the top 200 positions of each ranking, leading to 13 final libraries of approximately 1000 compounds each (Table 9). This represents a reasonable size for the structure-based phase when the libraries will be docked in the binding site of the relative targets. In contrast to ChEMBL compounds that are known to be active at least against the second target, ZINC compounds will need to be docked and assayed both on Hsp90 and on the second target. Chemical diversity of the focused libraries was assessed by reducing the compound structures to molecular scaffolds. The data are shown in table 10, denoting a significant chemical diversity.

Table 10. *Chemical diversity within focused ZINC compound libraries.*

Target	Top100 + cons. Top2000	Core Structures	Cyclic Carbon Scaffolds	Reduced Cyclic Carbon Scaffolds
ASK1 (MAPKKK5)	356	166	84	23
Bcl2	456	113	73	23
Braf	426	172	111	21
Eralpha	440	120	70	21
Erbeta	445	179	114	26
erbB2	521	147	74	17
mTor	421	264	174	38
SRC	493	225	146	37
VDR	467	125	75	26
Chk1	484	173	116	34
GR	401	52	35	17
KDR	507	255	166	32
PIM1	384	79	43	16
Target	**Top200 + cons. Top2000**	**Core Structures**	**Cyclic Carbon Scaffolds**	**Reduced Cyclic Carbon Scaffolds**
ASK1	765	332	144	28
Bcl2	851	207	133	29
Braf	833	301	191	29
Eralpha	839	218	127	32

Erbeta	868	340	211	38
erbB2	1037	234	109	23
mTor	848	510	301	51
SRC	980	420	276	43
VDR	947	217	121	31
Chk1	981	329	206	44
GR	675	109	66	19
KDR	1018	504	318	43
PIM1	619	152	65	22

3.1.5 Structure-based virtual screening, ChEMBL compounds

The 13 sets of ChEMBL compounds were docked in the Hsp90 binding site in order to search for potential multi-target actives. An ensemble docking was performed including five different conformations of Hsp90 and considering three or four conserved water molecules (Table 11). The use of multiple Hsp90 conformations helps to account for the protein binding site flexibility. Figure 4 highlights a detail of the ATP binding site conformational change between the three X-ray crystal structures considered in this study.

Table 11. *Hsp90 structures used for docking. The number of conserved water molecules considered for each structure is also reported.*

Hsp90 Structure PDB ID	Water molecules
1OSF	3
1UY6	3
1UY6	4
2VCI	3

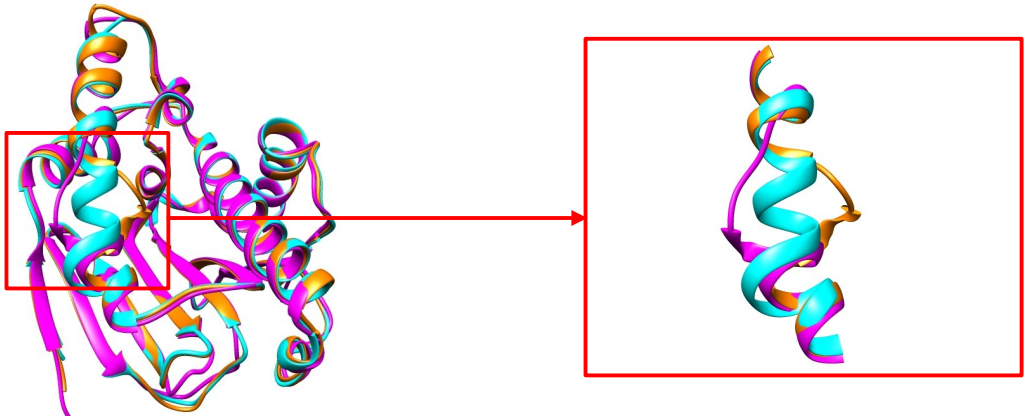

Figure 4. *Superposition of 1OSF (magenta), 1UY6 (cyan), and 2VCI (orange) X-ray crystal structure of Hsp90. A detail of the conformational change is shown.*

The docking calculations were carried out with AutoDock4 and post-processed with BEAR. The refined docking complexes were ranked according to their MM-PBSA energy score. Visual inspection of the docking complexes was carried out to check for the establishment of a hydrogen bond between the ligand and the residue Asp93 of Hsp90. This interaction is known to play a key role in Hsp90 inhibition. A molecular dynamics simulation of 20 ns in explicit water was also performed for the most promising complexes to validate the binding mode and assess more thoroughly the free energy of binding. The docking protocol was validated by checking for known Hsp90 actives in the first positions of the rankings. The data obtained are shown in Table 12. In the first 50 positions of the each ranking, two or more known Hsp90 actives were found. One or more Hsp90 actives were found in the to 10 positions of the rankings relative to the X-ray crystal structure. Only in the case of the homology model no Hsp90 actives were found in the first 10 positions of the ranking, suggesting lower confidence in those results compared to the X-ray crystal structures.

Table 12. *Hsp90 known inhibitors within the top 10 and 50 positions of virtual screening the rankings.*

Hsp90 Structure	Hsp90 actives within top 10 compounds	Hsp90 actives within top 50 compounds
1OSF	1	2
1UY6-3water	2	2
1UY6-4water	2	2
2VCI	1	5
Homology model	0	2

The final selection was affected by a rather low availability of ChEMBL compounds in commercial catalogs. Nevertheless, five of the best scoring compounds which displayed favorable interactions with Hsp90 binding site were purchased. The five compounds have activity annotations towards ER α, ER β, serine/threonine protein kinase ChK1 (ChK1) and Vitamin D receptor (VDR). The molecular structure of the five compounds is shown in Figure 5. Their ChEMBL activity annotations and energy scores are annotated in Table 11.

3.1.6 Biological assays, ChEMBL compounds

Biological assayswere performed to test the activity of the five compounds on Hsp90. All the compounds were tested at different concentrations for their ability to compete with Geldanamycin to bind the protein. Unfortunately, none of the selected compounds displayed any activity.

Table 11. *Activity annotations and energy scores (kcal/mol) obtained after BEAR and molecular dynamics in explicit solvent (MD).*

Compound	Second target	Activity	Hsp90 conformation	BEAR (MM-PBSA)	MD
CHEMBL370039	ER α	IC50 = 85700 nM	1UY6	-19.89	-21.6

CHEMBL370039	ER β	IC50 = 43000 nM	1UY6	-19.89	-21.66
CHEMBL585	ER α	Potency = 35 μM	1OSF	-17.43	-7.31
CHEMBL1344690	ER α	IC50 = 14432.57 nM	1UY6	-18.55	-25.42
CHEMBL211626	Chk1	IC50 = 300 nM	1OSF	-15.91	-20.88
CHEMBL119	VDR	Potency = 8.912 uM	Homology model	-26.92	-26.94

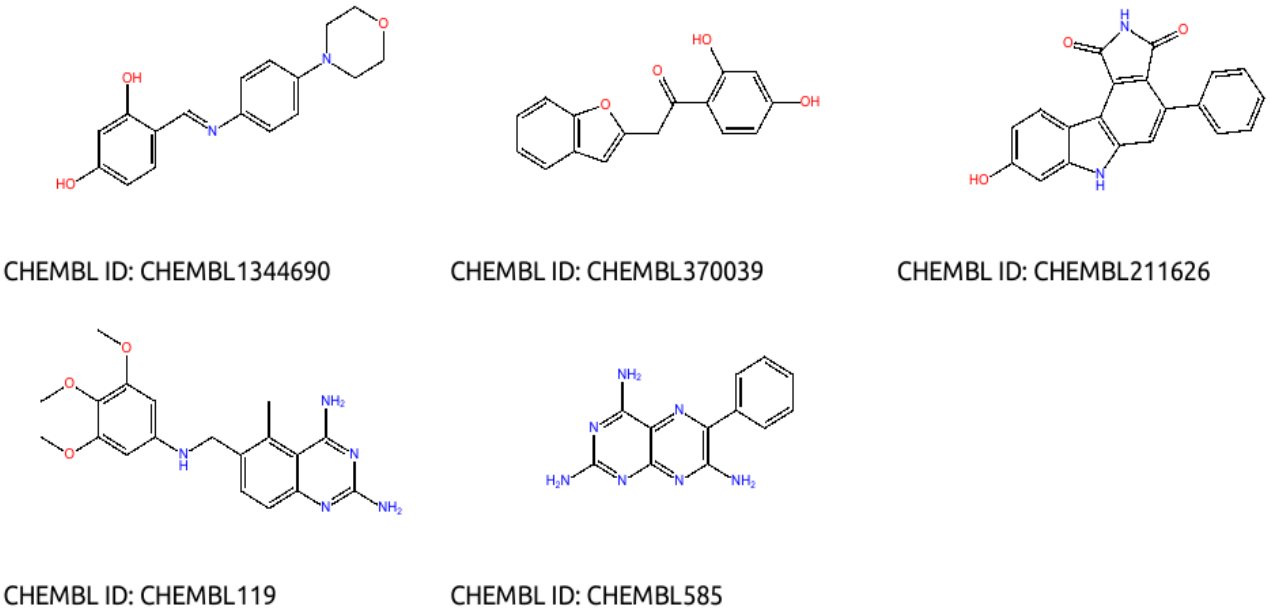

Figure 5. *Molecular structures and ChEMBL IDs of the five compounds chosen via structure-based virtual screening.*

3.1.7 Structure-based virtual screening, ZINC compounds

The ZINC libraries were docked inHsp90 and the relative second target with an improved protocol comprehending both AutoDock and Glide. Four second targets were prioritized due to biological assay feasibility:

- Glucocorticoid receptor (GR)
- Receptor tyrosine-protein kinase erbB-2 (erbB2)

- Serine/threonine-protein kinase B-raf (Braf)
- Proto-oncogene tyrosine-protein kinase Src (Src)

Table 12. *Second targets and PDB IDs of the X-ray crystal structures used for docking.*

Second Target	Structure PDB ID
erbB2	3PP0
Braf	3SKC
Braf	3D4Q
Src	2H8H
GR	1M2Z

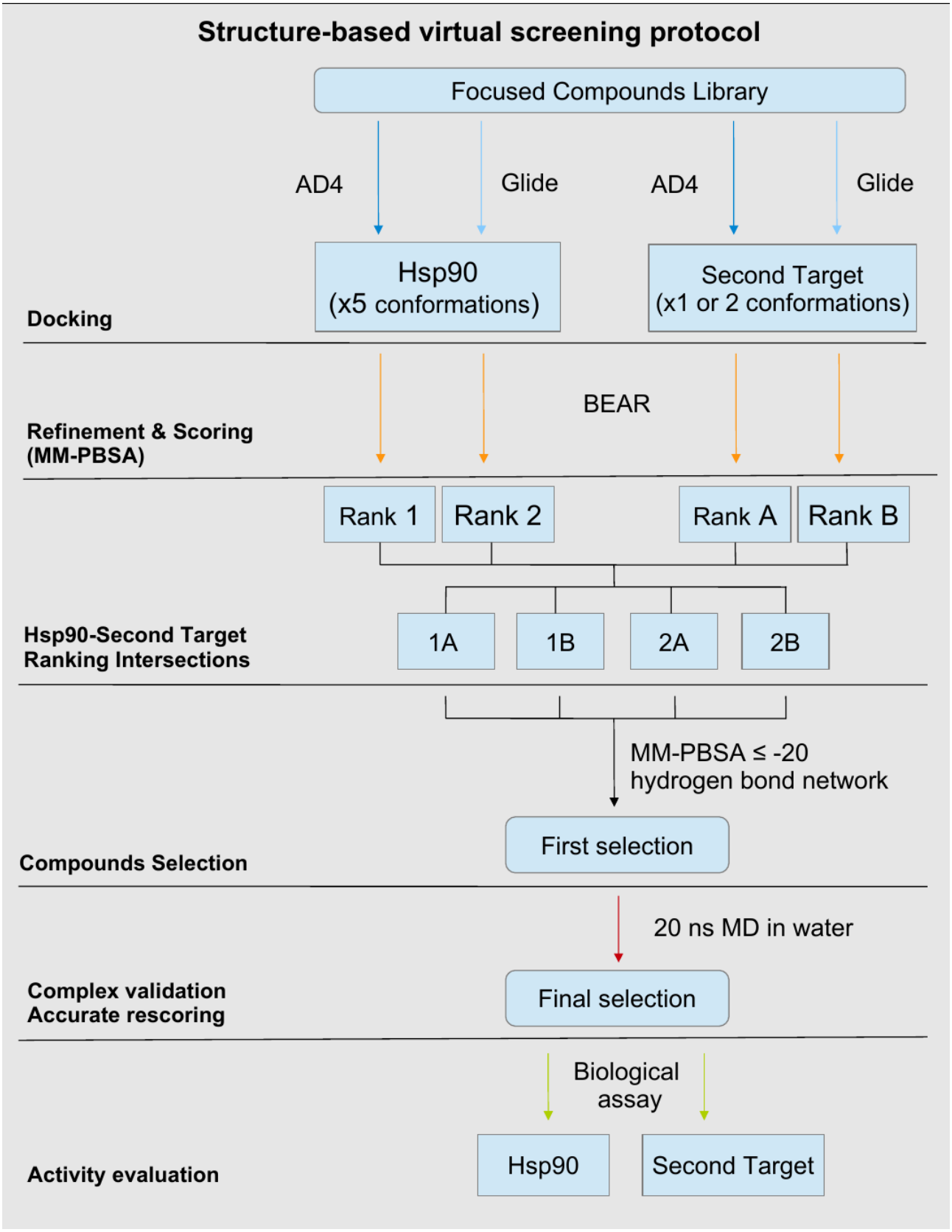

Figure 6. *Structure-based virtual screening protocol followed for the ZINC compound libraries.*

The virtual screening procedure is schematically represented in Figure 6. Eleven compounds plus an additional one (Gefitinib, entry id 3) were initially identified through the virtual screening procedure (Figure 7). Gefitinib, was added to the selection for comparison with O-desmetyl Gefitinib (entry id 2), the O-desmethyl metabolite of the drug Gefitinib. Gefitinib is a potent inhibitor of the epidermal growth factor receptor (erbB1). Eight of the twelve compounds were finally purchased for biological activity evaluation according to commercial availability and trying to maximize the chemical diversity of the selection. The energy scores obtained for the purchased compounds are summarized in Table 12. Molecular structures in Figures 7.

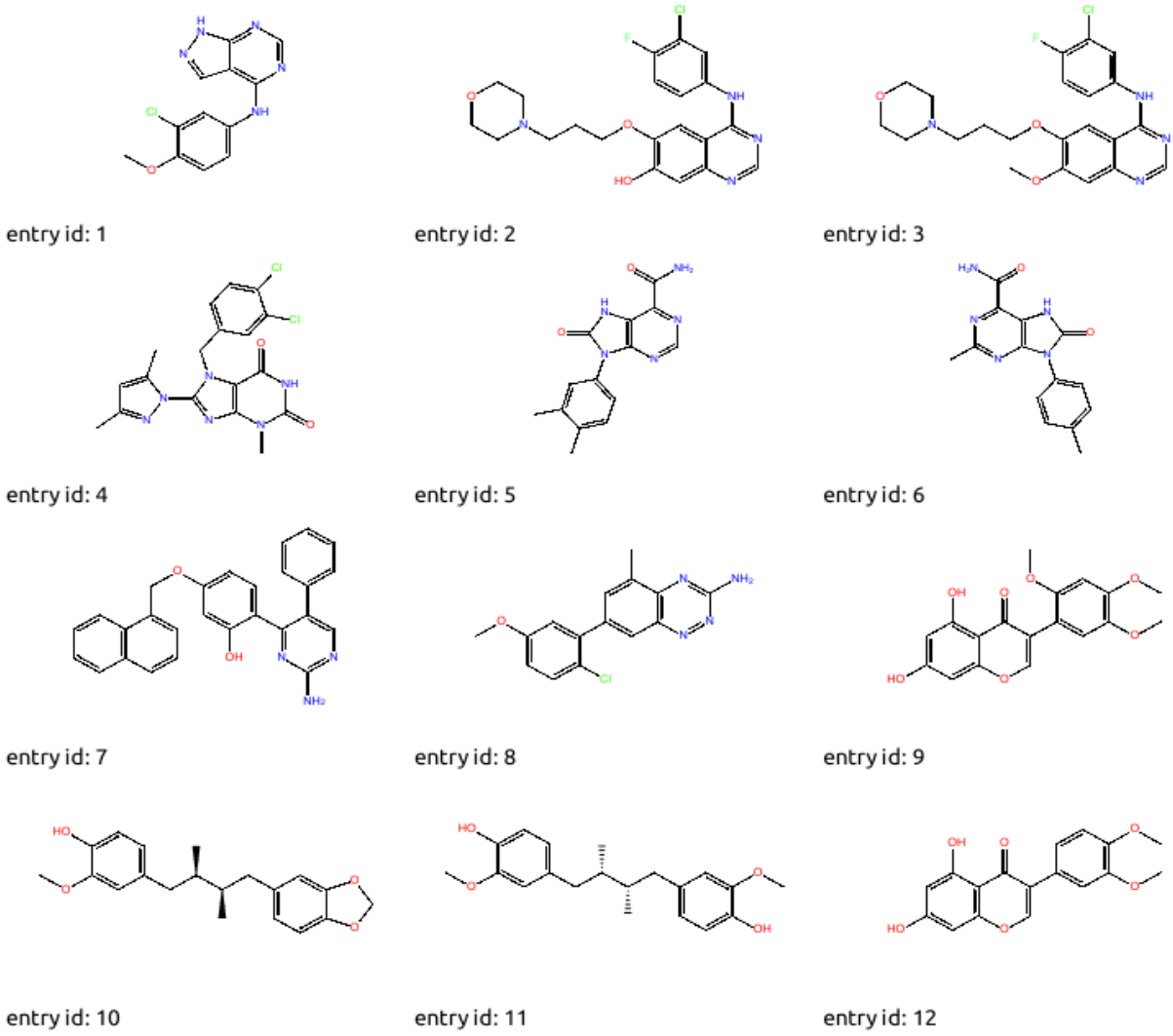

Figure 7. *Molecular structure of the 12 ZINC compounds initially identified via structure-based virtual screening.*

Table 12. ZINC compounds selection. S*econd targets and energy scores (kcal/mol) obtained after BEAR and molecular dynamics simulations (MD).*

Compound (entry id)	ZINC ID	Second target	BEAR 2ndT	Hsp90 confromation	BEAR Hsp90	MD Hsp90
1	11856423	erbB2	-27.05	1OSF	-21.81	-22.74
2	22056257	erbB2	-33.02	1UY6	-21.50	-32.34
3	Gefitinib*	Added for comparison (predicted inactive on Hsp90)				
4	12231331	B-raf	-29.16	1UY6	-30.24	-30.72
6	21036770	B-raf	-21.32	1OSF	-26.77	-18.28
7	6758932	SRC	-29.25	1UY6	-29.48	-26.76
10	14760153	GR	-31.31	1UY6	-20.78	-29.83
12	6556076	SRC	-23.03	1UY6	-23.95	-22.07
		GR	-23.05	2VCI	-23.80	-14.89

3.1.8 Biological assays, ZINC compounds

Biological assays were performed as briefly described in section 3.1.5 to assess the activity of the selected compounds on Hsp90. Moreover cell based assays were performed by incubating SkB3 cells with different concentrations of each compound to assess inhibition of the second targets.

Among the 8 compounds purchased, 4 were displayed activity in biological assays (Figure 8 and Table 13). Gefitinib and O-desmetyl Gefitinib (entry id 2 and 3) resulted active inhibitors of erbB2. The primary target of Gefintinib is erbB1 and it is not surprising that it is active also against erbB2, the same goes for its metabolite. In fact, erbB1 and erbB2 share significant homology (48% sequence identity) and similar ATP binding sites. Unfortunately, both compounds failed to exert any activity on Hsp90. Compound 4 displayed a slight activity at about 100 μMagainst Hsp90 while the result on the Braf were unclear.

Interestingly, the molecular structure of compound 4 is highly similar to known Braf inhibitors [79]. Compound 6 displayed good activity at about 3 μM on Braf but lacked activity on Hsp90. Compound 6 displayed good activity on Braf, but failed to do the same on Hsp90. According to these results and further binding mode visual inspection compound 4 and 6 were considered interesting for further elaboration via hit expansion. Their binding mode on Hsp90 and the second target were considered to have a potential for multi-target activity optimization. Figure 9 shows the binding mode predicted for the four compounds that displayed activity on either Hsp90 or the second target. Figure 10 shows the superposition of compound 4 with the Hsp90 inhibitor PU3 (extracted from 1UY6). The methylxanthine moiety of compound 4 adopts a conformation resembling the purine moiety of PU3, although they are not perfectly matched. The dichlorophenyl moiety of compound 4 overlaps the trimethoxybenzyl group of PU3. Most importantly, the hydrogen bond between PU3 and Asp93 can be established also by compound 4 by means of a secondary ammine.

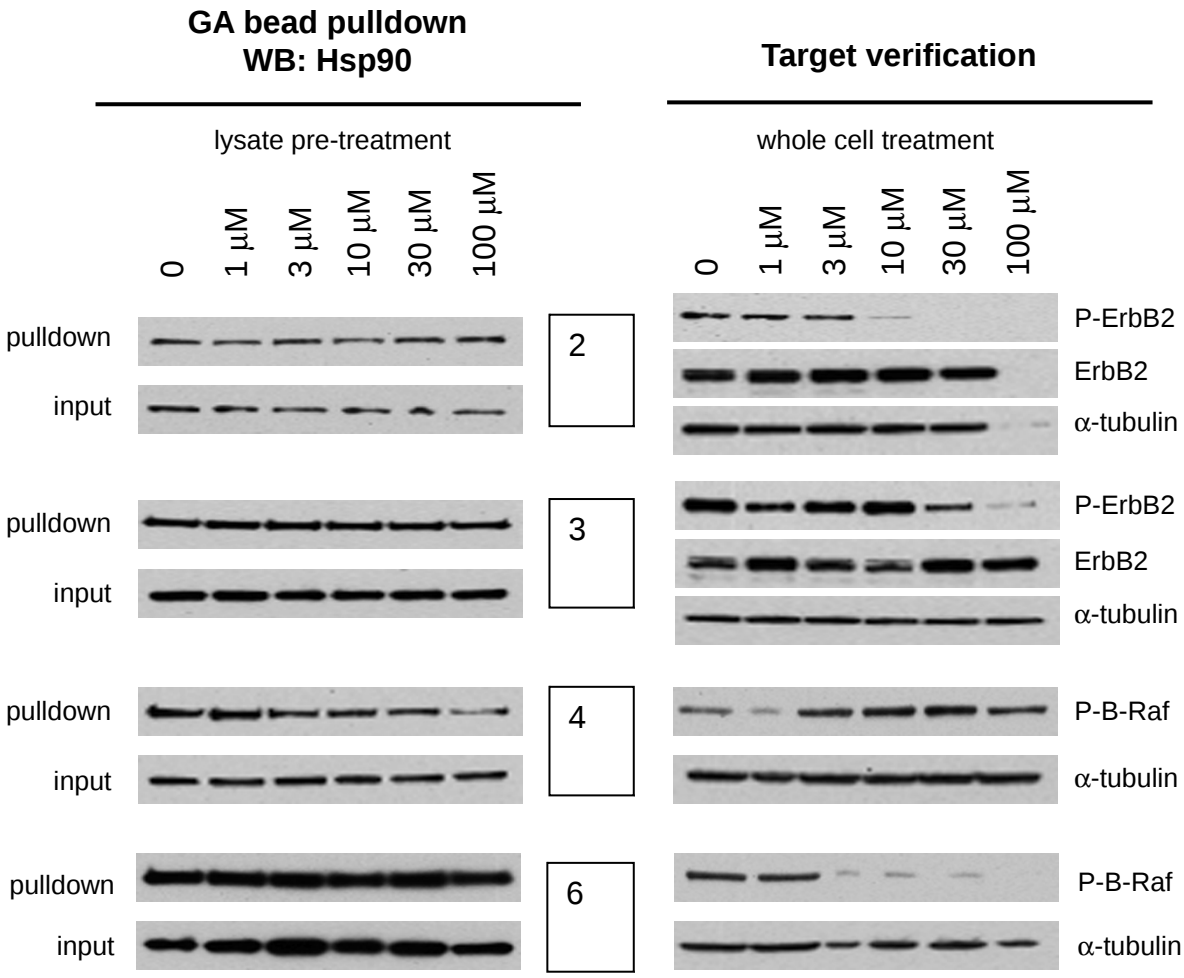

Figure 8. *Western blots resulting from biological assays performed to test the activity of*

compounds 2, 3, 4, and 6 on Hsp90 (left panels) and on the second targets (right panels).

Table 13. *Biological assays results.*

Compound	Hsp90 activity	Second target	Second target activity
2	Not active	erbB2	Active
3	Not active	erbB2	Active
4	Slightly active	Braf	Unclear (further testing)
6	Not active	Braf	Active

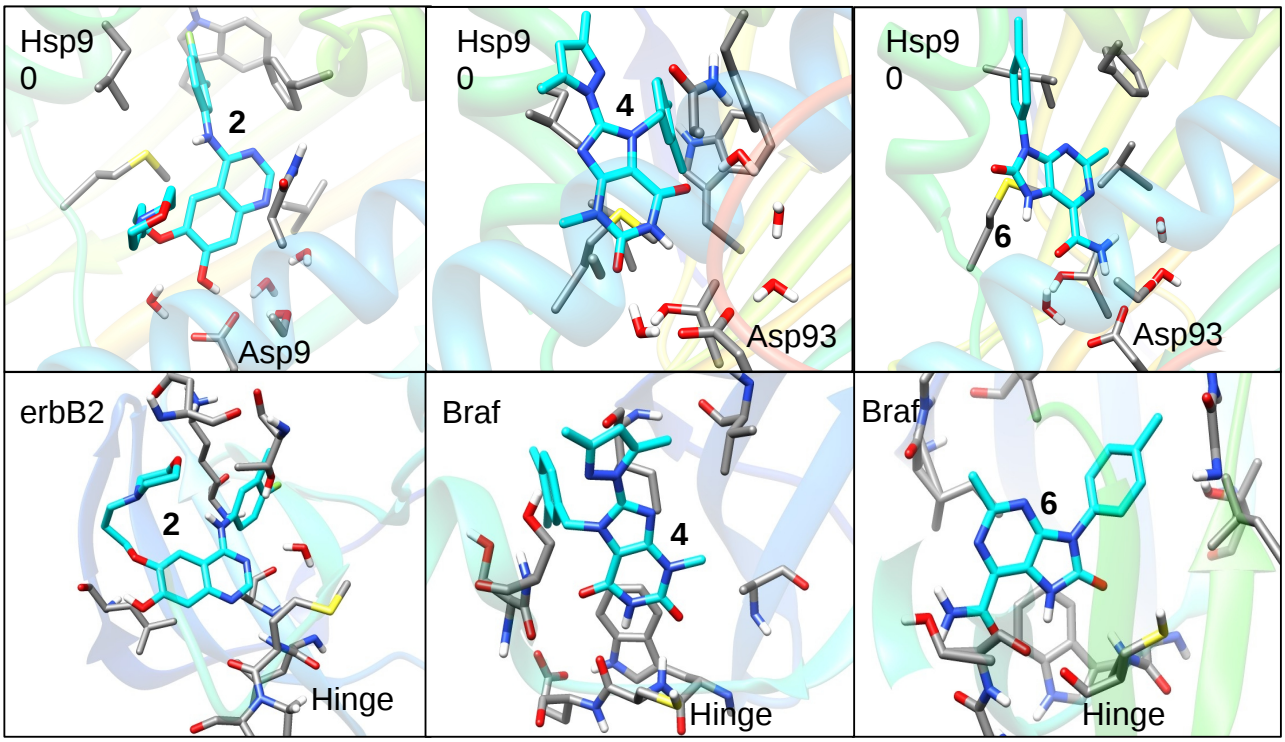

Figure 9. *Binding modes predicted by docking for the three ZINC compounds active on either Hsp90 or the second target selected in Hsp90. Binding modes in Hsp90 and the second targets are respectively represented in the top and bottom panels. Entry id of the compounds are reported. The key interaction sites (Hsp90 residue Asp93 and kinase hinge region) are indicated.*

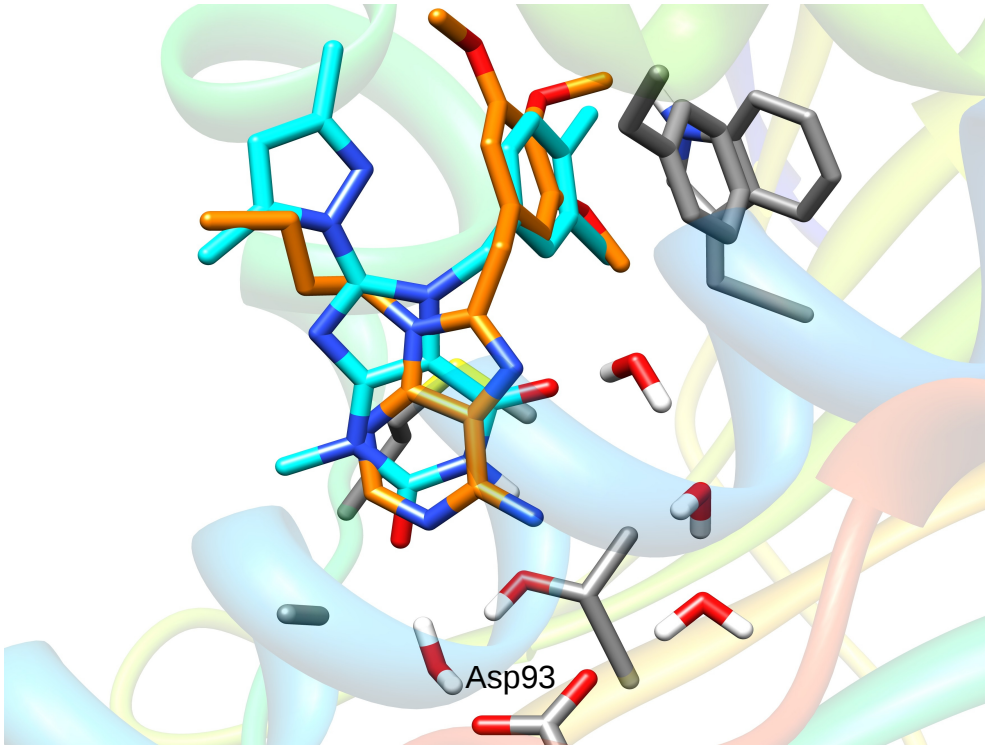

Figure 10. *Superposition of compound 4 (cyan) docked in the binding site of Hsp90 with the X-ray structure of the known inhibitor PU3 (orange).*

3.1.7 New libraries through hit expansion

Hit expansion was conducted on compounds 4 and 6 in the effort to obtain the desired multi-target activity profile. SS calculations were conducted to rank a database built by selecting up-to-date vendor catalogs. The screening procedure is schematically represented in Figure 11. A library including 2436 compounds was obtained for compound 4 a library of 1464 compounds for compound 6.

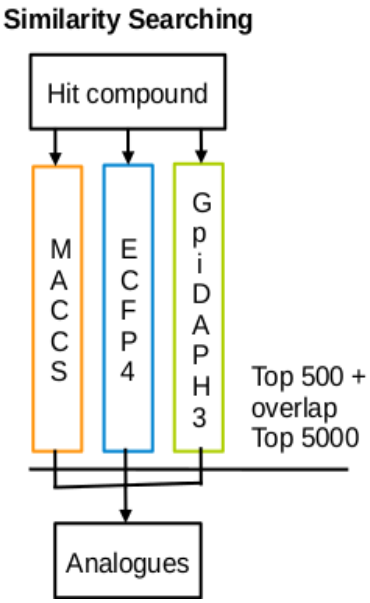

Figure 11. *Hit expansion procedure followed for compounds 4 and 6.*

3.1.8 New libraries from known dual-inhibitors expansion

The possibility to identify new hit compounds in the database of selected vendor catalogs was further explored by performing SS using known dual-inhibitors as reference. Three compounds with *high confidence* activity annotations for both Hsp90 and erbB2 (activities in Table 14, structures in Figure 12) were identified in ChEMBLdb. A procedure similar to the hit expansion described in section 3.1.7 was followed. A new library focused on Hsp90-erbB2, including 1927 compounds, was thus obtained.

Table 14. *Activtiy annotations of the known Hsp90-erbB2 dual inhibitor extracted from ChEMBLdb.*

ChEMBL ID	erbB2 activity	Hsp90 activity
CHEMBL109480	IC50 = 31 nM	IC50 = 20 nM
CHEMBL191635	IC50 = 280 nM	EC50 = 30 nM
CHEMBL193282	IC50 = 1700 nM	EC50 = 90 nM

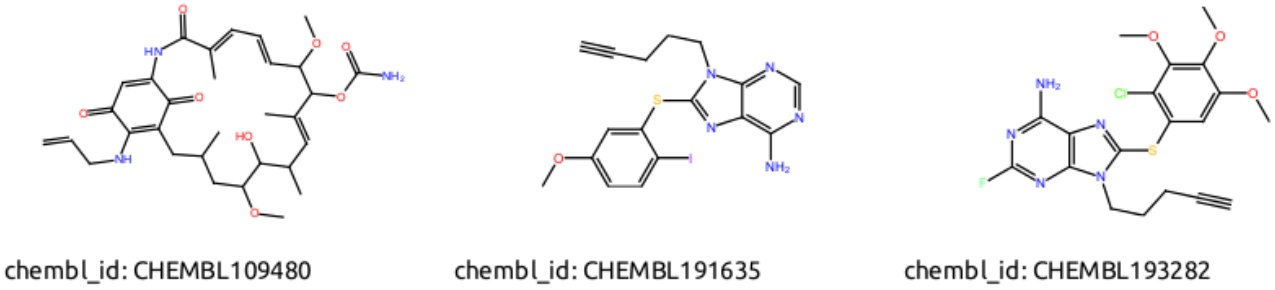

Figure 12. *Molecular structure of the known Hsp90-erbB2 dual-inhibitors extracted from ChEMBLdb.*

3.1.9 New libraries through rank fusion strategy

A new focused compound library was built for each of the target combinations earlier prioritized, that is Hsp90-Braf, Hsp90-erbB2, Hsp90-SRC, and Hsp90-GR, via rank fusion (details in methods). To this purpose, a new ligand-based protocol was set up and applied, schematically represented in Figure 13. Four new focused libraries including about 2000 compounds (Table 15) were obtained.

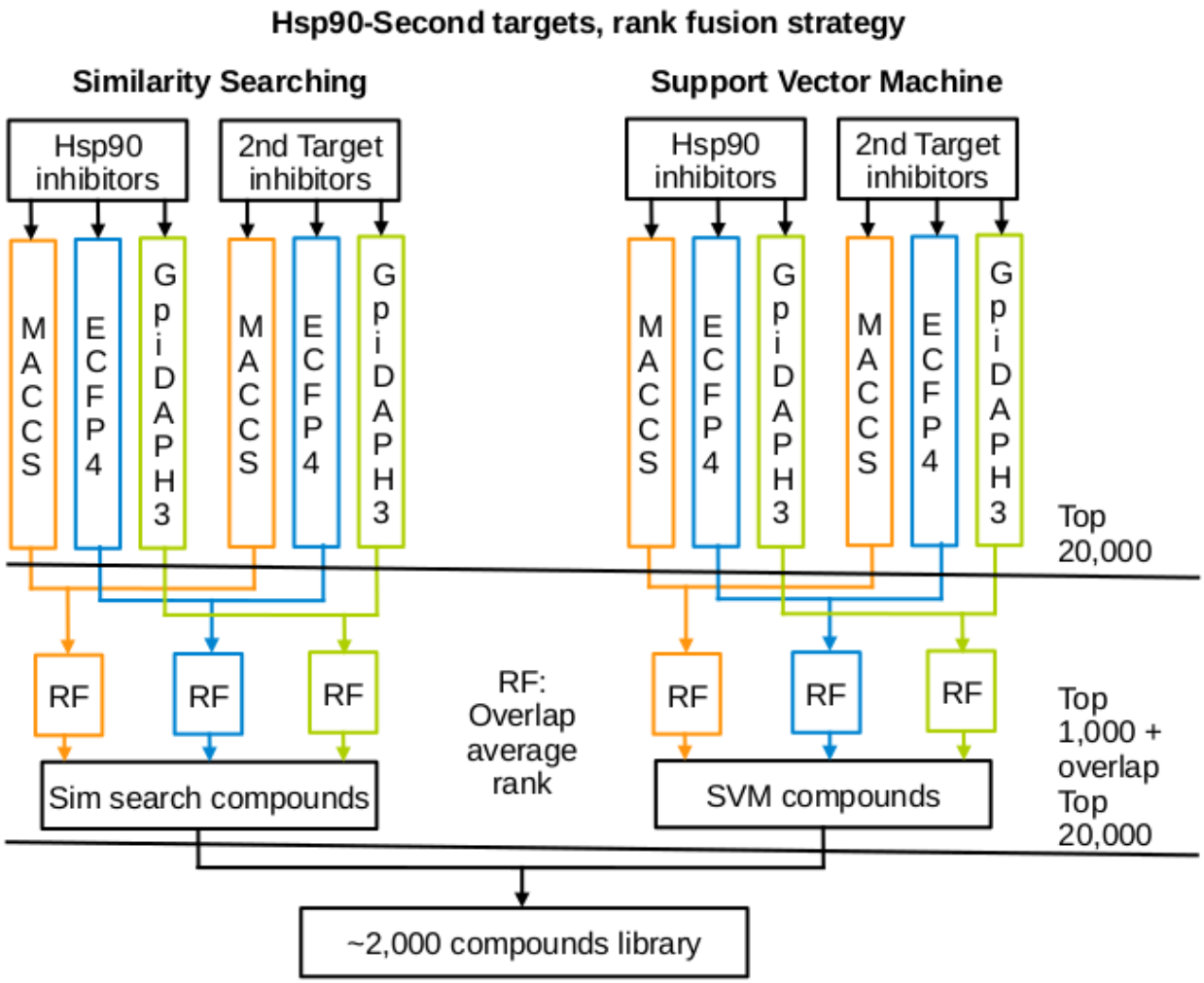

Figure 13. *Building of new focused libraries through rank fusion (RF) strategy.*

Table 15. *New focused libraries built by rank fusion for the four target combinations.*

Target combination	Number of compounds
Hsp90-Braf	2002
Hsp90-erbB2	2012
Hsp90-SRC	2038
Hsp90-GR	2000

3.1.10 Structure-based virtual screening of the new libraries

In total, seven new libraries were built. Two from hit expansion, one from expansion of known Hsp90-erbB2 dual-inhibitors, and four from rank fusion strategy. Each of these library was processed via docking (AutoDock and Glide) and post-docking (BEAR) as described in section 3.1.6. A new selection of 29 candidate dual-inhibitors was made and the compounds were sent for biological assays (Table 16). Twelve compounds were selected for Hsp90-Braf, 12 for Hsp90-erbB2, and three for Hsp90-SRC.

Table 16. *Twenty-nine compounds selected from new libraries.*

Compound (entry id)	Second target	BEAR Braf (MM-PBSA)	BEAR Hsp90 (MM-PBSA)	Library
13	Braf	-31.8	-27.82	Rank fusion
14	Braf	-33.63	-32.13	Rank fusion
15	Braf	-28.63	-27.79	Rank fusion
16	Braf	-31.52	-28.93	Rank fusion
17	Braf	-26.76	-30.88	Cmpd 4 exp.
18	Braf	-37.09	-30.9	Cmpd 4 exp.
19	Braf	-29.38	-32.15	Cmpd 4 exp.
20	Braf	-33.7	-27.93	Cmpd 4 exp.
21	Braf	-30.43	-24.26	Cmpd 6 exp.
22	Braf	-20.19	-34.96	Cmpd 6 exp.
23	Braf	-35.36	-29.63	Cmpd 6 exp.
24	Braf	-27.8	-30.01	Cmpd 6 exp.

25	Braf	-24.43	-29.14	Cmpd 6 exp.
26	erbB2	-27.06	-38.2	Rank fusion
27	erbB2	-40.27	-26.51	Rank fusion
28	erbB2	-31.51	-24.21	Rank fusion
29	erbB2	-28.22	-31.74	Dual-inh. exp.
30	erbB2	-35.15	-27.19	Dual-inh. exp.
31	erbB2	-11.43	-18.37	Dual-inh. exp.
32	erbB2	-34.85	-24.04	Dual-inh. exp.
33	erbB2	-42.29	-28.75	Dual-inh. exp.
34	erbB2	-26.37	-35.05	Dual-inh. exp.
35	erbB2	-21.19	-26.38	Dual-inh. exp.
36	erbB2	-30.76	-26.47	Dual-inh. exp.
37	erbB2	-39.55	-31.38	Dual-inh. exp.
38	erbB2	-22.41	-37.95	Dual-inh. exp.
39	SRC	-31.29	-28.84	Rank fusion
40	SRC	-37.29	-25.22	Rank fusion
41	SRC	-21.24	-30.82	Rank fusion

The twenty nine compounds are currently undergoing biological assays to asses their activity on both Hsp90 and the second targets.

3.2 Virtual screening on G-protein coupled receptors

A structure-based virtual screening approach including docking by AutoDock4 and post-docking processing by BEAR was validated for its ability to recognize GPCR known antagonists within a database of decoys. Moreover, the ability to rank in the first positions of the database multi-target active compounds was checked.

3.2.1 Redocking of cocrystallized Antagonists

The ability of the docking protocol to retrieve correct binding modes for A_{2a}, β_2, D_3, and H_1 was tested by performing redocking experimentse of the four cocrystallized antagonists. The root-mean-square deviations (rmsd) between the docked and the crystallographic poses range from 0.94 to 1.27 Å (Table 17). Considering that rmsd values ≤2 Å are usually taken as indicative of a correct docking pose, the docking parameters and the receptor structures used for redocking were appropriate. Furthermore, visual inspection of the docked binding modes confirmed that AutoDock4 was able to reproduce the pattern of molecular interactions seen in the crystal structures. The superposition between the docked and crystallographic binding modes confirmed that in all cases the docking method was able to retrieve correct binding geometries. In β_2, D_3, and H_1, the docked antagonists were involved in salt bridge formation between the conserved aspartic acid site and the protonated amine, and in A_{2a}, a correct hydrogen bond network between the primary amine of the antagonist ZM241385, Glu 169, and Asn 253 residues was predicted. These data provide evidence that AutoDock4 was able to predict binding modes consistent with published crystallographic information for these known GPCR antagonists.

Table 17. *Root Mean Square Deviation (rmsd) Values Obtained in the Redocking of the Cocrystallized Antagonists in Each Receptor Structure.*

receptor	PDB ID	cocrystallized antagonist	rmsd (Å)
β_2	2RH1	carazolol	1.27
A_{2a}	4EIY	ZM241385	1.67
A_{2a}	3EML	ZM241385	1.54

receptor	PDB ID	cocrystallized antagonist	rmsd (Å)
D_3	3PBL	eticlopride	1.03
H_1	3RZE	doxepin	0.94

3.2.2 Virtual Screening with AutoDock

Known antagonists extracted from GLIDA database and molecular decoys were docked into each receptor structure prepared as described above, using the same parameters used for redocking. Then, virtual screening performances were evaluated by calculating enrichment factors (EF). In the EF plots, the percentage rank of known ligands is plotted against the percentage position in the ranked database. The higher the enrichment factor, the better the performance of the method in identifying known ligands. The EF plots are shown in Figure 12. In addition, Table 18 reports the EF percentages at the 5% and 10% fractions of each ranked database, and Table 19 reports the percentages of correct binding modes of the known antagonists present in the GLIDA database after docking and postdocking. Binding modes of all antagonists were visually inspected to assess their geometrical consistency with the crystallographic binding modes. The ligand geometries were considered “correct” or “incorrect” by comparison with the binding mode of the crystallographic templates, paying particular attention to the formation of key molecular interactions with binding site residues according to the information in the literature and, when possible, comparing them with crystal structures of complexes with antagonists having similar structure. The EF plots (Figure 14) show that, except for β_2 and to a lesser extent A_{2a} (4EIY), AutoDock4 was generally able to provide enrichment curves better than the random selection curve. EF percentages (Table 18) ranging from 0 to 34 (EF5%), and from 5 to 55 (EF10%), were obtained. Regarding pose predictions, from 58% to 80% of the known antagonists had correct binding modes (Table 19). However, no correlation between the percent of correct binding modes reported in Table 19 and the EF percentages reported in Table 18 or shown graphically in Figure 14 could be observed. As an example, 80% of known β_2 antagonists had correct binding modes, but very low EF values were observed for this receptor.

3.2.3 Postdocking with BEAR

The AutoDock4 results were postprocessed with BEAR.

After the postdocking analyses, the EF plots were recalculated and included in Figure 14 for comparison. In the case of D_3, H_1, and A_{2a} (3EML), the EF plots obtained with MM-PBSA were qualitatively comparable or better with respect to those obtained with AutoDock4. Importantly, MM-PBSA remarkably improved the performance on A_{2a} (4EIY) and β_2 (2RH1) (Figure 14). The EF5% and EF10% values reported in Table 18 confirm that in all cases except for A_{2a} with the 3EML structure, a significant or dramatic improvement of EF values could be obtained after postdocking with BEAR. A_{2a} with the 4EIY structure gave higher EF values and higher percent of correct binding modes with respect to 3EML, suggesting that this higher resolution structure with the sodium ion may be more suited for virtual screening.

Table 18. *Enrichment factor (EF) Percentages of known active compounds retrieved by AutoDock4 (AD4) and BEAR at the 5% and 10% fractions of each ranked database.*

receptor	PDB ID	EF AD4 5% (%)	EF BEAR 5% (%)	EF AD4 10% (%)	EF BEAR 10% (%)
β_2	2RH1	5.00	40.00	5.00	75.00
A_{2a}	4EIY	0.00	26.09	13.40	47.83
A_{2a}	3EML	21.74	17.39	34.79	30.43
D_3	3PBL	16.70	26.20	31.00	42.86
H_1	3RZE	33.93	41.07	55.36	64.29

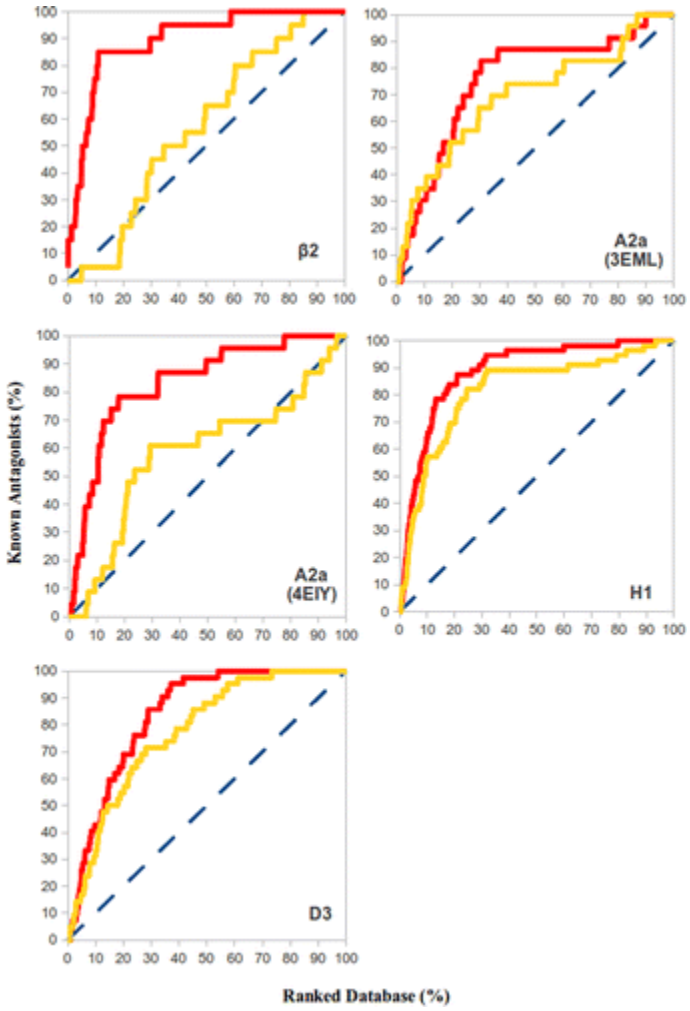

Figure 14. *Enrichment factor plots for each virtual screening. Known ntagonists retrieved (y-axis) are plotted against the ranked database (x-axis). Random selection (blue), Autodock4 enrichment (yellow), and BEAR MM-PBSA enrichment (red) curves are reported in each plot.*

The percentages of correct binding modes obtained after BEAR (Table 19) are identical or higher with respect to the ones obtained with AutoDock4. Furthermore, for many antagonists that were already classified as correct after AutoDock4, a better directionality of some key interactions with the receptor could be observed after structural refinement.

Table 19. *Percentages of correct binding modes of the known GLIDA antagonists observed after docking and postdocking.*

Receptor	PDB ID	correct binding modes AD4 (%)	correct binding modes BEAR (%)
β_2	2RH1	80.0	90.0
A_{2a}	4EIY	73.9	73.9
A_{2a}	3EML	65.2	65.2
D_3	3PBL	58.1	67.3
H_1	3RZE	64.3	66.1

3.2.4 Multitarget Approaches in GPCRs virtual screenings

In the GLIDA database used for docking, the H_1 and D_3 receptors had five antagonists in common, namely chlorpromazine, olanzapine, promazine, quetiapine, and ziprasidone. It is known that some antipsychotic drugs display affinity toward both receptors. Activity against D_3 is considered responsible for the therapeutic effect [80-81], while the binding of H_1 by antipsychotics is sometimes associated with some side effects like sedation and weight gain, nevertheless it mediates beneficial effects for mania, depression, and psychosis therapy through the treatment of insomnia [82].

In a virtual screening effort aimed at the discovery of small molecules with multi-target activity profiles, the candidate molecules would be ideally located within the best scoring positions of each desired target ranked list. Therefore, the ability of the screening protocol to identify these five antagonists in the first positions of each respective ranked list was investigated (Table 20). The five compounds were retrieved within the first 16% (H_1) and 20% (D_3) positions of each ranked database after docking and within the first 9% (H_1) and 13% (D_3) after postdocking. This finding allowed us to evaluate that the virtual screening was able to correctly identify multitarget ligands, at least within the limited set of receptors and database molecules analyzed in this study. It will be interesting to use this workflow on other test cases and with more extended databases of molecules. GPCRs, like protein kinases, will undoubtedly constitute a superfamily of targets in which polypharmacology approaches will play a key role. For this reason, future work should be directed toward predicting polypharmacology profiles of GPCR ligands.

Table 20. *Ranking positions (expressed as percentages with respect to each ranked database) of the five antagonists in common between the H 1 and D 3 receptors by AutoDock4 and BEAR.*

Antagonist	Ranking AD4 D_3 (%)	Ranking AD4 H_1 (%)	Ranking BEAR D_3 (%)	Rranking BEAR H_1 (%)
chlorpromazine	11.24	2.79	10.69	1.30
olanzapine	8.76	15.92	12.79	2.70
promazine	20.50	6.17	12.99	1.70
quetiapine	20.70	6.77	5.29	7.29
ziprasidone	0.90	2.19	3.80	8.59

3.3 Virtual screening on Cyclin-Dependent Kinase 2

Using the crystal structure of CDK2 in complex with the allosteric probe ANS (PDB code 1PXF), a virtual screening of commercially-available compounds in this newly discovered allosteric pocket was performed using a combination of docking (AutoDock 4) and post-docking (BEAR) methods.

3.3.1 ANS redocking

Before database screening, the docking protocol was tested for its ability to reproduce the crystallographic orientation of ANS. Re-docking of ANS retrieved a binding mode that was almost identical to that of the co-crystallized ligand (RMSD of 0.529 Å). Of the two ANS molecules bound to the crystal, AutoDock placed the ligand in the inner (residue 305.A in 3PXF, named ANS1) and conceivably more favourable binding site, with a score of -7.4 kcal/mol. According to BEAR, the binding free energies of ANS in the inner (ANS1) and outer (ANS2) pockets were -37.7 and -24.6 kcal/mol, respectively, confirming that the inner pocket provides higher binding, at least with this ligand.

3.3.2 Structure-based virtual screening, Asinex compounds

About 600,000 compounds from Asinex were processe thorugh structure-based virtual screening (details in methods section).

The selection of compounds for biological evaluation was made from the hundred top-scoring compounds ranked according to the MM-PBSA scoring function implemented in BEAR. In the top hundred compounds, MM-PBSA scores ranged from -59.1 to -47.1 kcal/mol. A common feature to all top-scoring compounds was a strong ionic interaction

with Lys33, *i.e.* the catalytic lysine of CDK2 that lines one side of the inner ANS1 pocket. Moreover, all compounds occupied the ANS1 pocket with a combination of complementary electrostatic (Lys33) and hydrophobic (Leu55, Phe146, Tyr15, Phe80, Leu78, Leu148, Leu66) interactions, many of them extending - in all or in part - to the outer ANS2 pocket thus creating additional favourable contacts. In the outer pocket, a number of top-scoring compounds formed additional electrostatic interactions with Lys56 (αC helix) and His71 (β4sheet), and hydrophobic interactions with Ile52, Leu76, Leu78, Leu37, Val69, and Ile35. To promote the selection of structurally diverse compounds, potential hits were grouped into chemical classes on the basis of Tanimoto similarity indexes and clustering analyses. Then, compound selection was performed by *i)* selecting at least one representative of each cluster of compounds; *ii)* the visual inspection of the binding modes; *iii)* the analysis of chemical groups interacting with Lys33. A total of thirty-five compounds were finally purchased and submitted to biological assays.

3.3.3 Biological assays

Biological assays were performed in the group of Massimo Broggini, Istituto di Ricerche Farmacologiche Mario Negri, in Milano.

Compounds were initially screened for their ability to bind CDK2 and displace ANS from the allosteric pocket. Intrinsically fluorescent compounds were not considered for further characterization. Competition experiments showed that seven compounds (compounds **1-7** in Chart 1) had a concentration-dependent ability to displace ANS from CDK2, with EC_{50} values in the micromolar range. Selected computational and biological properties of these compounds are reported in Table 21. Four of the seven hits have significantly diverse structures (having average pairwise Tanimoto similarity coefficients of around 0.2)

and have been classified as representatives of carbonyl, nitro, and carboxylic acid classes of compounds depending on the chemical group interacting with Lys33. ADMET predictions performed with QikProp (Schrödinger Suite 2011) showed that almost all the predicted properties of these compounds fall within the 95% range of similar values for known drugs. Therefore, the identified compounds may constitute valuable starting points for further hit-to-lead optimization.

Table 21. *Predicted free energies of binding, VS rankings, chemical classification, occupation of the ANS1 and ANS2 pockets, selected physico-chemical properties, and biological properties of the seven allosteric inhibitors discovered in the primary virtual screening (compounds 1-7) and of the nine analogues of compound 4 selected from hit expansion (compounds 4a-4i).*

Cpd #	Asinex #	$\Delta G'_{bind}$[a]	Rank #[b]	Class[c]	ANS1[d]	ANS2[d]	MW[e]	PSA[f]	logP[g] oct/w	EC_{50} (μM)[h]	IC_{50} (μM)[i] MDA-MB231	IC_{50} (μM)[i] ZR-75-1
1	BAS01060589	-55.9	2	carb. acid	x	x	492.0	132.9	4.8	7±3	>50	>50
2	BAS00631909	-51.6	14	nitro	x	x	515.4	102.8	5.6	3±1	20.5±4.4	16.7±2.2
3	BAS00590245	-51.4	16	carbonyl	x		514.3	64.1	4.7	45±16	18.6±2.5	25.9±1.0
4	BAS00380830	-50.9	21	nitro	x		438.5	108.9	5.0	71±28	4.0±0.1	4.5±0.2
5	BAS00434932	-49.5	29	nitro	x		385.5	61.5	4.2	27±6	>50	>50
6	BAS01060577	-49.2	33	nitro	x	x	449.9	134.2	3.7	32±7	49.2±1.2	>50
7	BAS01123443	-54.2	5	carb. acid	x	x	463.9	134.8	3.9	48±18	>50	>50
Hit expansion of compound **4**												
4a	BAS02102259	-50.8	2[l]	nitro	x		506.6	127.1	5.6	>100	21.1±1.5	43.3±9.5
4b	BAS00619651	-50.7	3	nitro	x		486.6	112.6	6.0	88±57	>50	>50
4c	BAS01404025	-49.8	6	nitro	x		502.6	121.6	5.7	49±21	>50	>50
4d	BAS02102292	-49.4	7	nitro	x		520.6	126.8	5.7	23±12	>50	>50
4e	BAS02102245	-49.4	8	nitro	x		520.6	124.8	5.7	46±14	>50	>50
4f	BAS01547732	-48.9	13	nitro	x		492.5	130.1	5.1	76±39	19.4±1.9	21.1±1.2
4g	BAS00111586	-48.8	15	nitro	x		446.5	111.2	5.0	60±37	6.5±1.0	5.6±0.5
4h	BAS00916022	-48.0	20	nitro	x		508.6	114.0	6.4	>100	>50	>50
4i	BAS00381203	-47.2	24	nitro	x		398.5	110.2	4.3	>100	23.4±5.5	21.4±0.3

EC_{50} values obtained range from 3 μM compound 2 to 71 μM for compound 4. Compounds 1 (EC_{50} of 7 μM) and 2 (EC_{50} of 3 μM) bound CDK2 with an affinity higher than ANS (K_d=37 μM)[11], while compounds 3 and 5-7 had similar affinity. To confirm the truly allosteric nature of these ligands, competition experiments between compounds 1-7 and ANS were also

performed in the presence of a potent ATP-competitive type I inhibitor that blocks the ATP site, *i.e.* staurosporine. For all seven compounds the presence of staurosporine did not modify significantly the ANS displacement curve, indicating that they do not occupy the ATP site but bind to a CDK2 allosteric site.

The seven compounds were then tested for their ability to inhibit the growth of human breast cancer cells *in vitro*, using two different human breast cancer cell lines, MDA-MB231 and ZR-75-1. As reported in Table 21, the most effective compound was 4, with a concentration able to inhibit the growth by 50% (IC_{50}) of 4 and 4.5 μM for MDA-MB231 and ZR-75-1, respectively. Four of the seven compounds (1, 5-7) did not show appreciable activity (IC_{50} >50 μM). Compounds 2 and 3 showed IC_{50} values around 20 μM. Compounds 1 and 2 were the most effective in displacing ANS from CDK2, but only 2 showed appreciable activity in cell-based assays. Moreover, compound 4 was the less efficient in binding CDK2 but the most active in the cell-based assays. These findings indicate the lack of a clear correlation between *in vitro* CDK2 binding characteristics of these compounds and their effect on cells, a commonly observed phenomenon generally arising from differential cell membrane permeability of compounds [83-2]. Nevertheless, compound 4 was able to effectively inhibit CDK2-mediated phosphorylation of Retinoblastoma at a concentration of 4 μM (its IC_{50} in cells).

3.3.4 Binding mode validation

The predicted binding modes of the seven active hits were confirmed by performing a more exhaustive sampling and clustering of docked orientations compared to the original conditions applied for virtual screening. The results are collectively reported in Tables S1

and S2 (available as supporting information), and the overall best-scoring solutions according to BEAR/MM-PBSA chosen as representative binding modes of each compound are graphically shown in Figure 15. Compounds 1 and 7 are predicted to bind the inner catalytic Lys33 with the 6-carboxylate group and the outer Lys56 group with the 2-nitro group of the 2-nitro-phenylsulfanyl moiety. The remaining interactions were mainly hydrophobic contacts with Tyr15, Ile35, Ile52, Leu55, Val69, Leu76, Leu78 and Phe80. Compounds 2 and 6 bind Lys33 with the nitro group of the 2-nitro-phenylsulfanyl moiety, and Lys56 and His71 with the carboxylate group. Being characterized by two polar heads separated by a hydrophobic core, head to tail orientations of compounds 1/7 *vs* 2/6 having each polar group interacting once with Lys33 and once with Lys56 are reasonable. Other interactions are mainly hydrophobic. Compound 3 binds Lys33 with the pyrimidin-4-one carbonyl and interacts favourably with Tyr15, Ile35, Leu55, Leu66, Leu78, Phe80, Leu148, Val154 and Val156. The nitro group of compound 4 interacts with Lys33, and the carbonyl group of the cyclohex-2-enone ring hydrogen bonds with the Tyr15 hydroxyl. The nitrophenyl ring establishes stacking interactions with Tyr15 while the rest of the molecule is in contact with Ile35, Ile52, Leu55, Val64, Leu66, Leu78, Val154 and Val156. Compound 5 binds Lys33 with the nitro group while the remainder of the molecule interacts with Tyr15, Ile35, Ile52, Leu55, Ile63, Val64, Leu66, Phe80, Phe146, Leu148, Val154 and Val156.

3.3.5 Hit expansion of compound 4

With the aim of potentially identifying additional biologically active hits and to provide a first exploration of their SAR, commercially available analogues of compound 4, *i.e.* the compound showing the most interesting activity in cancer cell-based assays, were searched for in the Asinex database. To this end, a focussed library of commercial

analogues of 4 was prepared by including compounds with a similarity Tanimoto coefficient (Tc) of at least 0.7 with this reference compound. This operation resulted in 2217 compounds, 1536 of which were not post-processed with BEAR in the primary virtual screening owing to the AutoDock energy score cut-off of -11 kcal/mol imposed when selecting molecules for post-docking refinement and rescoring. The entire focussed library (2217 compounds) was then submitted to post-docking using the same procedure described for the primary screening. According to MM-PBSA, compound 4 ranked first in this focussed library, but a number of analogues with interesting scores emerged, some of which were not processed before because their AutoDock scores were less favourable than the cut-off value. Nine analogue compounds (compounds 4a-4i, included in Chart 1) with favourable MM-PBSA scores and binding modes consistent with that of the parent compound 4 were finally selected for biological evaluation. Their chemical and biological properties are included in Table1.

Three compounds (4a, 4h, 4i) did not show appreciable ANS displacement activity (Table 1). The remaining six analogues had EC_{50} values ranging from 23 to 88 μM. In some cases, *e.g.* compound 4d, the affinity for CDK2 was slightly improved with respect to that of the parent compound 4. Moreover, the compounds showing appreciable ANS displacement activity were also tested in the presence of staurosporine to confirm that they bind in the allosteric ANS pocket. The displacement ability of the six tested compounds did not change in the presence of staurosporine, confirming their allosteric binding. Again, none of the compounds showed direct inhibition of the activity of the CDK2/cyclinA complex up to a concentration of 100μM (data not shown). These data are in line with those of initially tested compounds 1-7.

When tested in cell based assays on both MDA-MB231 and ZR-75-1 human breast cancer cell lines, compound 4g showed IC_{50} values similar to those of compound 4, while

compounds 4a, 4f and 4i had IC_{50} values around 20 μM. The remaining five compounds did not show appreciable antiproliferative activity (Table 21).

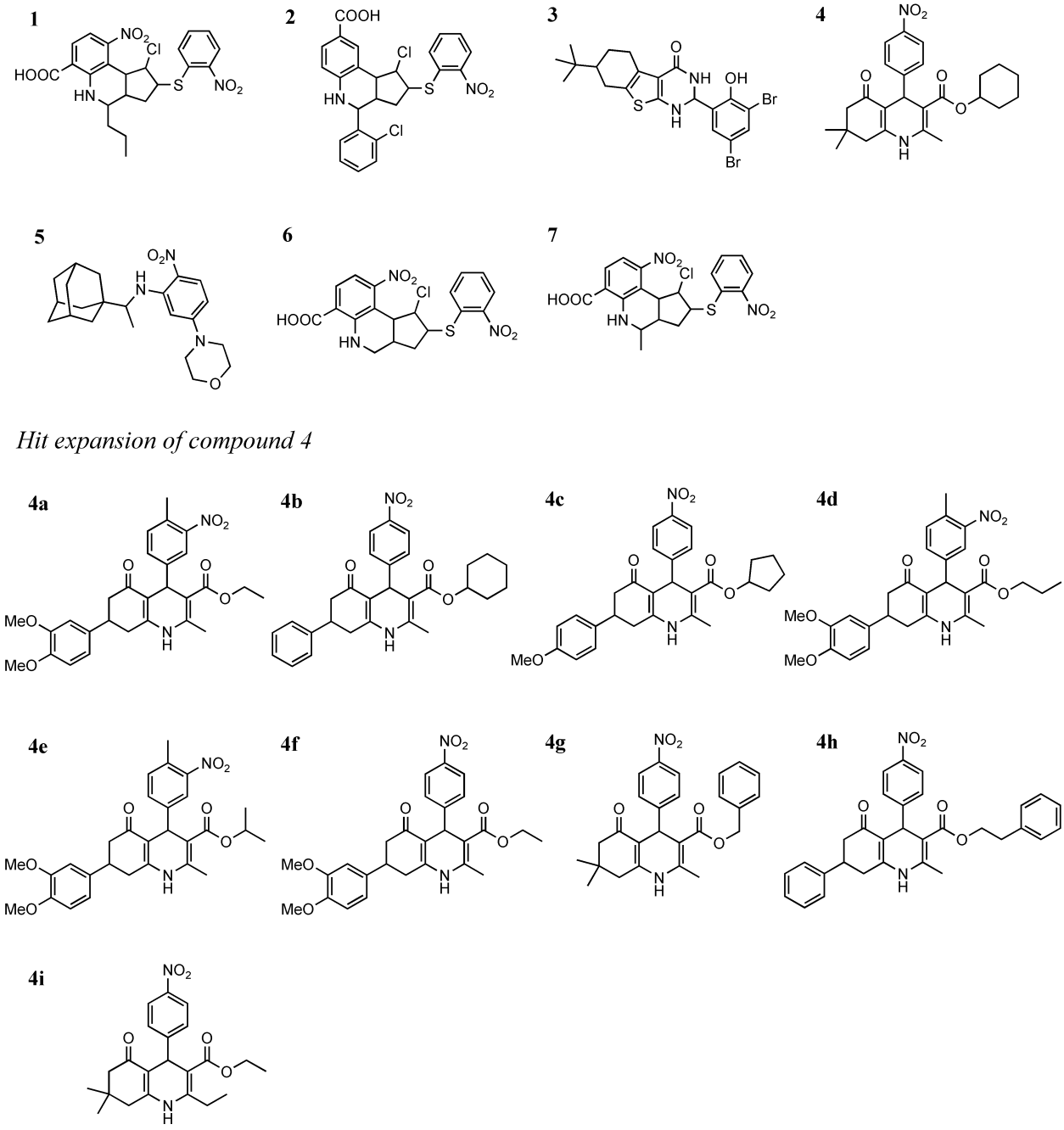

Figure 15. *Molecular structure of CDK2 allosteric inhibitor discovered via structure-based virtual screening.*

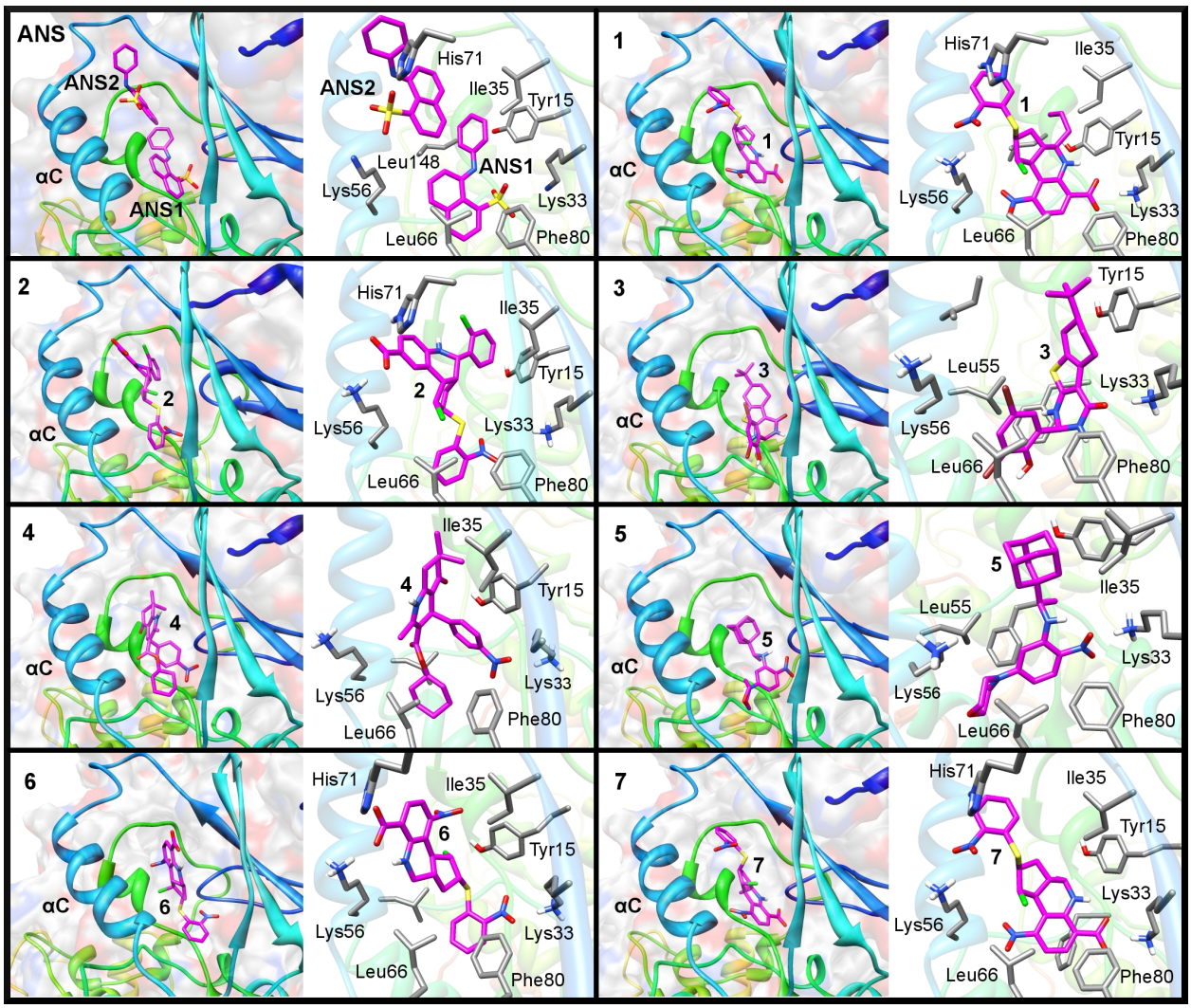

Figure 16. *Binding mode of ANS and of the allosteric inhibitors discovered.*

4. Conclusions

In my PhD work, computational methods including ligand- and structure-based methodologies have been extensively validated and applied *i)* for the identification of suitable target combinations for polypharmacology approaches, *ii)* for the design of ligands with potential multi-target activity, *iii)* for the ability to recognize active compounds within a database of molecular decoys, *iv)* and for the ability to predict new active compounds confirmed in biological assays. Several protein targets were investigated, including the molecular chaperone Hsp90, GPCRs (β_2, A_{2a}, D_3, H_1 receptors), and protein kinases (Braf, erbB2, Src, CDK2, and others).

Several computational methods including analysis of information deposited in public molecular databases, ligand- and structure-based virtual screenings, were combined to identify suitable target combinations for a polypharmacology approach within the Hsp90 interactome. Activity annotations of compounds active against the Hsp90 interactome have been extensively analyzed. Twenty-six known compounds with an experimentally-validated multi-target activity profile were identified, inhibiting Hsp90 and one or more second targets within a set of 17 Hsp90 interactors. A ligand-based virtual-screening protocol including SS and SVM calculations identified 81 targets that in combination with Hsp90 represent potential target combinations for dual-inhibitor design. Thirteen target combinations were selected to build focused libraries of ChEMBL and ZINC compounds. Three active compounds were identified through a structure-based protocol including docking and post-docking processing. In particular, two compounds of the ZINC library belonging to the Hsp90-Braf target combination displayed activity on either Hsp90 or Braf and were considered suitable for hit expansion. Therefore, a new round of ligand-based calculations was carried out to expand the hits identified and to build new focused compound libraries, that were then analyzed with structure-based calculations. Currently, twenty-nine compounds are undergoing biological assays.

A structure based-virtual screening protocol including docking with AutoDock4 and post-processing with BEAR was validated on four GPCRs with known crystal structures. Considering that GPCR binding sites are flexible and may experience conformational rearrangements for accommodating different antagonists, the results obtained confirm a

good performance of AutoDock4 in retrieving correct binding modes using a single receptor conformation but suggest that the scoring function had difficulties in discriminating true binder from decoys, even when ligands were docked correctly. Rescoring docking complexes with MM-PBSA improved the virtual screening results, implying optimistic expectations for prospective campaigns using these methods. The performance of the protocol in ranking multi-target active compounds on two receptors was also validated. The results obtained suggest that these screening tools may not only help in scoring favorably ligands specifically active on each receptor, but also in ranking favorably potential multi-target hits shared by different GPCRs. This result may be useful for selecting GPCRs hits with a desired multi-target profile, as well as for the prediction of potential undesired off-target effects.

The recent availability of the first crystal structure of a member of the CDK family with an open allosteric pocket in proximity of αC helix allowed us to apply a virtual screening protocol that led to the successful identification of the first-in-class allosteric inhibitors of CDK2. Using a combination of high-throughput docking and refinement and rescoring performed with our post-docking tool BEAR, we rationally selected and tested thirty-five top-ranked compounds and finally identified seven ligands able to bind CDK2 in the allosteric ANS pocket, corresponding to a hit rate of 20%.

The active ligands were able to displace ANS from CDK2 at micromolar concentrations and showed a mechanism fully compatible with the expected allosteric mechanism. Competition experiments performed in the presence of the ATP-competitive inhibitor staurosporine confirmed the truly allosteric nature of these ligands. Importantly, some of these ligands inhibited the proliferation of MDA-MB231 and ZR-75-1 breast cancer cells with IC50 values in the low micromolar range. In particular, compounds 4 and 2 had IC50 values of 4 μM and 20 μM, respectively. Hit expansion of compound 4 provided an additional ligand 4g with similar in vitro potency. These compounds constitute valuable hits for further optimization of drug potency, cellular penetration, and SAR investigations.

5. Publications

- Giulio Rastelli, Andrew Anighoro, Martina Chripkova, Laura Carrassa, Massimo Broggini “Structure-Based Discovery of the First Allosteric Inhibitors of Cyclin-Dependent Kinase 2” *Cell Cycle*, accepted for publication.
- Anighoro Andrew, Giulio Rastelli. “BEAR, a Molecular Docking Refinement and Rescoring Method” *Computational Molecular Bioscience*, **2013**, 3, 27-31.
- Andrew Anighoro, Dagmar Stumpfe, Kathrin Heikamp, Jürgen Bajorath, Giulio Rastelli. “Targeting the Hsp90 interactome using in silico polypharmacology approaches” *La Chimica e l'Industria*, **2013**, 4, 105-106.
- Andrew Anighoro, Giulio Rastelli. “Enrichment factor analyses on G-protein coupled receptors with known crystal structure” *Journal of Chemical Information and Modeling*, **2013**, 53, 739-743.
- Miriam Sgobba, Fabiana Caporuscio, Andrew Anighoro, Corinne Portioli, Giulio Rastelli. “Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations” *European Journal of Medicinal Chemistry*, **2012**, 58, 431-440.

6. Oral communications

- “3rd Meeting of the Paul Ehrlich MedChem Euro-PhD Network”, 27th -29th September 2013, Santa Margherita di Pula, Cagliari. Oral communication: “A computational polypharmacology protocol within the Hsp90 interactome”;
- “Computationally Driven Drug Discovery”, 4th-6th February 2013, hosted by the Italian Institute of Technology, Genova. Oral communication: “Targeting the Hsp90 interactome using in silico polypharmacology approaches”;
- “XII Giornata della Chimica dell’Emilia-Romagna” organized by the Italian Chemical Society, 17th December 2012, Ferrara. Oral (flash) communication: “Targeting the Hsp90 interactome using in silico polypharmacology approaches”, awarded as best communication.
- "XI Giornata della Chimica dell’Emilia-Romagna" organized by the Italian Chemical Society, Modena 28th October 2011. Poster: "Validazione di un metodo computazionale per lo screening di ligandi di recettori accoppiati a proteina G" ("Validation of a computational method for the screening of G-protein coupled receptors ligands").
- "1st Meeting of the Paul Ehrlich MedChem Euro-PhD Network", 7th Meeting of the European Network of Doctoral Studies in Pharmaceutical Sciences. 13th-15th July, 2011, hosted by three Universities of Madrid, Spain (Universidad de Alcalà, Universidad Complutense de Madrid and Universidad San Pablo CEU). Oral communication: "Validation of the performance of a post-docking method on multiple target structures"

7. Acknowledgments

I would like to express my very great appreciation to all the people that helped me to accomplish the work here reported.

Prof. Giulio Rastelli, my supervisor, for his guidance and encouragement throughout my PhD studies.

Prof. Dr. Jürgen Bajorath for hosting me as a visiting PhD student at the B-IT and for supervising the ligand-based part of the project.

Dr. Dagmar Stumpfe and Kathrin Heikamp for actively collaborating to perform activity annotation analysis and ligand-based calculations.

Dr. Leonard Neckers and his collaborators for their valuable work in conducting the biological assays on Hsp90.

Dr. Massimo Broggini for his valuable work in conducting biological assays on CDK2.

Dr. Fabiana Caporuscio for her useful advices and assistance.

Finally, I wish to thank my parents and my sister for their support and encouragement throughout my studies.

8. References

[1] J. Arrowsmith, *Nature Reviews Drug Discovery* **2012**, *11*, 17–18.

[2] G. Rastelli, *Pharm. Res.* **2013**, *30*, 1458–1463.

[3] A. R. Leach, B. K. Shoichet, C. E. Peishoff, *J. Med. Chem.* **2006**, *49*, 5851–5855.

[4] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat Rev Drug Discov* **2004**, *3*, 935–949.

[5] M. Sgobba, F. Caporuscio, A. Anighoro, C. Portioli, G. Rastelli, *European Journal of Medicinal Chemistry* **2012**, *58*, 431–440.

[6] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J Comput Chem* **2009**, *30*, 2785–2791.

[7] 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, et al., *J. Med. Chem.* **2004**, *47*, 1739–1749.

[8] T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, J. L. Banks, *J. Med. Chem.* **2004**, *47*, 1750–1759.

[9] G. Rastelli, G. Degliesposti, A. Del Rio, M. Sgobba, *Chem Biol Drug Des* **2009**, *73*, 283–286.

[10] A. Anighoro, G. Rastelli, *Computational Molecular Bioscience* **2013**, *03*, 27–31.

[11] P. Ripphausen, B. Nisius, J. Bajorath, *Drug Discov. Today* **2011**, *16*, 372–376.

[12] H. Geppert, M. Vogt, J. Bajorath, *J Chem Inf Model* **2010**, *50*, 205–216.

[13] G. Maggiora, M. Vogt, D. Stumpfe, J. Bajorath, *J. Med. Chem.* **2013**, DOI 10.1021/jm401411z.

[14] MACCS Structural Keys; Symyx Software: San Ramon, CA, **2009**.

[15] D. Rogers, M. Hahn, *J Chem Inf Model* **2010**, *50*, 742–754.

[16] *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, **2013**.

[17] A. L. Hopkins, *Nat. Chem. Biol.* **2008**, *4*, 682–690.

[18] A. D. W. Boran, R. Iyengar, *Curr. Opin. Drug Discov. Devel.* **2010**, *13*, 297–309.

[19] J. Mestres, E. Gregori-Puigjané, *Trends Pharmacol. Sci.* **2009**, *30*, 470–474.

[20] X. Jalencas, J. Mestres, *Med. Chem. Commun.* **2012**, *4*, 80–87.

[21] J.-J. Lu, W. Pan, Y.-J. Hu, Y.-T. Wang, *PLoS ONE* **2012**, *7*, e40262.

[22] M. Rask-Andersen, M. S. Almén, H. B. Schiöth, *Nat. Rev. Drug Discov.* **2011**, *10*, 579–590.

[23] Y. Hu, J. Bajorath, *J Chem Inf Model* **2010**, *50*, 2112–2118.

[24] P. C. Echeverría, A. Bernthaler, P. Dupuis, B. Mayer, D. Picard, *PLoS ONE* **2011**, *6*, e26044.

[25] L. Neckers, P. Workman, *Clin. Cancer Res.* **2012**, *18*, 64–76.

[26] X. Lu, L. Xiao, L. Wang, D. M. Ruden, *Biochem. Pharmacol.* **2012**, *83*, 995–1004.

[27] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, et al., *Nucleic Acids Res.* **2012**, *40*, D1100–1107.

[28] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, M. K. Gilson, *Nucleic Acids Res* **2007**, *35*, D198–D201.

[29] R. Lappano, M. Maggiolini, *Nat Rev Drug Discov* **2011**, *10*, 47–60.

[30] M. C. Lagerström, H. B. Schiöth, *Nat Rev Drug Discov* **2008**, *7*, 339–357.

[31] R. C. Stevens, V. Cherezov, V. Katritch, R. Abagyan, P. Kuhn, H. Rosen, K. Wüthrich, *Nat Rev Drug Discov* **2013**, *12*, 25–34.

[32] B. K. Shoichet, B. K. Kobilka, *Trends Pharmacol. Sci.* **2012**, *33*, 268–272.

[33] F. M. McRobb, B. Capuano, I. T. Crosby, D. K. Chalmers, E. Yuriev, *J Chem Inf Model* **2010**, *50*, 626–637.

[34] Y. Okuno, A. Tamon, H. Yabuuchi, S. Niijima, Y. Minowa, K. Tonomura, R. Kunimoto, C. Feng, *Nucleic Acids Res.* **2008**, *36*, D907–912.

[35] G. Rastelli, A. Del Rio, G. Degliesposti, M. Sgobba, *J Comput Chem* **2010**, *31*, 797–810.

[36] M. D. Parenti, G. Rastelli, *Biotechnol. Adv.* **2012**, *30*, 244–250.

[37] J. Zhang, P. L. Yang, N. S. Gray, *Nat. Rev. Cancer* **2009**, *9*, 28–39.

[38] R. Eglen, T. Reisine, *Pharmacol. Ther.* **2011**, *130*, 144–156.

[39] L. Palmieri, G. Rastelli, *Drug Discov. Today* **2013**, *18*, 407–414.

[40] M. Rabiller, M. Getlik, S. Klüter, A. Richters, S. Tückmantel, J. R. Simard, D. Rauh, *Arch. Pharm. (Weinheim)* **2010**, *343*, 193–206.

[41] V. Lamba, I. Ghosh, *Curr. Pharm. Des.* **2012**, *18*, 2936–2945.

[42] Z. Fang, C. Grütter, D. Rauh, *ACS Chem. Biol.* **2013**, *8*, 58–70.

[43] S. Lapenna, A. Giordano, *Nat Rev Drug Discov* **2009**, *8*, 547–566.

[44] *Cancer Discovery* **2013**, *3*, 4–4.

[45] S. Ortega, M. Malumbres, M. Barbacid, *Biochim. Biophys. Acta* **2002**, *1602*, 73–87.

[46] A. Echalier, J. A. Endicott, M. E. M. Noble, *Biochim. Biophys. Acta* **2010**, *1804*, 511–519.

[47] S. Betzi, R. Alam, M. Martin, D. J. Lubbers, H. Han, S. R. Jakkaraj, G. I. Georg, E. Schönbrunn, *ACS Chem. Biol.* **2011**, *6*, 492–501.

[48] M. P. Martin, R. Alam, S. Betzi, D. J. Ingles, J.-Y. Zhu, E. Schönbrunn, *Chembiochem* **2012**, *13*, 2128–2136.

[49] C. Cortes, V. Vapnik, *Mach. Learn.* **1995**, *20*, 273–297.

[50] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, R. G. Coleman, *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.

[51] J. M. Jez, J. C.-H. Chen, G. Rastelli, R. M. Stroud, D. V. Santi, *Chem. Biol.* **2003**, *10*, 361–368.

[52] L. Wright, X. Barril, B. Dymock, L. Sheridan, A. Surgenor, M. Beswick, M. Drysdale, A. Collier, A. Massey, N. Davies, et al., *Chem. Biol.* **2004**, *11*, 775–785.

[53] P. A. Brough, W. Aherne, X. Barril, J. Borgognoni, K. Boxall, J. E. Cansfield, K.-M. J. Cheung, I. Collins, N. G. M. Davies, M. J. Drysdale, et al., *J. Med. Chem.* **2008**, *51*, 196–218.

[54] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman, AMBER 10, University of California, San Francisco, **2008**.

[55] K. Aertgeerts, R. Skene, J. Yano, B.-C. Sang, H. Zou, G. Snell, A. Jennings, K. Iwamoto, N. Habuka, A. Hirokawa, et al., *J. Biol. Chem.* **2011**, *286*, 18756–18765.

[56] S. Wenglowsky, K. A. Ahrendt, A. J. Buckmelter, B. Feng, S. L. Gloor, S. Gradl, J. Grina, J. D. Hansen, E. R. Laird, P. Lunghofer, et al., *Bioorg.Med.Chem.Lett.* **2011**, *21*, 5533–5537.

[57] J. D. Hansen, J. Grina, B. Newhouse, M. Welch, G. Topalov, N. Littman, M. Callejo, S. Gloor, M. Martinson, E. Laird, et al., *Bioorg.Med.Chem.Lett.* **2008**, *18*, 4692–4695.

[58] L. F. Hennequin, J. Allen, J. Breed, J. Curwen, M. Fennell, T. P. Green, C. Lambert-van der Brempt, R. Morgentin, R. A. Norman, A. Olivier, et al., *J.Med.Chem.* **2006**, *49*, 6465–6488.

[59] R. B. Bledsoe, V. G. Montana, T. B. Stanley, C. J. Delves, C. J. Apolito, D. D. Mckee, T. G. Consler, D. J. Parks, E. L. Stewart, T. M. Willson, et al., *Cell(Cambridge,Mass.)* **2002**, *110*, 93–105.

[60] V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S.

Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, et al., *Science* **2007**, *318*, 1258–1265.

[61] V. P. Jaakola, M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. Chien, J. R. Lane, A. P. Ijzerman, R. C. Stevens, *Science* **2008**, *322*, 1211–1217.

[62] W. Liu, E. Chun, A. A. Thompson, P. Chubukov, F. Xu, V. Katritch, G. W. Han, C. B. Roth, L. H. Heitman, A. P. IJzerman, et al., *Science* **2012**, *337*, 232–236.

[63] E. Y. Chien, W. Liu, Q. Zhao, V. Katritch, G. W. Han, M. A. Hanson, L. Shi, A. H. Newman, J. A. Javitch, V. Cherezov, et al., *Science* **2010**, *330*, 1091–1095.

[64] T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G. W. Han, et al., *Nature* **2011**, *475*, 65–70.

[65] V. P. Jaakola, M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. Chien, J. R. Lane, A. P. Ijzerman, R. C. Stevens, *Science* **2008**, *322*, 1211–1217.

[66] G. Lebon, T. Warne, P. C. Edwards, K. Bennett, C. J. Langmead, A. G. W. Leslie, C. G. Tate, *Nature* **2011**, *474*, 521–null.

[67] U. Pieper, B. M. Webb, D. T. Barkan, D. Schneidman-Duhovny, A. Schlessinger, H. Braberg, Z. Yang, E. C. Meng, E. F. Pettersen, C. C. Huang, et al., *Nucleic Acids Res.* **2011**, *39*, D465–474.

[68] A. Fiser, R. K. Do, A. Sali, *Protein Sci.* **2000**, *9*, 1753–1773.

[69] A. Fiser, A. Sali, *Bioinformatics* **2003**, *19*, 2500–2501.

[70] J. C. Shelley, A. Cholleti, L. L. Frye, J. R. Greenwood, M. R. Timlin, M. Uchimaya, *J Comput Aided Mol Des* **2007**, *21*, 681–691.

[71] J. R. Greenwood, D. Calkins, A. P. Sullivan, J. C. Shelley, *J Comput Aided Mol Des* **2010**, *24*, 591–604.

[72] S. Betzi, R. Alam, M. Martin, D. J. Lubbers, H. Han, S. R. Jakkaraj, G. I. Georg, E. Schonbrunn, *Acs Chem.Biol.* **2010**, *6*, 492–501.

[73] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J Comput Chem* **2004**, *25*, 1605–1612.

[74] Suite 2011: Maestro, version 9.2, Schrödinger, LLC, New York, NY, **2011**.

[75] RDKit: Cheminformatics and Machine Learning Software. http://www.rdkit.org/

[76] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J Cheminform* **2011**, *3*, 33.

[77] Suite 2011: QikProp, version 3.4, Schrödinger, LLC, New York, NY, **2011**.

[78] M. P. Martin, R. Alam, S. Betzi, D. J. Ingles, J.-Y. Zhu, E. Schönbrunn, *Chembiochem*

2012, *13*, 2128–2136.

[79] C. Luo, P. Xie, R. Marmorstein, *J. Med. Chem.* **2008**, *51*, 6121–6127.

[80] E. V. Gurevich, Y. Bordelon, R. M. Shapiro, S. E. Arnold, R. E. Gur, J. N. Joyce, *Arch. Gen. Psychiatry* **1997**, *54*, 225–232.

[81] E. S. Burstein, J. Ma, S. Wong, Y. Gao, E. Pham, A. E. Knapp, N. R. Nash, R. Olsson, R. E. Davis, U. Hacksell, et al., *J. Pharmacol. Exp. Ther.* **2005**, *315*, 1278–1287.

[82] S. M. Stahl, *CNS Spectr* **2008**, *13*, 1027–1038.

[83] M. P. Gleeson, A. Hersey, D. Montanari, J. Overington, *Nat Rev Drug Discov* **2011**, *10*, 197–208.