

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

A Pipeline to Enhance Ligand Virtual Screening: Integrating Molecular Dynamics and Fingerprints for Ligand and Proteins

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1657385> since 2018-01-14T00:56:59Z

Published version:

DOI:10.1021/acs.jcim.5b00169

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

This is the author's final version of the contribution published as:

Francesca Spyraakis, Paolo Benedetti, Sergio Decherchi, Walter Rocchia, Andrea Cavalli, Stefano Alcaro, Francesco Ortuso, Massimo Baroni, Gabriele Cruciani. A Pipeline to Enhance Ligand Virtual Screening: Integrating Molecular Dynamics and FLAP. JOURNAL OF CHEMICAL INFORMATION AND MODELLING. 55(10) pp: 2256-2274.

DOI: 10.1021/acs.jcim.5b00169

The publisher's version is available at:

<http://pubs.acs.org/doi/10.1021/acs.jcim.5b00169>

When citing, please refer to the published version.

A Pipeline to Enhance Ligand Virtual Screening: Integrating Molecular Dynamics and FLAP

Francesca Spyrakis ^{1*}, Paolo Benedetti ², Sergio Decherchi ^{3,4}, Walter Rocchia ³,
Andrea Cavalli ^{5,6}, Stefano Alcaro ⁷, Francesco Ortuso ⁷, Massimo Baroni ⁸,
Gabriele Cruciani ^{2*}.

¹ Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 103, 41125, Modena, Italy; ² Department of Chemistry, Biology and Biotechnology, University of Perugia, Via Elce di Sotto 8, 06123, Perugia, Italy; ³ CONCEPT Lab, Istituto Italiano di Tecnologia, via Morego 30, 16163, Genova, Italy; ⁴ BiKi Technologies s.r.l., via XX Settembre 33, 16121, Genova, Italy; ⁵ CompuNet, Istituto Italiano di Tecnologia, via Morego 30, 16163, Genova, Italy; ⁶ Department of Pharmacy and Biotechnology, University of Bologna, via Belmeloro 6, 40126, Bologna, Italy; ⁷ Department of Health Science, University Magna Graecia, Campus "S. Venuta", Viale Europa, Germaneto, 88100, Catanzaro, Italy; ⁸ Molecular Discovery Limited, 215 Marsh Road, Pinner Middlesex-London HA5-5NE, United Kingdom.

*** Corresponding authors**

Gabriele Cruciani

Department of Chemistry, Biology and Biotechnology, University of Perugia, Via Elce di Sotto 8, 06123, Perugia, Italy.

phone: 0039 075 5855629

e-mail: gabri@chemiome.chm.unipg.it

Francesca Spyrakis

Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 103, 41125, Modena, Italy.

phone: 0039 059 2058604

e-mail: francesca.spyrakis@unimore.it

ABSTRACT

The importance of taking into account protein flexibility in drug design and virtual ligand screening (VS) has been widely debated in the literature, and molecular dynamics (MD) has been recognized as one of the most powerful tools for investigating intrinsic protein dynamics. Nevertheless, deciphering the amount of information hidden in MD simulations and recognizing a significant minimal set of states to be used in virtual screening experiments can be quite complicated.

Here we present an integrated MD-FLAP (Molecular Dynamics - Fingerprint for Ligand and Proteins) approach, comprising a pipeline of Molecular Dynamics, Clustering and Linear Discriminant Analysis, for enhancing accuracy and efficacy in VS campaigns. We first extracted a limited number of representative structures from tens of ns of MD trajectories by means of the k-medoids clustering algorithm as implemented in the BiKi Life Science Suite.¹ Then, instead of applying arbitrary selection criteria, i.e. RMSD, pharmacophore properties, enrichment performances, we let the Linear Discriminant Analysis algorithm, implemented in FLAP,² to automatically choose the best performing conformational states among medoids and X-ray structures. Retrospective virtual screenings confirmed that ensemble receptor protocols outperform single rigid receptor approaches, proved that computationally generated conformations comprise the same quantity/quality of information included in X-ray structures, and pointed to the MD-FLAP approach as a valuable tool for improving VS performances.

INTRODUCTION

The concept of flexibility inherently permeates proteins intrinsic nature, and knowledge of the intrinsic dynamics of macromolecules has direct implications in understanding the relationship between structure and function.^{3,4} Against this scenario, the role of macromolecule flexibility in drug design and discovery has been widely debated in the literature,³⁻⁸ and several studies have highlighted the need to properly account for protein dynamics to improve computational drug design predictions.⁹⁻¹⁴ Actually, numerous evidences have revealed the limitations of the old lock-and-key theory and more sophisticated models have therefore been elaborated. In particular the plasticity of biomolecules has been accounted for by the induced fit hypothesis (based on the induced structural changes occurring upon ligand binding) or by the conformational selection hypothesis (based on the presence of a pre-existing equilibrium ensemble of structures).^{5,6,15,16}

The use of single static structures (typically provided by X-ray crystallography or NMR) clearly provides a limited representation of proteins nature and behavior, and is one of the major limitations in virtual screening experiments, where the recognition of active compounds is greatly affected by the receptor conformation.⁷

Different strategies and recipes have been proposed to increase the success rate of VS campaigns.¹⁷⁻²⁷ For instance, Multiple Receptor Conformations (MRC) approach demonstrated to give better results in cross docking and VS, even when a limited number of structures is selected.^{17-19,26-28} Taken the MRC beneficial for granted, the difficulty relies more in the definition of the optimal set of structures to be chosen among a number of conformations, in particular when running extensive MD simulations, which typically provide thousands of snapshots for a single receptor.²⁹ Regardless of the methodology used to run the simulations and of the virtual screening approach adopted, the key point for successful VS campaign is represented by the sampling accuracy and by the capability of selecting the most representative conformations.^{30,31}

Ensembles of experimental receptor structures were first used by Knegtel *et al.* in 1997, who obtained good performances for ras p21 and in particular for HIV-1 protease.²² Crystal structures provide a valuable starting point for drug design endeavours,²⁷ however they rarely account for intrinsic protein flexibility, which can be of paramount importance to properly identify novel scaffolds and new lead candidates. It has been also underlined that in a real and not-retrospective screening study, there is no *a priori* way to know which the most representative conformations are.²²

Rueda *et al.* devised a recipe for conformation selection by analysing a benchmark of 1086 X-ray structures for 99 different therapeutically relevant protein targets, and comparing the performances of the single and ensemble-conformation docking.²⁷ The authors proved that, when possible, a ligand-guided approach represents the best strategy to identify the conformations better able to discriminate between binders and non-binders. The enrichment factors or the AUC can be used to rank *a posteriori* the different conformations to be used in independent dockings.³² They also found that adopting conformational ensembles does not systematically outperform the use of a single well-performing structure and that it is generally not possible to surpass the best single conformation, which is however unknown *a priori*.^{17,33} Significant improvements with MRC can be reached when poor structures are available, such as in case of homology-built models. When no ligand information is available, an ensemble of randomly chosen 3-5 conformations can provide better results improving

ligand recognition with respect to a single static state, as also suggested by Verdonk *et al.*³⁴

Bottegoni *et al.* carried out a retrospective screening study against 36 X-ray targets comparing the Single Receptor Conformation (SRC) and the MRC VS, focusing on i) the different separation power, ii) the number of actives identified and the chemical diversity of the hits, iii) the contribution of the single receptor conformation to the MRC-VS overall performance.¹⁸ Their analyses proved that both the MRC-score and MRC-rank methodologies outperformed the SRC and that using multiple receptor conformations with respect to one static structure significantly increases the variety of the chemical scaffolds. Better performances could be likely achieved by introducing more different and ligand-unbiased conformations coming, for instance, from MD simulations.

Again, the authors underlined that the conformational selection in absence of any ligand activity information still represents a challenging and major issue.

Similar conclusions were drawn by Korb *et al.*, who investigated the performances of ensemble docking in virtual screening and pose prediction, as a function of ensemble size, for eight X-ray solved targets using GOLD and three different scoring functions.³⁵ The authors observed that MRC docking always outperformed the use of the worst single structure, giving, in some cases, even better results with respect to the best single protein structure. Better performances were also obtained in terms of chemotype diversity. A strong dependence of docking results from scoring functions and sampling accuracy was highlighted, since different performances were observed for the same set of structures when using different scoring functions. The pressing need of developing ensemble selection protocols for choosing the best conformations given a target and a scoring function was again finally underlined.

In this work we propose an integrated Molecular Dynamics - FLAP approach for the automatic identification of the most relevant conformations to be used in VS experiments, when ligands activity information is available. Molecular dynamics simulations were run on the Purine Nucleoside Phosphorylase (PNP) enzyme, on the tyrosine-protein kinase ABL1, and on the transmembrane A_{2A} in an explicit membrane model. A clustering algorithm was used to extract from the trajectories a limited number of proteins conformations. The most relevant and useful conformations, in the clustered structures, for the identification of active ligands among thousands of decoys were finally chosen by a Linear Discriminant Analysis (LDA) implemented in the FLAP software.^{2,36} Different benchmarking for comparing the performances of FLAP with respect to other VS tools have been already performed, always reporting promising results,^{36,37} and have been supported again here (see Supporting Information, Figure S1 and Table S1).

We demonstrated that, when activity data is available, taking into account protein flexibility in VS simulations can successfully remove the inherent structural bias of static structures co-crystallized with specific ligands, allowing the identification of different and new chemical structures. We proved how computationally generated conformations can behave as well as experimentally solved ones and, finally, recognized the MD-FLAP approach a valuable tool for the dynamic investigation of highly flexible systems as kinases, and for the automatic selection of the most informative conformational states in virtual screening.

RESULTS AND DISCUSSION

Purine Nucleoside Phosphorylase

The PNP protein³⁸⁻⁴⁰ was used as first case study to extensively investigate the performances of the proposed integrated MD-FLAP approach.

PNP catalyzes the phosphorolysis of 6-oxypurine nucleosides and deoxynucleosides to generate α -D-(deoxy) ribose 1-phosphate and the purine base, which is then recycled or oxidized to uric acid.³⁹ A deficient PNP activity has been associated to an elevated blood concentration of D-guanosine, and to the consequent apoptosis of dividing T-cells.^{41,42} Different PNP inhibitors have been developed and used in the treatment of T-cell cancers and autoimmune disorders, including psoriasis, rheumatoid arthritis, multiple sclerosis and some of them are in clinical trials for gout.⁴³ Among others, 7-deazaguanine (PDB code 3iny), 2-fluoroadenine (PDB code 3gb9), and transition state analogues (TSA) such as Immucillin-H (TSA for the bovine PNP), DADMe-ImmH, DATMe-ImmH and SerME-ImmH (TSAs for the human PNP) proved their capability to strongly inhibit the target, showing K_i in the picomolar range.⁴⁴⁻⁴⁶

PNP is a homotrimer bearing a HPO_4^{2-} ion in each binding site.⁴⁷ The ternary complex formed by PNP, DATMe-ImmH (Scheme 1, compound 1) and HPO_4^{2-} is reported in Figure 1a. As shown in Figure 1b, the ligand makes several hydrogen bonds with the residues lining the binding site, i.e. Glu201, Asn243, and His257, and the HPO_4^{2-} ion. These interactions stabilize the protein-ligand complex and, together with their TSA nature, justify the observed high affinity.

Such strong inhibitors do not represent the typical initial scenario of a virtual screening campaign, where one generally starts with micromolar affinity scaffolds and target a low micromolar or nanomolar activity level. Reasonably, higher the affinity of the ligand towards the target, higher is the geometric/electrostatic complementarity of the binding site for the cognate ligand and stronger is the structural bias induced on the co-crystal. We will later discuss this possible bias towards analogous scaffolds present in a database, demonstrating that upon its removal, i.e. by working with a reduced dataset, our approach still grants significantly improved results over plain VS.

We first performed standard VS experiments and then proposed a new approach combining the contribution of Molecular Dynamics, clustering and Linear Discriminant Analysis. VS experiments were carried out screening the DUD-E dataset, containing 225 known active ligands and 6950 decoys

The first experiment was run using a Single Receptor Conformation approach on the 3k8o crystal structural (hereafter named SRC-3k8o VS). This represented the baseline for all further results. SRC-3k8o gave a good enrichment with an AUC value of 0.85 and a good performance also in the early stage of the screening. As reported in Table 1 and shown in Figure 2a, at the 1% of false positives, the 40% of true positives, i.e. actives, were already found (1% ROC enrichment equal to 0.40). As well, almost 60% of positive (5% ROC enrichment equal to 0.58) were identified at the 5% of screened false positives. The good early performance of the SRC-3k8o VS is also shown in Figure 2a', reporting the percentage of true positives found in the first screened 500 molecules. For instance, the 40% of true positives was already found in the first 100 screened molecules, showing that if the first 100 ranked molecules were purchased, then 40 of them would have been active against the target. The quality of the VS performances is also supported by the predicted ligands poses, which generally

resembles the one of the co-crystallized DATMe-ImmH reported in Figure 1. As an example, the FLAP-predicted orientation of two active molecules into 3k8o binding site is reported in Figure 3. The bi-dentate interaction with Asn243 is well conserved in both molecules, as well as, when possible, the interactions with the phosphate, Glu201 and His257.

Starting from such a good performance we focused on the following issues:

- i) accounting for protein flexibility via MD simulations to improve the screening performance in term of quantity (AUC and ROC enrichments) and quality (chemical diversity);
- ii) using and testing different X-ray structures to provide better predictions;
- iii) exploiting information coming from MD and choosing the best performing templates.

The aim of the following experiments was to build a predictive LDA model using the aforementioned 3k8o X-ray structure and a number of conformations extracted from a MD trajectory run on the same protein.

We run 38 ns MD for the PNP trimer (the whole backbone RMSD is reported in Figure S2). Upon completion of the simulation, we checked the conformational space explored by the loop facing the binding site, namely residues Val246-Lys266. Figure 4 reports the RMSD over time of the backbone atoms of the aforementioned loop for each PNP monomer, with respect to the first frame, i.e. the structure after the last NPT equilibration step. The graph shows that, consistently, the monomers explored an analogous space around the value of 1.5 Å, allowing a significant sampling of loop conformations.

Then, a total of 10 clusters were computed and the medoids (namely the clusters centres; Figure S3) were used as reference structures for the virtual screening. Both the MD simulations box setup and the clustering³⁸ were performed using the BiKi LifeScience environment.¹

To assess the diversity of the space spanned by the clustering results, we computed the RMSD of the different monomer loops for each cluster, with respect to the apo loop conformation (PDB code 1m73).⁴⁸ Figure 5 shows that each monomer had a different conformation with respect to the apo loop, thus confirming the enrichment in terms of sampled structures. In particular, monomer 1 (red line) shows the minimal distortion with respect to the apo structure. For this reason, in order to explore the most diverse space with respect to the 3k8o crystal and to reduce the overall computational burden, we decided to perform all the subsequent VS experiments on monomer 1. Nevertheless, the full trimeric unit was always retained since the binding site is located at the dimerization interface. Figure 6 reports the super-imposed structures of the loops present in the clusters, including the apo form of the loop (blue coloured).

In a 'standard' VS approach, all the frames would have been tested in order to identify the ones able to give the best performances in terms of enrichment.²⁹ Here, instead, we first pre-filtered the trajectory by clustering, and then asked the LDA-based protocol implemented in FLAP to automatically choose the combination of three different conformations among the medoids and the 3k8o X-ray structure, and of three FLAP scores, better able to discriminate between active compounds and decoys. The selected structures were the X-ray 3k8o and the medoids 7 and 9, while the FLAP scores were N1*O, DRY*O and H*O*N1 (see Material and Methods for a more detailed description of FLAP scores).

The ROC curve given by the LDA-MRC-3k8o VS is shown in Figure 2a (dark grey line) and the corresponding AUC value (0.94) in Table 1. The improvement given by the MRC approach is quantified by the gain in the AUC value, increasing from 0.85 in the SRC-3k8o VS to 0.94 in the LDA-MRC-3k8o VS. Nevertheless the early recognition did not improve appreciably as the ROC enrichments at 0.5%, 1% and 2% only showed a small gain. At 5% the enhancement became more evident, i.e. 0.75 with respect to 0.59. To verify that the AUC and partial ROC enrichments were not only given by a subset of the FLAP probes but by the multiple conformations allowed during the screening procedure, we build an LDA-SRC-3k8o model using only the 3k8o structure as template but three different scores. Lower enrichment values were obtained. The global AUC was equal to 0.88 and the ROC enrichment at 0.1% and 5%, 0.21 and 0.52, respectively. Data are reported in Table 1. The graph showing the comparison between the 3 templates/3 scores (LDA-MRC-3k8o) and the 1 template/3 scores approach is reported in Figure S4.

The most significant improvements were actually recognized in terms of results quality and of chemical diversity of the ranked molecules, rather than in terms of quantity, i.e. AUC and ROC enrichments. In Table 2 the rank position obtained in the SRC-3k8o and in the LDA-MRC-3k8o VSs by some of the most different actives with respect to the co-crystallized DATMe-ImmH are reported. Main changes in the chemical structure are related to the nature and to the position of the substituents, as the base scaffold interacting with Asn243 is, generally, well conserved in all the active molecules. The five reported compounds were all poorly ranked in the SRC-3k8o VS, while they occupied relevant ranking position in the LDA-MRC-3k8o VS, e.g. the poorest ranked was located at position 241. To further prove the capability of the LDA-based MRC approach to improve chemical diversity with respect to standard SRC approaches, active molecules and DATMe-ImmH were clusterized in classes. A fingerprint- and substructure-based clusterization was performed according to the algorithm developed by Stahl and Mauser, using parameters described in the original publication.⁴⁹ Fourteen different classes were obtained and the average rank of each class components calculated for both the SRC-3k8o and the LDA-MRC-3k8o screenings. Results are reported in Figure 7 and show how, in general, the average rank of the different classes is lower when the LDA-based MRC approach is adopted. Only two classes were better ranked by the SRC approach, with one of them, i.e. class 2, including the co-crystallized DATMe-ImmH. The higher probability of identifying chemically different molecules with the LDA-based MRC approach was then quantified by calculating the mean and the variance of the rankings across the identified ligand classes. Values of 1482 ± 1410 and 501 ± 463 were obtained, respectively, for the SRC-3k8o and the LDA-MRC-3k8o screenings, showing that on average the ranking is better and that, more importantly, the variability of the ranking is much reduced across the ligand classes for the MRC approach. This again supported the capability of the proposed method of enlarging the chemical space of the first ranked compounds.

Choosing the right template

As previously discussed, the availability of such a stable and good performing starting structure does not represent the standard situation in VS campaigns. More commonly, VS experiments are run using co-crystallized structures with high μM range affinity ligands, or apo forms of the target protein. To assess how much VS performances can be affected by the initial structure, we performed VS experiments in SRC and MRC

mode using an apo form and three different structures of PNP in complex with different ligands.

As reported in the Methods section, the apo structure of the trimer (not present in the PDB repository) was obtained by running steered molecular dynamics on each of the monomers of 3k8o to reach the apo form of the PNP monomer (PDB code 1m73). The co-crystallized ligands in the selected holo forms were: 8-azaguanine,⁵⁰ acyclovir⁵¹ and 2-mercapto-(3H)-quinazoline⁵² (PDB codes 1v41, 1pwy and 3d1v; Scheme 1, compounds 2, 3, 4, respectively), showing K_d values equal to 20, 90, 324 μ M, respectively. These structures were chosen because, first, the respective ligands are in the micro-molar range, thus resembling a VS-like scenario, secondly they exhibit a relative diversity of the co-crystallized ligands. The performances of the SRC VS on the apo, the 1v41, 1pwy and 3d1v complexes are reported in Table 1 and in Figures 2b-e (light grey lines). With the only exception of 1v41, all structures performed better than the original 3k8o in terms of global AUC, giving values equal or close to 0.90. Nevertheless, only 3d1v gave results comparable to 3k8o in terms of ROC enrichments at the 0.5% and 1% of screened false positives. While the 3k8o and 3d1v structures allowed the identification of more than 30% of positives at the 0.1% of screened false positives, no more than 18% of positives was, in fact, detected when using the other structures. Better performances were given at 5% ROC enrichment by 1pwy, detecting 57% of positives with respect to the 59% and 61% identified by 3k8o and 3d1v, while the apo form and the 1v41 complex only reached the 43% and 40%, respectively. Interestingly, the best performing structure, 3d1v, would not have been chosen *a priori* as starting structure for any SBVS because of a number of inconsistencies in the same structure, as missing atoms and missing bonds.

We then tried to investigate if the MD trajectory encoded information could help in choosing the most proper structure or could, at least, remove the bias given by the use of one structure with respect to another.

Similarly to the previously reported experiment (MRC on medoids plus 3k8o X-ray), we performed MRC VS experiments, using as possible structure templates the ten medoids and each different X-ray structure at a time. Again, we asked the LDA to automatically choose the best combination of the three different conformations and the three FLAP scores better able to discriminate between actives and decoys. Finally, we removed the contribution of any X-ray structure and run the LDA model by using as possible templates the only ten medoids extracted from the MD/clustering procedure. The obtained models were LDA-MRC-apo, LDA-MRC-1v41, LDA-MRC-1pwy, LDA-MRC-3d1v, LDA-MRC-med. Results are reported in Table 1 and Figure 2 (dark grey lines).

The comparison of the AUC and of the ROC enrichments reported in Table 1 intriguingly shows that the same performances were obtained regardless the presence of the different X-ray structures in the panel of possible templates. Actually, the contribution of the aforementioned crystallographic structures was completely disregarded, since the same three medoids, i.e. 2, 4, 8, and the same three FLAP scores, i.e. H*DRY, H*N1*H, H*O*H, were chosen in all the cases. Thus, even in the absence of any X-ray structure (LDA-MRC-med model), the VS performances did not change.

The comparison of the SRC-VS and MRC-VS performances, reported in Figure 2b-e (light grey and dark grey lines, respectively) highlights how better results were always obtained when using multiple conformations with respect to a single structure. In particular, very good performances were reported for LDA-MRC-apo and LDA-

MRC-1v41, with respect to SRC-apo and SRC-1v41. Better results were also given by LDA-MRC-1pwy, especially in the early screening stage (see Figure 2b'-e'). In Figure 8, the percentage of active molecules found in the first 500 screened ones by the LDA-MRC-3k8o, LDA-MRC-med and the LDA-MRC-Xray VSs is reported. In the last model, the previously screened X-ray structures, i.e. 3k8o, apo form, 1v41, 1pwy, 3d1v, were used as possible templates. The combination represented by the three structures 3k8o, 1pwy and 3d1v and by the three FLAP score H*DRY, N1 and H*O*H was chosen as the best discriminating one. The three models performed very well and similarly in the very early stage of the screening up to the first 100 screened molecules. Then, slightly better results were provided by the LDA-MRC-3k8o, which still demonstrated to give the best performance (see Table 1).

Removing the ligand induced bias

The well-known induced fit theory is based on the belief that in a protein-ligand complex the final protein conformation is partially induced by the structure of the ligand. As the ligand enters the binding site, the protein adjusts itself in order to reach the highest possible complementarity and to generate the most stable complex. In contrast, theories based on conformational ensembles postulate the presence of a pre-existing equilibrium of structures ensemble, among which the ligand chooses the most suitable and complementary one.⁵ More recent theories describe ligand binding as a combination of a conformer selection-stage, followed by minor adjustments within the binding pocket modulated by the ligand (induced-fit stage).⁵³

We previously commented on the possible structural bias given by the DATMe-ImmH ligand on the 3k8o structure, considering the very high affinity and the high number of hydrogen bonds formed between the ligand and the residues lining the pocket. This high similarity could lead to an overestimation of the SRC-3k8o VS performances, since the dataset is populated by many active compounds structurally similar to the co-crystallized one.

Aimed to remove the ligand induced bias, to have a reliable estimation of the single structure screening quality, and to provide a proper comparison with the MRC-based approach, we reduced the original dataset by removing ligands similar to DATMe-ImmH. We thus performed a ligand-based virtual screening of the whole original DUD-E PNP dataset, using as template the DATMe-ImmH ligand. Molecules were ranked according to the FLAP distance to the template. The first ranked 1%, including both actives and decoys, was removed from the dataset. We then screened the reduced dataset against the single 3k8o structure and against the LDA model built using as template the same 3k8o structure and the ten medoids. The comparison of the new SRC-3k8o' and LDA-MRC-3k8o' models performance is reported in Figure 9, as well as the comparison of the results provided by the original SRC-3k8o and LDA-MRC-3k8o VS. Data are reported in Table 3.

The removal of the co-crystallized ligand did change the predictions. The so well performing SRC-3k8o model significantly worsened and the new SRC-3k8o' was able to give a global AUC of only 0.75 (light blue line, panel a), with respect to the original 0.85 (light grey line, panel a). The performances in the early screening stage also significantly got worse since the ROC enrichment at the 0.5% of false positive decreased from 0.33 to 0.16. Similarly, at 5%, the ROC enrichment reached the value of 0.26, with respect to the original 0.59. Interestingly, the new LDA-MRC-3k8o' model did not experience the same worsening and still provided appreciable predictions. As shown in Figure 9a, the global AUC moved from 0.94 (dark grey line,

panel a) to 0.92 (blue line, panel a). We only observed a slight drop in the ROC enrichment at 0.5%, being 0.18 with respect to 0.28, which was then regained at 5%, moving to 0.62 with respect to the original 0.75.

Figure 9b reports the percentage of active compounds found in the 500 ranked molecules by the VS run using the four aforementioned models, i.e. SRC-3k80' (light blue line), SRC-3k80 (light grey line), LDA-MRC-3k80' (blue line), LDA-MRC-3k80 (dark grey line). Here too, we could appreciate the poor performance provided by the SRC VS once the database has been deprived of the molecules most similar to the co-crystallized DATMe-ImmH. The percentage of actives decreased, in fact, from 58 to 32. Conversely, the MRC model experienced a much less pronounced decreasing, i.e. from 73 to 64, being able to sample different conformations and to recognize different scaffolds.

The screening data obtained for the other models are represented in Table 3 and in Figure 5 SI. With respect to the original dataset, all SRC models gave worse performances, even if to a less extent than SRC-3k80'. In the case of SRC-1v41', molecules were ranked according to the H*N1 FLAP score, giving definitely better AUC values with respect to the distance from the template. The most different results, in particular, were obtained for the SRC-apo' and the SRC-1pwy' models, both in terms of global AUC and in percentage of actives ranked in the early VS range. Again, we did not observe any big difference when the reduced dataset was screened towards the LDA generated models, thus supporting the importance of considering the intrinsic protein dynamics, and the quality of the computationally-generated conformations.

The percentage of active molecules found in the first 500 screened ones by the new LDA-MRC-3k80', LDA-MRC-med' and the LDA-MRC-Xray' VSs was again compared and is reported in Figure 10. The best results were provided by the LDA-MRC-med' VS (blue line), which performed better than the LDA-MRC-3k80' VS (where 3k80 X-ray was retained as LDA template; magenta line) in particular in the early screening stage. Interestingly, even better predictions were obtained with respect to the LDA-MRC-Xray' VS (red line), suggesting that, in this case, computationally generated conformations can contain more useful information than experimentally solved structures. As in the VS performed on the whole dataset, the LDA-MRC-3k80' outperformed the LDA-MRC-Xray' VS, but here the LDA-MRC-med' VS gave even better results (Figure 8 and Figure 10).

Looking at the phosphate as an active player

We then considered the phosphate as an active player in VS experiments and showed that by including in the LDA models medoids with and without the phosphate, results could be further increased.

A significant amount of active ligands are transition state analogous stabilized in the binding site by the formation of electrostatic interactions with the negatively charged HPO_4^{2-} ion. Nevertheless, also phosphonate derivatives are known PNP inhibitors.⁵⁴⁻⁵⁶ Their potency has been mainly attributed to the capacity of filling three discrete binding sites in the enzyme pocket, i.e. a guanine binding region, a hydrophobic pocket and a phosphate binding site.⁵⁴ We can reasonably hypothesize that the phosphate group occupies the region where the co-crystallized HPO_4^{2-} ion is normally located, as also supported by crystallographic evidences.^{57,58} This would imply the phosphate absence, or displacement, in the binding site, and a PNP not in the fully saturated regime.

We included in the configurational space the possibility of a missing phosphate. In particular, since the DUD-E dataset used in this work contains different phosphonate derivatives we decided to investigate how the VS performances could change with the phosphate being removed from the binding pocket. We thus built the LDA-MRC-med20 model, giving to FLAP as possible templates the medoids extracted by the MD/clustering approach with and without the HPO_4^{2-} ion, i.e. twenty medoids. Among the twenty structures the three most able to discriminate between active molecules and decoys were medoids 2 and 4 in presence of HPO_4^{2-} and medoid 3 in absence of HPO_4^{2-} . The three selected FLAP scores were H*DRY, H*O*N1 and H*O*DRY. The comparison of the ROC curves and of the percentage of active molecules found in the first 500 screened ones by the previous LDA-MRC-med model and by the last LDA-MRC-med20 is shown in Figure 11 panels a and a'. Data are reported in Table 4. Despite the global AUC only increased from 0.93 to 0.96, extremely relevant was the increment of the partial ROC enrichments. The percentage of actives at 0.5% of screened false positives moved from 32% to 46%, while at ROC 5% 80% of actives were identified, with respect to the previous 66%. Panel a' in Figure 11 clearly shows how the actives percentage raised in the early screening stage. Similar results were obtained when the model was used to screen the reduced dataset. While the global AUC moved from 0.92 to 0.95, at ROC 5% we were able to identify 80% of active compounds with respect to 62%.

Again the MD-clustering-LDA approach allowed us to choose the most informative structures, not only in terms of pocket flexibility but also of binding site architecture, and to accommodate within the binding site the most diverse ligand types with and without a phosphate group.

The most interesting aspect is that the best performing protein conformations among the chosen templates are both composed of PNP with and without the HPO_4^{2-} ion into the binding site, which may suggest that considering the cofactors as an integral part of the approach can further increase the explored configurational space and the VS performance.

Models robustness and validation

All the previously reported LDA models were calculated on the same datasets used for VS predictions. No overfitting was expected because of the low number of variables (3 templates and 3 scores), with respect to the number of samples, i.e. 7229 for the whole dataset and 7080 for the reduced one.

To prove this assumption, we randomly separated the compounds in a training and test set, equally dividing the original dataset. The LDA-MRC-PNP model was calculated on the training set, using again 3 templates and 3 scores, and validated on the test set. The result of the screening is reported in Figure 12 (panel a, purple line) and has been compared with the enrichment obtained screening only the test set towards the single receptor conformation, i.e 3k8o, (panel a, pink line). AUC and partial ROC enrichment values are reported in Table 5. Again, better results were obtained when using an LDA-MRC approach with respect to a SRC methodology. In particular, a significantly better ROC enrichment of 0.81 was obtained at 5%, with respect to an original value equal to 0.67. To prove that the enrichment gain was mainly given by the use of different protein conformations and not by the combination of multiple FLAP scores, we built an LDA-SRC-PNP model using the only 3k8o structure but three different scores (panel a, magenta line). The total AUC did not change with respect to the SRC VS, i.e. 0.81. Slight changes occurred in the

ROC curve but no significant gain could be observed, in particular, in the early enrichment phase.

Also, to design a more realistic scenario, in which only a few similar actives are known, we selected the 48 active compounds populating the same class of the original co-crystallized ligand and used them, plus 50% of decoys, to build a further LDA-MRC model. This model was validated on the corresponding test set, containing all the remaining actives and the remaining 50% of decoys, and used to screen the test set. In this case the LDA selected only medoid 7 as the template giving the best performance. The obtained AUC and ROC enrichment values were comparable to those given by the previous LDA-MRC-PNP model, in particular in the very early enrichment. The global AUC was 0.91 with respect to 0.92, while the ROC enrichments at 0.5%, 1%, 2% and 5%, respectively 0.24, 0.30, 0.35 and 0.60, with respect to 0.27, 0.38, 0.69, 0.81.

Exploring the versatility of the MD-FLAP approach

The integrated MD-FLAP approach was then applied to other two cases to better test its capability of improving VS performances on totally different targets. In particular, the tyrosine-protein kinase ABL1 and the adenosine A_{2A} receptor (here abbreviated as ABL1 and A2A, respectively) were chosen as representative test cases. The corresponding actives/decoys datasets were retrieved from the DUD-E database.

ABL1. The tyrosine-protein kinase ABL1 is involved in different cell differentiation processes as cell division and adhesion. Mutations in the ABL1 gene are mainly associated to chronic myeloid leukemia (CML), caused by the overactivation of the tyrosine kinase c-ABL due to the presence of the BCR-ABL fusion gene.⁵⁹

Kinases are well-known highly flexible targets, whose activation is mainly controlled by conformational changes in three conserved structural motifs at the active site: the activation loop (A-loop), the Asp-Phe-Gly (DFG) motif, and the α C-helix.⁶⁰ This structural flexibility has assumed particular importance in the development of specific kinase inhibitors able to target kinase-specific conformations. For ABL1, in particular, four different conformations were identified for the DFG motif: the DFG-in active conformation, the DFG-out inactive conformation, the DFG-flipped conformation and the Src-like inactive conformation.⁶¹ In the active state, the aspartate residue of the DFG motif points toward the active site (DFG-in), where it coordinates a catalytic magnesium ion. In the inactive conformation, the aspartate moves outside the active site while the phenylalanine moves inward. This DFG-out or DFG-flipped conformation is incompatible with Mg²⁺ binding and catalysis. The well-known imatinib, for instance, binds the DFG-out conformation, originally proposed to account for imatinib specificity towards ABL1.^{62,63}

The ABL1 structure deposited in the DUD-E database corresponds to the DFG-flipped conformation (PDB code 2hzi⁶¹). The DUD-E ligand dataset for ABL1 contained 295 actives and 10884 decoys. These were randomly separated in a training and a test set composed, respectively, by 5584 and 5595 compounds, with a similar actives/decoys ratio.

The SRC-ABL1 virtual screening was performed on the single reference X-ray structure. The corresponding ROC curve is shown in Figure 12 (panel b, pink line), while data are reported in Table 5. The global AUC reached a value of 0.69, with very poor partial enrichments in the early screening stage. The ROC enrichment at 0.5%, in fact, was equal to 0.03 (the lowest value ever found in all the VS here reported),

which means that only the 3% of actives was identified at 0.5% of screened false positives. No significant improvements were observed at 1% ROC enrichment; at 5% a poor value of 0.23 was obtained. With respect to the PNP case, the overall AUC, and in particular, the early recognition showed a significant worsening in performance. This could be reasonably due to the highest level of flexibility experienced by ABL1 in comparison to PNP and, likely, by the capability of the dataset ligands to bind significantly different target conformations.

To identify these possible conformations and apply the proposed MD-FLAP approach, we run a 100 ns scaled MD simulation on ABL1, similarly to what performed in a recent work.⁶⁴ In the referred paper we used an enhanced sampling technique to improve the sampling of the binding site together with the ligand, whereas here we accelerated the phase space sampled by the kinase activation loop. Ten representative conformations were then extracted by clustering the trajectory on the activation loop (see the Methods section for further details). Then we let the LDA-based protocol choose the combination of three different conformations among the medoids and the 2hzi X-ray structure, and of three FLAP scores, better able to discriminate between active compounds and decoys on the training set. The obtained model was validated on the test set, and used to screen the same test set to get the best performances in prediction. The selected structures were medoids 1, 4 and 6 (the X-ray structure was not chosen as template), while the FLAP scores were H, DRY and DRY*O.

The ROC curve given by the LDA-MRC-ABL1 VS is shown in Figure 12 (panel b, purple line) and the corresponding data are reported in Table 5. The improvement given by the MRC approach is quantified by the gain in the AUC value, increasing from 0.69 in the SRC-ABL1 VS to 0.85 in the LDA-MRC-ABL1 VS, but, in particular, by the ROC enrichments better accounting for the method improved performances in the early screening stage. Thus, the ROC enrichment at 0.5% moved from 0.03 to 0.22 (a seven fold performance increase) while at 5% a value of 0.55 was obtained with respect to 0.23 originally provided by the SRC-ABL1 VS. The LDA-SRC-ABL1 model (panel b, magenta line), built using as template the X-ray structure and three different FLAP scores, did not show any significant improvement with respect to the SRC-ABL1 model, in particular in the early screening stage. Slight improvements can be observed in the central and final region of the curve. These observations again supported the importance of including more conformations in VS campaigns when dealing with flexible targets.

A2A. The adenosine A_{2A} receptor belongs to the family of G-protein coupled receptors (GPCRs), known to respond to an impressive plethora of different stimuli. A2A, in particular, is ubiquitously expressed in humans and is recognized as a major mediator of anti-inflammatory responses. Also it has been identified as a promising pharmacological target for the development of agonists for myocardial perfusion imaging and as anti-inflammatory agents, and antagonists for the treatment of neurodegenerative disorders, such as Parkinson's disease.⁶⁵

As for all GPCRs, the computational simulation of the receptor dynamics in the cell membrane is far from trivial. The complexity of building a reliable model for *in silico* membrane protein simulations resides in the varying nature of the membrane composition, and in the difficulty of providing an accurate parametrization of the different membrane components in the force field.⁶⁶ In this perspective simplified models using single species lipid bilayers have been generally adopted.^{67,68} More recently bilayers containing two or more different lipids have been used.^{66,69} Here we built a POPC bilayer and run a 100 ns MD trajectory.

As better described in the Methods section, the A2A model used in MD simulations was built starting from PDB codes 3uzc⁷⁰ and 4eiy⁷¹ for modelling, respectively, the trans membrane region and the extracellular and sodium ion. 4eiy structure was used instead of the 3eml⁷² structure deposited in the DUD-E dataset, because of the presence of the same co-crystallized ligand and of a better resolution.

Ten representative conformations were extracted by clustering the trajectory on the Phe168 and Leu249 residues, having a specific role in defining the shape of the transmembrane receptor pocket.

The DUD-E dataset for A2A contained 842 actives and 32026 decoys, which were randomly separated in training and test set containing, respectively, 16422 and 16446 compounds. The training set was used to build the LDA-MRC-A2A model, using again a 3 templates/3 scores approach. The original X-ray structure, medoids 2 and 6 were chosen as templates while N1*O, DRY and H*O*DRY were identified as the most representative FLAP scores.

The comparison of the SRC-A2A and of the LDA-MRC-A2A VS is reported in Figure 12 (panel c, pink and purple lines, respectively) and the corresponding data in Table 5. The SRC-A2A VS gave a very good AUC value of 0.81 and quite good partial enrichments also in the early screening phase, being able of identifying 41% of actives at the 5% of screened false positives (5% ROC enrichment). Very slight improvements were observed in the global AUC when the VS was performed using the LDA-MRC-A2A, i.e. 0.86 with respect to the previous 0.81, and in the partial ROC enrichments, with 49% of actives found at 5% ROC enrichment, with respect to 41%. With respect to PNP and ABL1, the A2A receptor might, in fact, experience a lower conformational flexibility, if we consider the packing effect of the membrane bilayer and the almost total α -helix nature of the transmembrane receptor.

The screening performed with the LDA-SRC-A2A model showed no better performances with respect to the SRC-A2A one, confirming that the slight improvement observed in the MRC screening was given by the use of multiple receptor conformations. Also, we have to note that no water molecule was included in VS calculations. Nevertheless their importance in mediating receptor-agonist/antagonist interactions, the prediction of waters network behaviour in GPCRs still remains rather challenging and time-consuming in VS campaigns. Recent higher levels of theory predictions showed the importance of water molecules when dealing with triazine compounds within A2A.⁷³

Assessing MD enhanced Virtual Screening

In a recent work, Nichols and collaborators²⁹ assessed the performance of virtual screening in presence of configurations sampled from molecular dynamics trajectories. The authors used a full set of MD snapshots and assessed the change in performance in terms of AUC and ROC enrichments. They reported that MD snapshots on average helped improving VS results but still, most of the snapshots gave worse performances with respect to X-ray structures. In addition, clustering alone did not seem enough to improve VS performances.

The approach here proposed combines together an unsupervised and a supervised selection of the most representative protein conformations. We first found the significant free energy basins by clustering (unsupervised selection) and then, using the LDA methodology (supervised selection), we identified the protein conformations able to better separate actives from decoys. This double filtering allowed us to retain

from the dynamics only the most potentially useful and effective information for improving VS figures.

This strategy could be significantly useful for any VS campaign willing to include receptor flexibility in presence of ligand activities. Regardless the nature of the flexibility information, being experimentally or computationally determined, the LDA could be able to identify the most informative structures and to improve virtual screening performances by considering protein dynamics. It has been reported in literature that, sometimes, including flexibility in docking or VS analyses could add noise rather than real information,¹⁷ in addition to a higher cost in terms of calculation time. This could be due to different reasons: i. the protein target is not experiencing significant conformational adjustments; ii. the conformational space is not properly sampled; iii. too many conformations have been considered in the simulations.

In this perspective the integrated MD-FLAP approach could verify if the protein undergoes fruitful adjustments for VS (if different X-ray structure are available) or if the explored conformational space does not convey any useful information. If no enrichment improvement could be achieved by allowing different binding site conformations, a model considering only one template will give the same performances of a model built using more structures. No better performances will be obtained by increasing the number of templates upon a certain level, where only the calculation time would increase.

METHODS

In the following section we discuss the details of the overall Virtual Screening protocol. The methodology consists in enhancing the virtual screening performance by generating feasible configurations of the targets orthosteric binding site, via plain or scaled molecular dynamics. In particular, to collect the most representative structures, we clustered the molecular dynamics trajectories and selected the most probable conformations belonging to free energy basins. Along with these structures, for the PNP case, we collected additional crystallographic structures available in the PDB repository, to possibly enhance the screening performance and investigate the contribution given by either experimental or computationally generated structures. All the simulation setup was carried using the BiKi LifeScience software suite.¹

Molecular dynamics guided conformational sampling and clustering

PNP. The structure of PNP complexed with DATMe-Immucillin H (PDB code 3k8o³⁹) was retrieved from the Protein Data Bank (www.pdb.org); the ligands were removed and the phosphate ions located in the binding sites treated as mono-protonated.^{38,47}

The protein was parameterized by Amberr99SBildn force field⁷⁴ whereas the phosphate group was parameterized according to,³⁸ where an ab-initio RESP charges fitting was performed. Gromacs 4.6.1 was used to run MD simulations.⁷⁵ The water model employed was TIP3P and the final box size was 111 x 102 x 111 Å in the X, Y and Z directions, respectively. The solvated system was preliminary minimized by 5000 steps of steepest descent. As in the subsequent equilibration the integration step was equal to 2 fs. The Verlet cut-off scheme, the Bussi-Parrinello thermostat,⁷⁶ LINCS for the constraints (all bonds), and the Particle Mesh Ewald for electrostatics, with a short range cut-off of 11 Å, were applied.

The system was equilibrated in four subsequent steps: 100 ps in NVT ensemble at 100 K, 100 ps in NVT ensemble at 200 K, 100 ps in NVT ensemble at 300 K and, finally, a 1 ns long NPT simulation to reach the pressure equilibrium condition. In the first two equilibration steps harmonic positional restraints were set on the backbone of the protein with a spring constant of 1000 kJ/mol/Å². The production run was carried on in the NVT ensemble at 300 K, without any restraint for approximately 40 ns.

Upon completion of the simulation we run a clustering process on the molecular dynamics trajectory. We employed the algorithm proposed in³⁸ and used the backbone RMSD with optimal alignment (Kabsch algorithm)⁷⁷ between frames as metric for distance computation. The clustering was performed on the whole protein backbone. In order to carry out the experiments in a reasonable amount of time, combining speed and accuracy, the following VS analyses were performed only on monomer 1, also according to the lower RMSD variation of the loop facing the binding site with respect to the apo form (Figure 5, red line).

Considering that in the PDB repository a full homotrimeric PNP apo structure was not available we performed a steered molecular dynamics on each of the monomers of the 3k8o structure, to obtain the apo form of the PNP monomer (PDB code 1m73). For this aim we used Gromacs 4.6.1,⁷⁵ Plumed2⁷⁸ and the RMSD collective variable. We run a 1 ns simulation in NVT ensemble and steered the heavy atoms of residues in the flexible loop/alpha-helix facing the binding site, namely from Val246 to Lys266. An equal weight was given to the atoms during the steering process and a target value of 0 in RMSD was employed.

Together with this structure, the structures of PNP complexed with 8-azaguanine,⁵⁰ acyclovir⁵¹ and 2-mercapto(3H) quinazoline⁵² were retrieved from the PDB (PDB codes 1v41, 1pwy and 3d1v, respectively) and aligned to the PNP-DATMe-Immucillin H complex, to enrich the set of conformations to be used in the virtual screening campaign.

ABL1. The chain A of the PDB code structure 2hzi was used as ABL1 model. We followed the same equilibration protocol as for PNP. To run the production simulation we scaled the MD potential energy with a factor of 0.5. At the same time, to only accelerate the activation loop (Asp381-Glu409) and maintain the overall fold of the protein, we restrained all the protein backbone except the activation loop with an harmonic positional restraint (spring constant equal to 50 kJ/mol/nm²). Some of the authors recently employed an analogous strategy to accelerate protein-ligand unbinding and study the unbinding kinetics.⁶⁴ The simulation was run for 100 ns. This acceleration significantly increased the diversity of the sampled configurations and allowed the binding site to accommodate both type I and type II inhibitors. Strictly speaking the sampled configurations are at high temperature and thus, from a thermodynamics point of view, we should reweight their occurrences. However here we were more interested in the sampling itself, regardless of the relative free energy difference between the basins. Interestingly, this acceleration method can be easily applied to several computational pipelines and will be further investigated in future works. The clustering was done setting 10 clusters and using as metric the RMSD between the different conformations.

A2A. The A2A model was built starting from PDB codes 3uzc and 4eiy. We used the 3uzc structure because of the absence of the apocytochrome b(562)RIL present in 4eiy and employed to obtain the crystal. Ala277, being in the nearby of the binding

site, was back-mutated to serine. At the same time the resolved extracellular loop and the sodium ion were recovered from structure 4eiy. We built a POPC membrane bilayer with 75x75 Å size and with a water layer of 30 Å on each side. To assemble the bilayer we used the CharmmGUI Server^{79,80} and we pre-equilibrated the template via NAMD.⁸¹ To define the pose and to perform the physical insertion of the protein within the membrane we applied the Protein Membrane tool present in BiKi LifeScience.¹ The equilibration protocol was analogous to the one previously described. The production was run for 100 ns in NPT ensemble and the pressure was kept anisotropic (X,Y plane).

All molecular dynamics simulations were run on a Intel Xeon CPU E3-1245 v3 @3.40GHz workstation equipped with two NVIDIA GeForce GTX 780, running a RedHat Linux operating system based on kernel 2.6.32-431 x86_64.

The DUD-E compound datasets

The database containing active ligands and decoys for the three different targets, PNP, ABL1 and A2A were retrieved from the DUD-E dataset (<http://dude.docking.org/>),⁸² an advanced version of the DUD dataset,⁸³ containing also tautomers and protomers. In the case of PNP 225 active ligands (clustered into 103) and 6950 decoys were present. For ABL1 the dataset contained 409 actives (clustered in 182) and 10885 decoys, while for A2A the dataset included 842 actives (clustered in 482) and 32026 decoys. Each dataset was then processed by FLAP and up to 25 conformers for each compound built, using the default RMSD threshold equal to 0.30. Molecular Interaction Fields were then calculated for each conformer and stored in the FLAP database used to perform the subsequent virtual screening analyses.

Virtual Screening

FLAP. The virtual screening experiments were performed with FLAP (Fingerprints for Ligands and Proteins),^{36,84} developed and licensed by Molecular Discovery Ltd. (www.moldiscovery.com). Several VS campaigns have been successfully performed with FLAP and reported in literature.⁸⁵⁻⁸⁸

FLAP describes small molecules and protein binding sites in terms of 4-point pharmacophoric fingerprints, extracted from the Molecular Interaction Fields (MIFs) calculated by GRID.^{89,90} Typically, the used probes are H (mapping the shape), DRY (evaluating hydrophobic affinities), O and N1 (mapping H-bond donor and acceptor regions, respectively). The information contained in the MIFs is extracted and condensed in quadruplets of pharmacoforic points, used to compare, align and superimpose different chemical entities, which can be either small molecules or macromolecules, usually described in terms of pockets. This procedure allows the selection of the most interesting candidates with chemical and structural complementarity with the receptor binding site (SBVS), or similarity with known ligands (LBVS). In SBVS studies, the algorithm calculates the receptor GRID-MIFs and describes the binding pocket in terms of MIF points quadruplets. As well, for each screened compound, atoms are combined in quadruplets. Matching quadruplets belonging to the pocket and to the ligands are used to overlay the compounds onto the receptor model during the screening process. This superimposition is quantitatively scored by considering the corresponding MIFs similarity, which is collapsed in eighteen different scores.

Probe scores are calculated for a given superimposition by directly comparing the volumes of the oriented MIFs, both on a probe-by-probe basis, as well as for probe combinations. For every solution being scored, FLAP first calculates scores which represent the degree of volume overlap for each of the probes (and of the corresponding generated MIFs) being used individually, i.e. H, DRY, O and N1, and then combines these scores in order to produce probe-combination scores. For instance, the DRY*O score corresponds to the product of the DRY and O MIFs overlap. In addition, FLAP also calculates two Global scores, the Global Sum, which is produced by summing all the scores of the individual probes together, and the Global Product, produced by multiplying all the scores of the individual probes together. The Global Product score is corrected when one of the terms is zero, to avoid a score equal to zero if three of four fields matched perfectly. The Probe Score finally saved for every individual probe and for every probe combination is the one giving the highest similarity.

Once the Probe scores for the individual probes and their combinations have been calculated, including the Global Sum and Global Product, FLAP will calculate a Distance Score, representing the overall similarity derived by a combination of the overlap degree between the single H, DRY, O and N1 MIFs computed for the candidates and the template, that is, the protein binding site.⁸⁴

All these different scores can be used to rank the compounds with respect to the template.

PNP VS. In the case of the SBVS performed on the different single X-ray structures of PNP, compounds were ranked according to the FLAP Distance from the template.⁸⁴ In all the screenings the binding site was identified by the ligand DATMe-Immucillin H extracted from the complex with PNP (PDB code 3k8o). All the other X-ray structures and the MD-generated medoids were aligned with the aforementioned complex. To take into account protein flexibility the ten medoids extracted from the whole MD trajectory by the clustering algorithm were also used as possible templates in the VS simulations, as well as the X-ray structures.

The Linear Discriminant Analysis approach implemented into FLAP was used to select the most representative templates among the medoids and the X-ray structures. The template selection is performed to maximize the capability of the corresponding probe scores linear combinations to discriminate between actives and decoys.

LDA analyses are commonly used to define the relationship between a non-metric dependent variable (Y) and metric independent variables (X), being respectively Y the class assigned to each ligand (activity or inactivity) and X the templates and the probe scores produced by FLAP.

The effectiveness of a discriminating model relies on its capability to predict which class an object belongs to, by defining a new variable called the discriminant function score. Linear discriminant function scores are computed by determining the coefficients for the independent variables able to maximize the distance between the classes (actives and inactives) corresponding to the dependent variable.

Mathematically the linear discriminant function is similar to a linear regression equation in which the independent variables are multiplied by the aforementioned coefficients and then summed to produce a score. From a geometrical point of view, assuming data is D-dimensional, a linear discriminant function define (D-1) dimensional hyperplanes to discriminate the objects in different classes and define the boundary between these groups.

FLAP produces a multitude of LDA models by making all possible combinations between the specified number of probe scores for every possible combination of the

specified number of template candidates (independent variables). Each model is then evaluated, in fitting and in prediction, using the Leave One Out (LOO) approach. Once a large number of LDA models have been generated and validated, the user can select the model to be used in the following VS analyses. When carrying out virtual screening using an LDA model, FLAP will produce "Activity Class" predictions for each investigated molecule, besides the usual output forms, and a corresponding LDA-R score ranking the compounds from the most active (highest score) to the most inactive (lowest score). When the inclusion of multiple structures improves VS predictions, the LDA-R score ranking gives the highest enrichment.

For the PNP test case different LDA models were calculated using, each time, the ten medoids extracted from the MD trajectory by the clustering algorithm and a different X-ray structure, i.e. LDA-MRC-3k8o (3k8o X-ray structure + 10 medoids); LDA-MRC-apo (apo generated structure + 10 medoids); LDA-MRC-1v41 (1v41 X-ray structure + 10 medoids); LDA-MRC-1pwy (1pwy X-ray structure + 10 medoids); LDA-MRC-3d1v (3d1v X-ray structure + 10 medoids). LDA models were also calculated for the ten medoids only and for the five analyzed X-ray structures, i.e. LDA-MRC-med (ten medoids) and LDA-MRC-Xray (3k8o, apo, 1v41, 1pwy, 3d1v). A 3 templates/3 scores approach was used for all the LDA models.

The related VS were then performed using as templates the aforementioned models, i.e. MRC-VS-3k8o; MRC-VS-apo; MRC-VS-1v41; MRC-VS-1pwy; MRC-VS-3d1v; MRC-VS-med; MRC-VS-Xray.

The LDA-MRC-med20 models were built using the ten medoids extracted by the BiKi clustering with and without the HPO_4^{2-} ion. Again, a 3 templates/3 scores approach was used.

A ligand-based VS was run in order to reduce the original DUD-E dataset and remove compounds most similar to the co-crystallized DATMe-Immucillin H ligand. The ligand was used as template and the dataset compounds were ranked according to the FLAP Distance from the template. The first 1% of ranked molecules was removed from the dataset, irrespectively if they were actives or decoys.

The original DUD-E datasets was then randomly separated in training and test sets, by using the random option implemented in "sort" (GNU coreutils, version 8.17). The same procedure was applied to generate training and test sets for the other two cases, ABL and A2A.

An LDA-MRC-3k8o model (simply named LDA-MRC-PNP in Table 5 and within the text) was then calculated for the training set. Again a 3 templates/3 scores approach was used. The model was then validated on the test set and the corresponding MRC-VS performed on the test set to get the best AUC in prediction. The LDA-SRC-PNP model, reported in Table 5 and used to evaluate the contribution of multiple FLAP scores with respect to a standard SRC model, was built using as template the only X-ray 3k8o structure, but allowing the use of three different scores. Similarly, the model was built on the training set and validated on the test set, and the corresponding VS run to get the best AUC in prediction

ABL1 VS. In the case of tyrosine-protein kinase ABL1 the SRC VS was performed on the X-ray reference structure 2hzi upon removal of the co-crystallized PD180970 ligand. Compounds were ranked according to the DRY*N1 score, giving the best performances and corresponding to the product of the DRY and N1 MIFs overlap. Considering the high level of flexibility allowed by the scaled MD simulation and generally experienced by kinases, the binding site of the X-ray structure and of the

MD-generated medoids was identified by automatically calculating the pocket volume and shape with the flapsite algorithm implemented into FLAP.⁹¹ Thus, differently from the PNP and A2A cases, no co-crystallized ligand was used to define the binding site to avoid any user-dependent pocket definition.

The reference X-ray structure, and the ten medoids extracted from the scaled MD trajectory were used as possible templates to calculate the LDA-MRC-ABL1 model using, again, a 3 templates/3 scores approach. The LDA-MRC-ABL1 model was built on the training set and validated on the test set. The MRC-VS was performed on the test set in order to have the best AUC in prediction. The LDA-SRC-ABL1 model was built on the training set using one template, i.e. the reference 2hzi X-ray structure, and three FLAP scores. The model was validated on the test set and, again, used to have the best AUC in prediction

A2A VS. For the adenosine A_{2A} receptor the SRC VS was run on the model used to setup and perform MD simulations. For further details about the structure preparation see the previous “*Molecular dynamics guided conformational sampling and clustering*” section.

As in the PNP case, compounds were ranked according to the FLAP Distance from the template, i.e. the protein binding site. The co-crystallized ligand was used to define the binding site of the original X-ray structure and of the MD-generated conformation upon alignment with the X-ray structure.

Again, the X-ray structure and the ten medoids extracted from the plain MD trajectory were used as possible templates to calculate the LDA-MRC-A2A model using a 3 templates/3 scores approach. The LDA-MRC-A2A model was built on the training set and validated on the test set. The MRC-VS was performed on the test set in order to have the best AUC in prediction. The LDA-SRC-A2A model was built on the training set using one template, i.e. the model structure used in MD simulations, and three FLAP scores. The model was validated on the test set and, again, used to have the best AUC in prediction

The description of the benchmarking performed with AutoDock Vina⁹² and Glide,^{93,94} with respect to FLAP, on the PNP case is reported in the Supplementary Information.

Virtual screening analyses and any other FLAP calculation were run on a 8-cores Intel i7-3632QM @2.20GHz with 8GB 1600MHz of RAM, and a Linux operating system based on kernel version 3.13.6-200 x86_64.

Figures of merit

Different metrics can be used to evaluate the efficacy of docking/screening methodologies to discriminate between actives and decoys.

We adopted here the Receiver Operating Characteristic curve (ROC), able to graphically display the trade-off between the true positive rate (TPR, positives correctly classified/total positives) and the false positive rate (FPR, negatives correctly classified/total negatives) of the used classifiers. The area under the ROC curve (AUC) numerically quantifies the performance of a classifier, i.e. a score or combinations of them.

A ROC curve can be calculated using the following equation:

$$ROC = \frac{1}{(nN)} \sum_{k=2}^N F_a(k) [F_i(k) - F_i(k-1)]$$

where n corresponds to the number of actives, N to the total number of molecules in the database, $F_a(k)$ is the accumulation curve, representing how many true binders obtained a rank better or equal to a given one in a docking run, and the subscript i stands for inactive molecules.⁹⁵

In order to better discriminate the different VS methods performances in the early recognition stage we used the ROC enrichments, quantifying the area covered by the curve at the 0.5%, 1%, 2% and 5% of the screened false positives.

CONCLUSIONS

In the current paper we showed that a pipeline composed of Molecular Dynamics for sampling, clustering for filtering, and LDA to build a predictive model can substantially improve VS performances. Experimental results on Purine Nucleoside Phosphorylase, tyrosine-protein kinase ABL1 and adenosine A_{2A} receptor confirmed the effectiveness of the approach. Additional analysis allowed us to take into account ligand-induced structural biases, as well as the possible advantage of performing configurational space sampling by alternatively keeping or removing cofactors in the binding site. It resulted that the supervised inclusion of MD information significantly improves VS performance in particular when dealing with highly flexible targets, and overcomes the possible drawbacks associated to the choice of a single, possibly bad-performing, structure. Lastly the MD-clustering-LDA integrated pipeline allowed a proper and rapid (via current GPU based MD) extraction of MD-generated conformations.

ABBREVIATIONS

A_{2A}, adenosine A_{2A} receptor; ABL1, tyrosine-protein kinase ABL1; AUC, area under the ROC curve; ImmH, Immucillin-H; LBVS, ligand-based virtual screening; LDA, linear discriminant analysis; LOO, leave one out; MD, molecular dynamics; MRC, multiple receptor conformations; PNP, purine nucleoside phosphorilase; ROC, receiver operating characteristic curve; SBVS, structure-based virtual screening; SRC, single receptor conformation; TSA, transition state analogues; VS, virtual screening.

ASSOCIATED CONTENT

SUPPORTING INFORMATION

The FLAP benchmarking with respect to AutoDock Vina and Glide. The RMSD backbone for the PNP MD trajectory. The medoids superimposition for the PNP test case. The ROC curves comparison for the LDA-SRC-3k8o and the LDA-MRC-3k8o virtual screenings. The AUC and ROC enrichments obtained for the PNP reduced dataset. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHORS INFORMATION

Corresponding Authors

Gabriele Cruciani

Phone: 0039 075 5855629. E-mail: gabri@chemiome.chm.unipg.it.

Francesca Spyrakis

Phone: 0039 059 2058604. E-mail: francesca.spyrakis@unimore.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We kindly acknowledge Dr. Mattia Mori for the helpful discussions and for the compounds clusterization support. We thank Roberto Gaspari for preparing the A2A structure.

FIGURES & TABLES

Scheme 1.

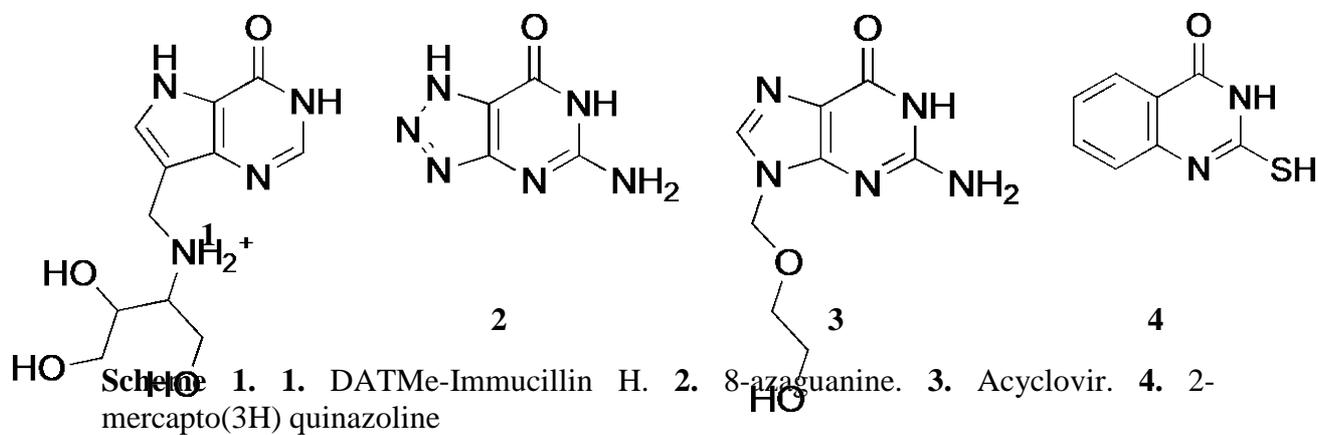


Figure 1.

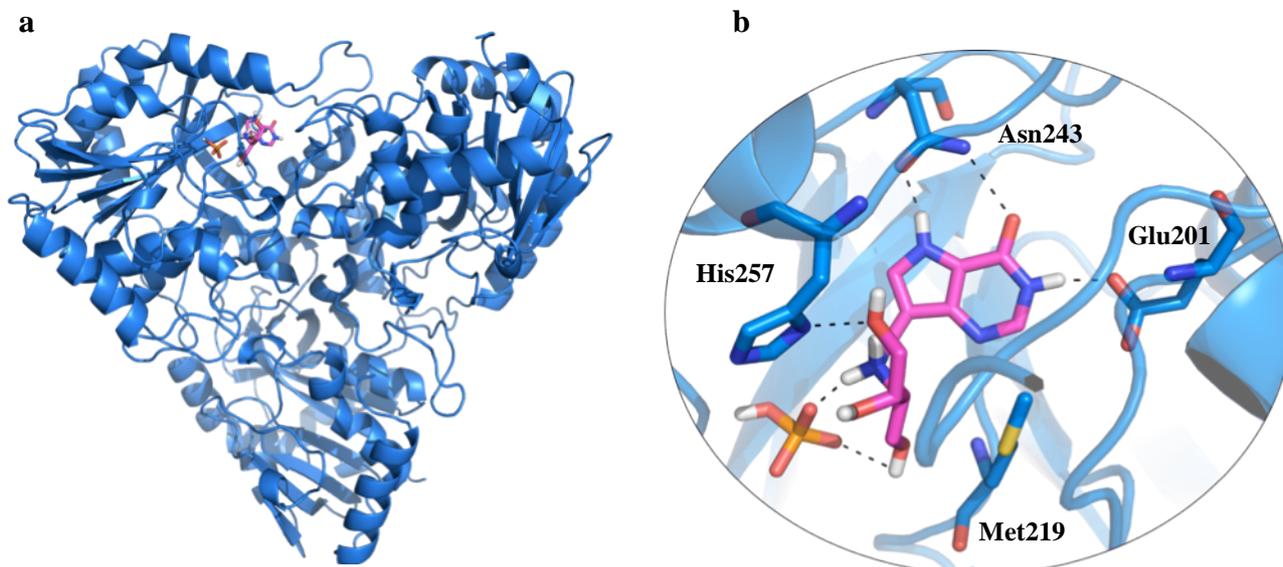


Figure 1. a. Overall folding of the PNP homotrimer. The DATMe-Immucillin H ligand bound to monomer one and the phosphate ion are shown in capped sticks. **b.** Close up of the binding site. The ligand, the phosphate and the residues lining the binding site are shown in capped sticks. The hydrogen bonds formed between the ligand and the pocket residues are displayed in black dash lines.

Figure 2.

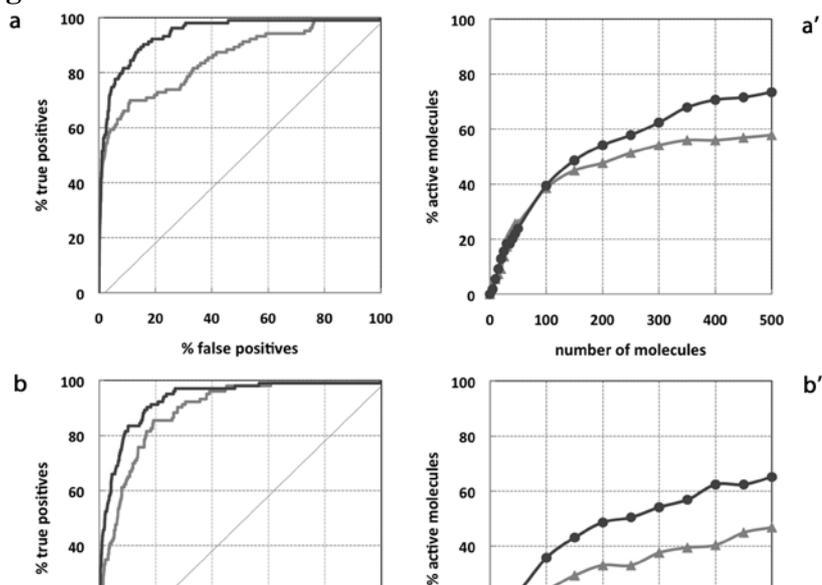


Figure 2. Comparison of the ROC curves (left panels) and of the percentage of active molecules found in the first 500 ranked molecules (right panels) for the single-structure-VS and for the MRC-VS.

ROC curves (**a**) and active molecules % (**a'**) of the SRC-3k8o VS (light grey line) and of the LDA-MRC-3k8o VS (dark grey line). ROC curves (**b**) and active molecules % (**b'**) of the SRC-apo VS (light grey line) and of the LDA-MRC-apo VS (dark grey line). ROC curves (**c**) and active molecules % (**c'**) of the SRC-1v41 VS (light grey line) and of the LDA-MRC-1v41 VS (dark grey line). ROC curves (**d**) and active molecules % (**d'**) of the SRC-1pwy VS (light grey line) and of the LDA-MRC-1pwy VS (dark grey line). ROC curves (**e**) and active molecules % (**e'**) of the SRC-3d1v VS (light grey line) and of the LDA-MRC-3d1v VS (dark grey line).

Figure 3.

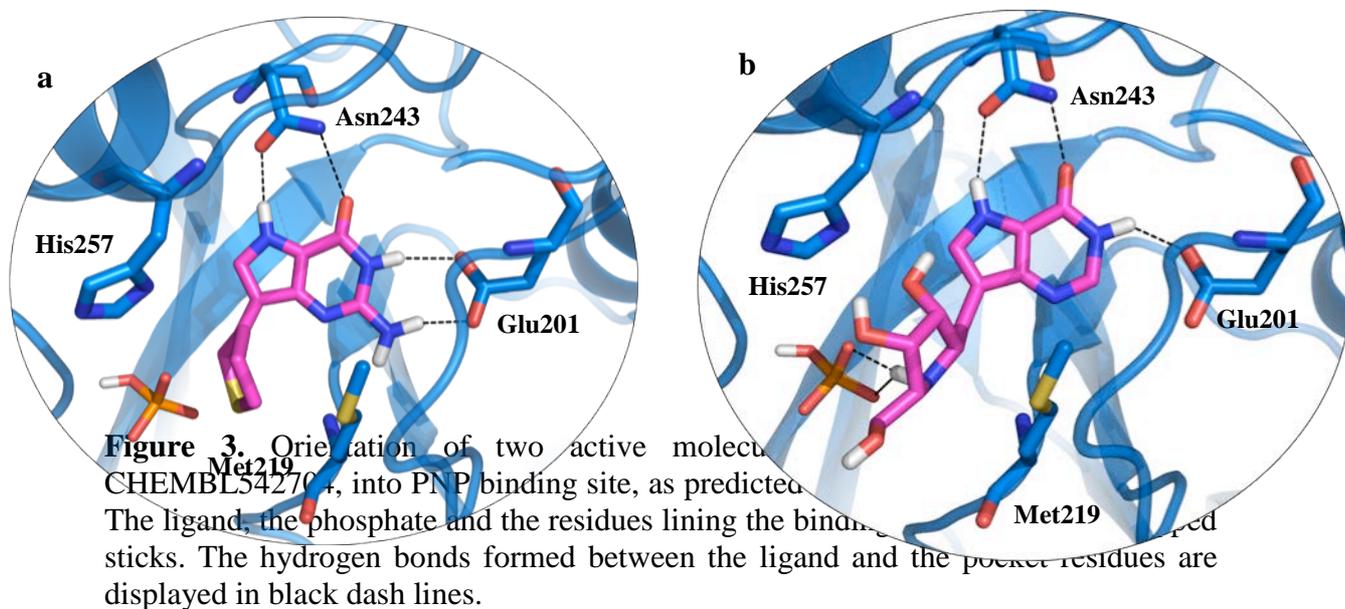


Figure 4.

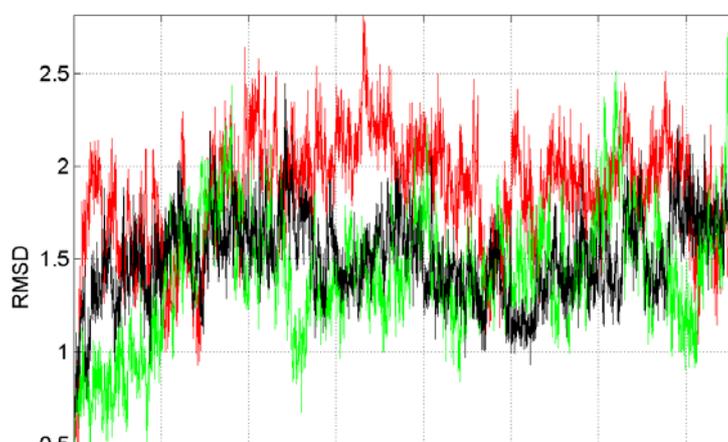


Figure 4. RMSD of the backbone atoms of the binding site frontal loop with respect to the first production frame. Each color encodes a specific loop, one for each binding site.

Figure 5.

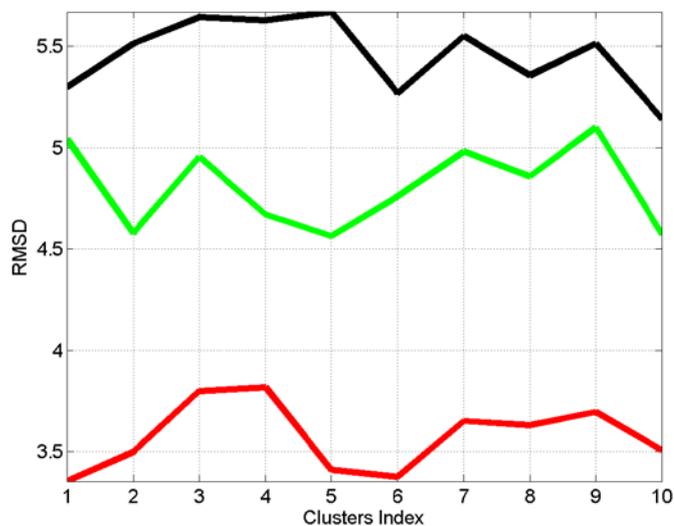


Figure 5. F
red line, m
value betw
Data show

, monomer 1 in
ster the RMSD
is represented.

Figure 6.

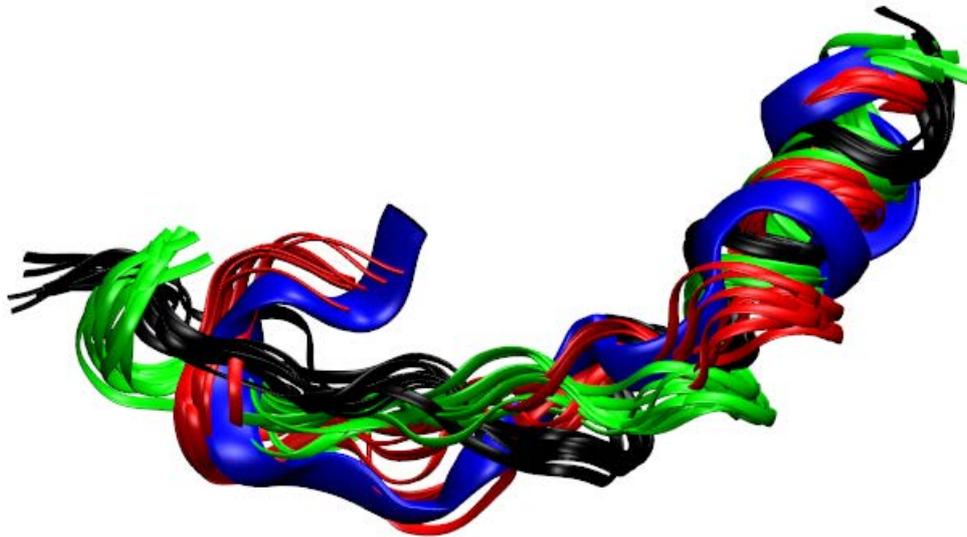


Figure 6. Over-imposition with respect to the RMSD of the clusters medoids conformations of the binding site loop. In blue the apo structure, the other colors map the monomers as per the previous figure.

Figure 7.

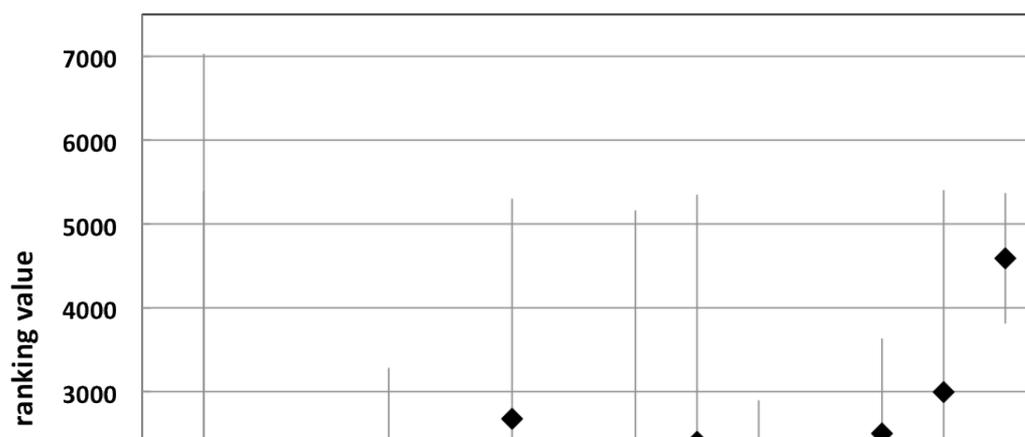


Figure 7. Average rank value for each active clustered class, calculated for the SRC-3k8o (black diamonds) and the LDA-MRC-3k8o (grey squares) screenings. The vertical lines represent the range of ranking values for each different class in the two virtual screenings.

Figure 8.

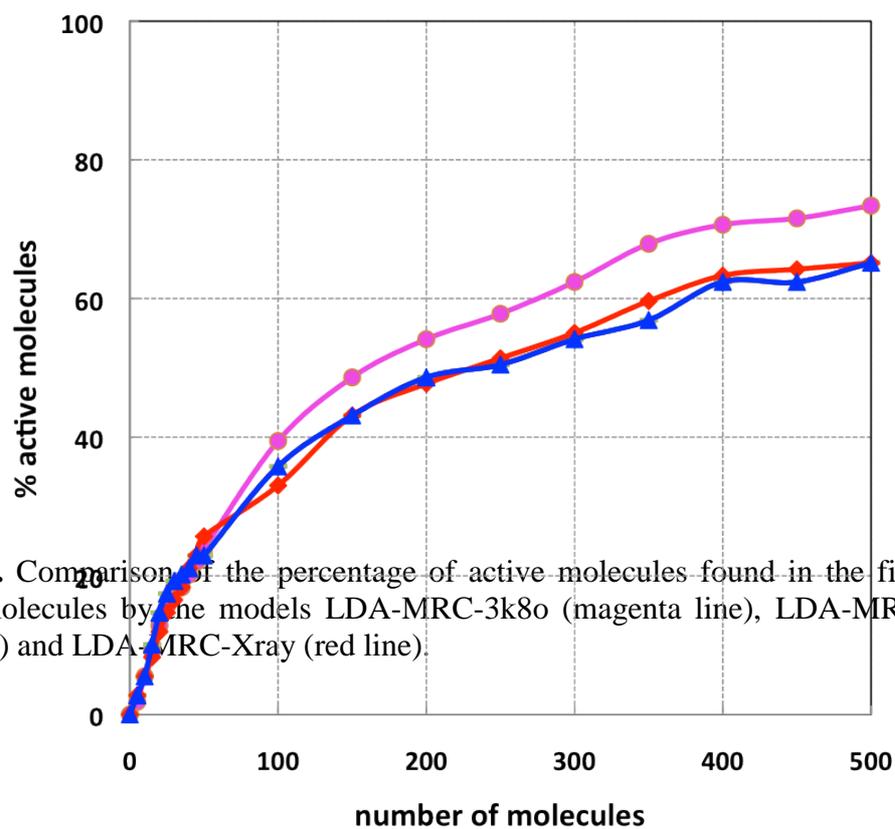


Figure 8. Comparison of the percentage of active molecules found in the first 500 ranked molecules by the models LDA-MRC-3k8o (magenta line), LDA-MRC-med (blue line) and LDA-MRC-Xray (red line).

Figure 9.

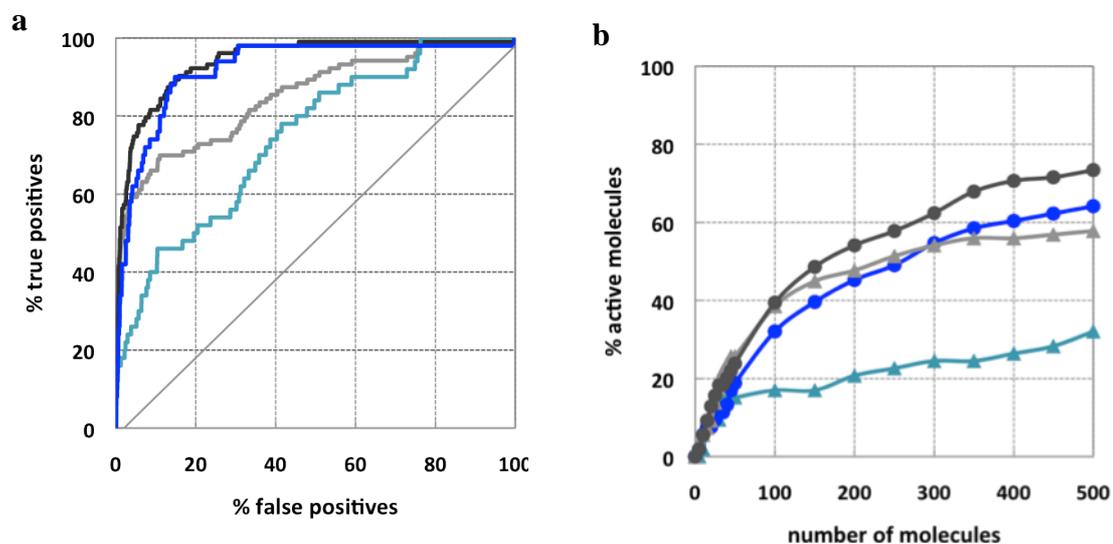


Figure 9. Comparison of the ROC curves (left panel, **a**) and of the percentage of active molecules found in the first 500 ranked molecules (right panel, **b**) by the models SRC-3k8o and LDA-MRC-3k8o (light grey and dark grey lines, respectively) when screening the entire dataset, and by the models SRC-3k8o' and LDA-MRC-3k8o' (light blue and blue lines, respectively) when screening the reduced dataset.

Figure 10.

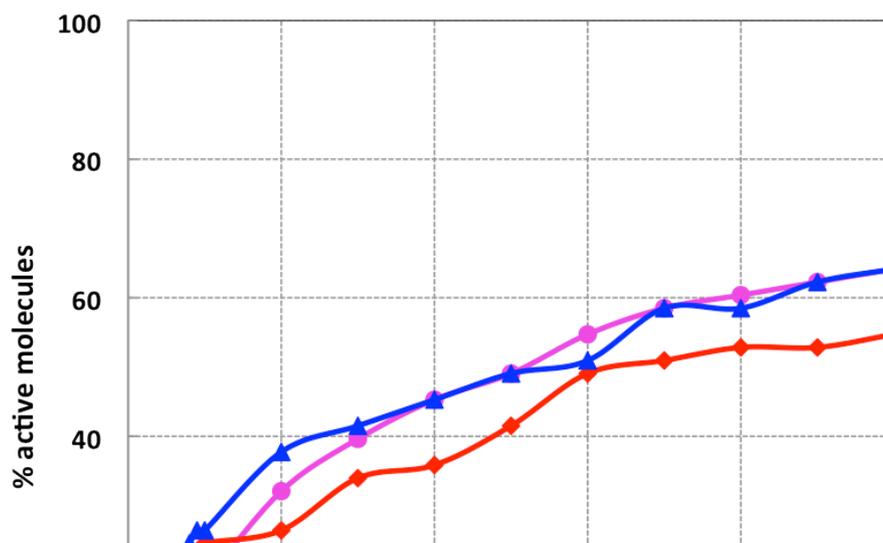


Figure 10. Comparison of the percentage of active molecules found in the first 500 ranked molecules by the models LDA-MRC-3k8o' (magenta line), LDA-MRC-med' (blue line) and LDA-MRC-Xray' (red line). The VS were performed against the reduced dataset.

Figure 11.

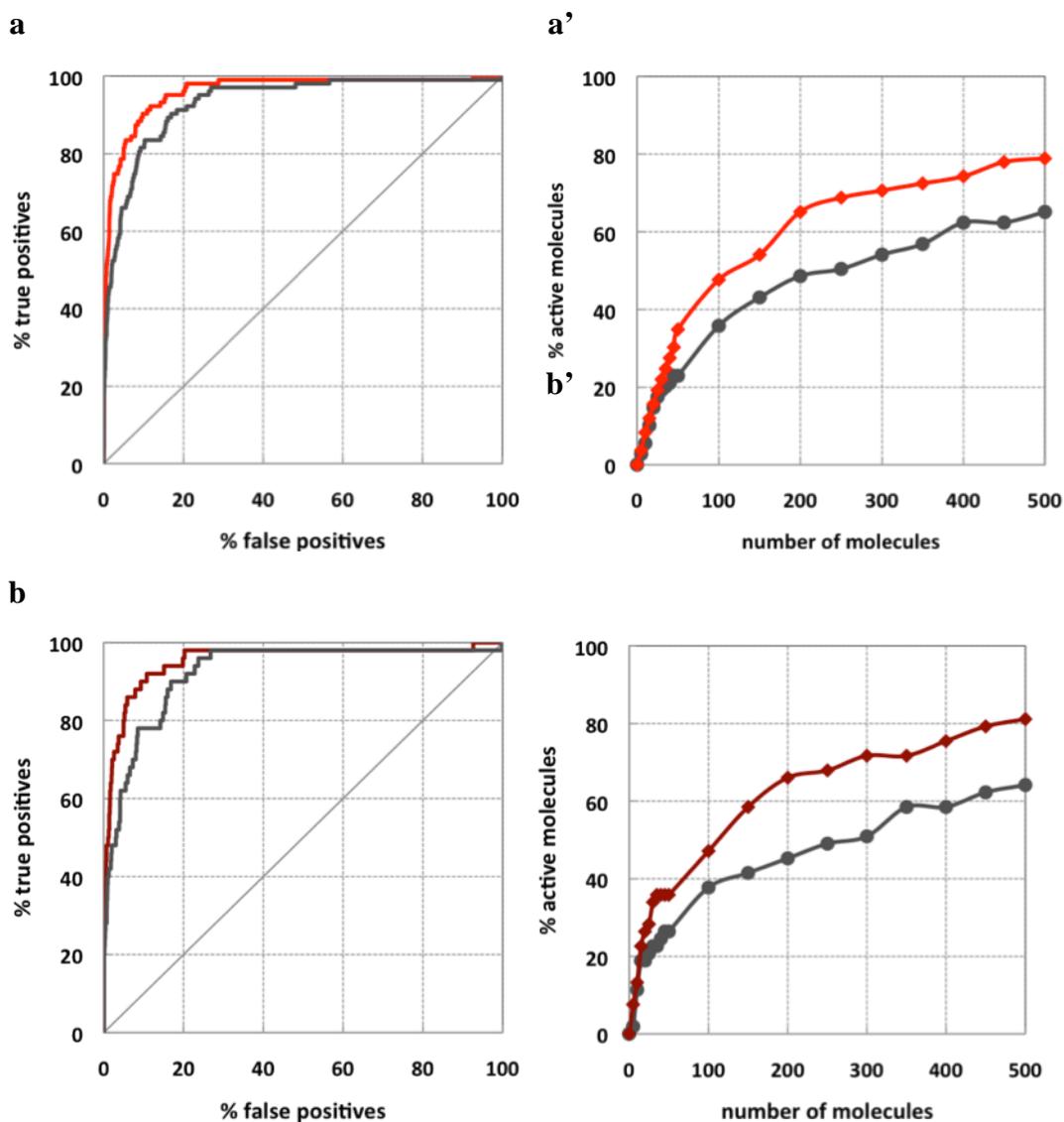


Figure 11. Comparison of the ROC curves (left panels) and of the percentage of active molecules found in the first 500 ranked molecules (right panels) molecules by the models LDA-MRC-med (grey lines panels **a** and **a'**), LDA-MRC-med20 (red lines panels **a** and **a'**), LDA-MRC-med' (grey lines panels **b** and **b'**) and LDA-MRC-med20' (dark red lines panels **b** and **b'**).

Figure 12.

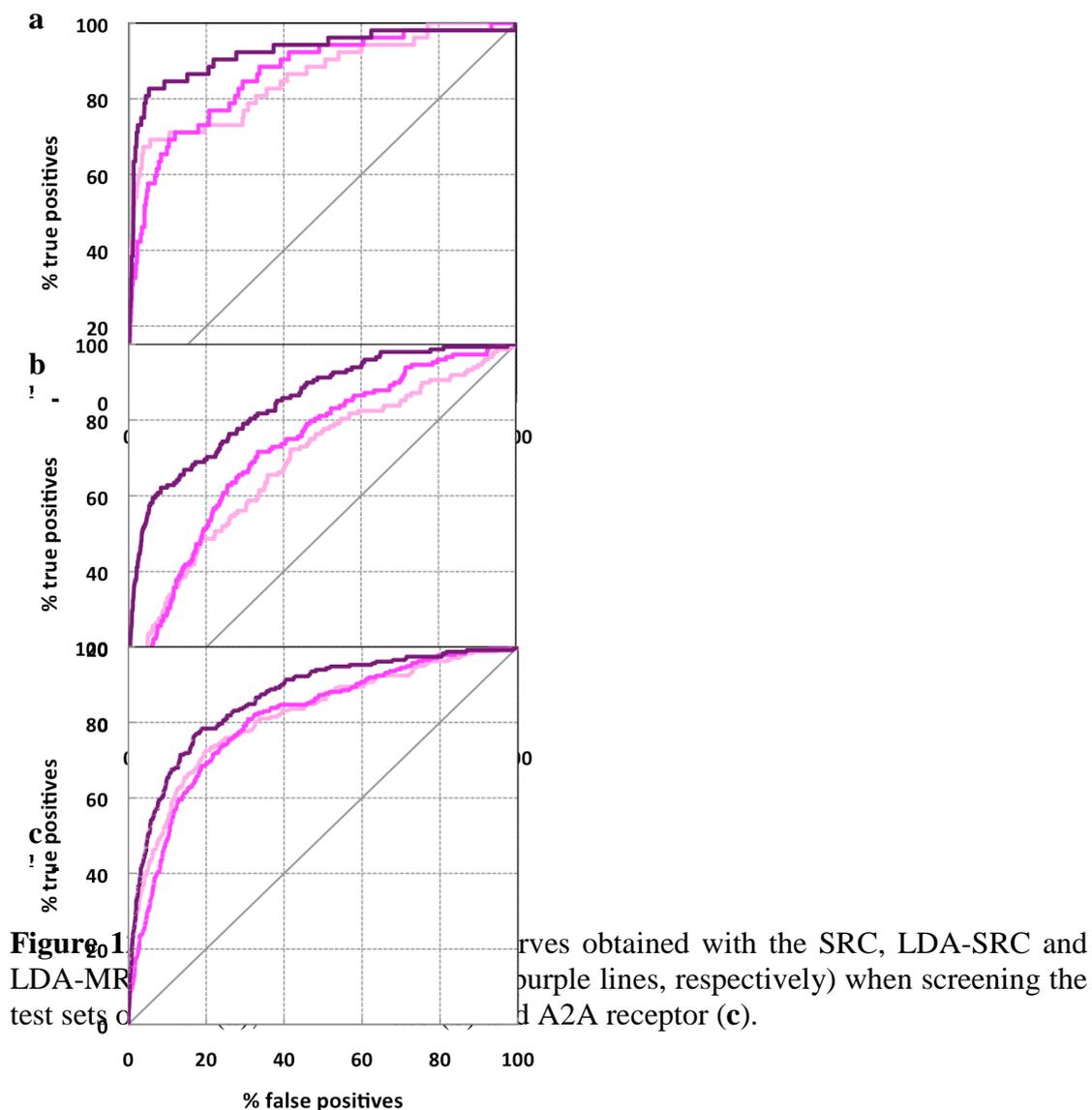


Figure 12. ROC curves obtained with the SRC, LDA-SRC and LDA-MRC (black, red and purple lines, respectively) when screening the entire database (a) and A2A receptor (c).

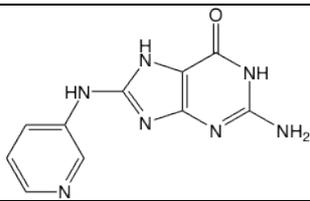
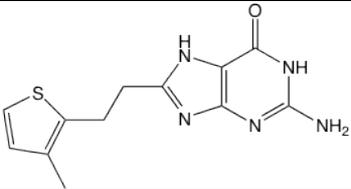
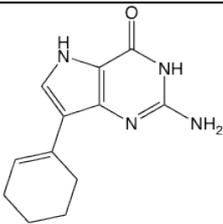
Table 1. Total AUC and ROC enrichments calculated for the SRC and MRC virtual screenings performed on the entire database^a

VS	AUC	0.5 %	1 %	2 %	5 %
----	-----	-------	-----	-----	-----

SRC-3k8o	0.85	0.33	0.42	0.50	0.59
SRC-apo	0.89	0.18	0.25	0.33	0.43
SRC-1v41	0.79	0.13	0.22	0.30	0.40
SRC-1pwy	0.89	0.16	0.24	0.46	0.57
SRC-3d1v	0.89	0.33	0.38	0.48	0.61
LDA-MRC-3k8o	0.94	0.29	0.46	0.57	0.75
LDA-SRC-3k8o	0.88	0.21	0.26	0.35	0.52
LDA-MRC-apo	0.93	0.32	0.41	0.50	0.66
LDA-MRC-1v41	0.93	0.32	0.41	0.50	0.66
LDA-MRC-1pwy	0.93	0.32	0.41	0.50	0.66
LDA-MRC-3d1v	0.93	0.32	0.41	0.50	0.66
LDA-MRC-med	0.93	0.32	0.41	0.50	0.66
LDA-MRC-Xray	0.92	0.32	0.36	0.50	0.68

^aROC enrichments were estimated at 0.5%, 1%, 2% and 5% of screened false positives.

Table 2. Ranking positions in the SRC-3k8o and LDA-MRC-3k8o VSs of selected active compounds

compound	chemical structure	SRC-3k8o VS ranking position	LDA-MRC-3k8o VS ranking position
CHEMBL159266		4202	16
CHEMBL350773		2976	114
CHEMBL301203		672	120

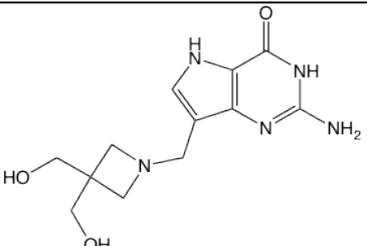
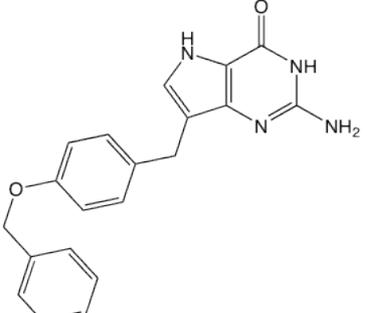
CHEMBL407776		1926	56
CHEMBL83158		606	241

Table 3. Total AUC and ROC enrichments calculated for the SRC and MRC virtual screenings performed on the reduced database^a

VS	AUC	0.5 %	1 %	2 %	5 %
SRC-3k8o'	0.75	0.16	0.16	0.18	0.26
SRC-apo'	0.84	0.12	0.14	0.18	0.24
SRC-1v41' ^b	0.73	0.12	0.16	0.24	0.32
SRC-1pwy'	0.83	0.08	0.10	0.28	0.38
SRC-3d1v'	0.85	0.22	0.26	0.36	0.50
LDA-MRC-3k8o'	0.92	0.18	0.32	0.42	0.62
LDA-MRC-apo'	0.92	0.28	0.40	0.46	0.62
LDA-MRC-1v41'	0.92	0.28	0.40	0.46	0.62
LDA-MRC-1pwy'	0.92	0.28	0.40	0.46	0.62
LDA-MRC-3d1v'	0.92	0.28	0.40	0.46	0.62
LDA-MRC-med'	0.92	0.28	0.40	0.46	0.62
LDA-MRC-Xray'	0.88	0.26	0.26	0.36	0.56

^aROC enrichments were estimated at 0.5%, 1%, 2% and 5% of screened false positives.

^bFor the SRC-1v41' model, the H*DRY score was used to have the best enrichment.

Table 4. Total AUC and ROC enrichments calculated for the MRC virtual screenings performed on the total (LDA-MRC-med and LDA-MRC-med20 models) and on the reduced database (LDA-MRC-med' and LDA-MRC-med20' models)^a

VS	AUC	0.5 %	1 %	2 %	5 %
LDA-MRC-med	0.93	0.32	0.41	0.50	0.66
LDA-MRC-med20	0.96	0.46	0.53	0.70	0.80
LDA-MRC-med'	0.92	0.28	0.40	0.46	0.62
LDA-MRC-med20'	0.95	0.42	0.48	0.66	0.80

^aROC enrichments were estimated at 0.5%, 1%, 2% and 5% of screened false positives.

Table 5. Total AUC and ROC enrichments calculated for the SRC, SRC-LDA and MRC-LDA virtual screenings performed on the test sets for PNP, ABL and A2A^a

VS	AUC	0.5 %	1 %	2 %	5 %
SRC-PNP	0.86	0.40	0.42	0.56	0.67
LDA-SRC-PNP	0.86	0.23	0.31	0.37	0.58
LDA-MRC-PNP	0.92	0.27	0.38	0.69	0.81
SRC-ABL1	0.69	0.03	0.07	0.10	0.23
LDA-SRC-ABL1	0.73	0.03	0.05	0.10	0.18
LDA-MRC-ABL1	0.85	0.22	0.30	0.39	0.55
SRC-A2A	0.81	0.15	0.19	0.28	0.41
LDA-SRC-A2A	0.81	0.08	0.10	0.17	0.32
LDA-MRC-A2A	0.86	0.13	0.20	0.30	0.49

^aFor LDA-SRC models only the original X-ray was used as template while three FLAP scores were allowed. For LDA-MRC models a 3 templates/3 scores combination was adopted. All LDA models were built on the training sets and

validated on the corresponding test sets. ROC enrichments were estimated at 0.5%, 1%, 2% and 5% of screened false positives.

REFERENCES

1. <http://www.bikitech.com> (accessed Jul. 21, 2015),
2. <http://www.moldiscovery.com> (accessed Jul. 21, 2015),
3. Cozzini, P.; Kellogg, G.E.; Spyraakis, F.; Abraham, D.J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L.A.; Morris, G.M.; Orozco, M.; Pertinhez, T.A.; Rizzi, M.; Sotriffer, C.A. Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *J. Med. Chem.* **2008**, *51*, 6237-6255.
4. Spyraakis, F.; BidonChanal, A.; Barril, X.; Luque, F.J. Protein Flexibility and Ligand Recognition: Challenges for Molecular Modeling. *Curr. Top. Med. Chem.* **2011**, *11*, 192-210.
5. Boehr, D.D.; Nussinov, R.; Wright, P.E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5*, 789-796.
6. Teague, S.J. Implications of Protein Flexibility for Drug Discovery. *Nat. Rev. Drug Discov.* **2003**, *2*, 527-541.
7. Teodoro, M.L.; Kavradi, L.E. Conformational Flexibility Models for the Receptor in Structure Based Drug Design. *Curr. Pharm. Des.* **2003**, *9*, 1635-1648.
8. Sinko, W.; Lindert, S.; McCammon, J.A. Accounting for Receptor Flexibility and Enhanced Sampling Methods in Computer-Aided Drug Design. *Chem. Biol. Drug Des.* **2013**, *81*, 41-49.
9. Bowman, A.L.; Nikolovska-Coleska, Z.; Zhong, H.; Wang, S.; Carlson, H.A. Small Molecule Inhibitors of the Mdm2-P53 Interaction Discovered by Ensemble-Based Receptor Models. *J. Am. Chem. Soc.* **2007**, *129*, 12809-12814.
10. Carlson, H.A.; Masukawa, K.M.; Rubins, K.; Bushman, F.D.; Jorgensen, W.L.; Lins, R.D.; Briggs, J.M.; McCammon, J.A. Developing a Dynamic Pharmacophore Model for Hiv-1 Integrase. *J. Med. Chem.* **2000**, *43*, 2100-2114.
11. Durrant, J.D.; Urbaniak, M.D.; Ferguson, M.A.; McCammon, J.A. Computer-Aided Identification of Trypanosoma Brucei Uridine Diphosphate Galactose 4'-Epimerase Inhibitors: Toward the Development of Novel Therapies for African Sleeping Sickness. *J. Med. Chem.* **2010**, *53*, 5025-5032.
12. Lin, J.H.; Perryman, A.L.; Schames, J.R.; McCammon, J.A. Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632-5633.
13. Schames, J.R.; Henschman, R.H.; Siegel, J.S.; Sotriffer, C.A.; Ni, H.; McCammon, J.A. Discovery of a Novel Binding Trench in Hiv Integrase. *J. Med. Chem.* **2004**, *47*, 1879-1881.
14. Sinko, W.; de Oliveira, C.; Williams, S.; Van Wynsberghe, A.; Durrant, J.D.; Cao, R.; Oldfield, E.; McCammon, J.A. Applying Molecular Dynamics Simulations to Identify Rarely Sampled Ligand-Bound Conformational States of Undecaprenyl Pyrophosphate Synthase, an Antibacterial Target. *Chem. Biol. Drug Des.* **2011**, *77*, 412-420.
15. Ma, B.; Shatsky, M.; Wolfson, H.J.; Nussinov, R. Multiple Diverse Ligands Binding at a Single Protein Site: A Matter of Pre-Existing Populations. *Protein Sci.* **2002**, *11*, 184-197.
16. Tsai, C.J.; Kumar, S.; Ma, B.; Nussinov, R. Folding Funnels, Binding Funnels, and Protein Function. *Protein Sci.* **1999**, *8*, 1181-1190.
17. Barril, X.; Morley, S.D. Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures. *J. Med. Chem.* **2005**, *48*, 4432-4443.
18. Bottegoni, G.; Rocchia, W.; Rueda, M.; Abagyan, R.; Cavalli, A. Systematic Exploitation of Multiple Receptor Conformations for Virtual Ligand Screening. *PLoS One* **2011**, *6*, e18845.
19. Cavasotto, C.N.; Abagyan, R.A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337*, 209-225.

20. Craig, I.R.; Essex, J.W.; Spiegel, K. Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 511-524.
21. Huang, S.Y.; Zou, X. Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. *Proteins* **2007**, *66*, 399-421.
22. Knegtel, R.M.; Kuntz, I.D.; Oshiro, C.M. Molecular Docking to Ensembles of Protein Structures. *J. Mol. Biol.* **1997**, *266*, 424-440.
23. Park, S.J.; Kufareva, I.; Abagyan, R. Improved Docking, Screening and Selectivity Prediction for Small Molecule Nuclear Receptor Modulators Using Conformational Ensembles. *J. Comput. Aided Mol. Des.* **2010**, *24*, 459-471.
24. Polgar, T.; Baki, A.; Szendrei, G.I.; Keseru, G.M. Comparative Virtual and Experimental High-Throughput Screening for Glycogen Synthase Kinase-3beta Inhibitors. *J. Med. Chem.* **2005**, *48*, 7946-7959.
25. Polgar, T.; Keseru, G.M. Ensemble Docking into Flexible Active Sites. Critical Evaluation of Flexe against Jnk-3 and Beta-Secretase. *J. Chem. Inf. Model.* **2006**, *46*, 1795-1805.
26. Rao, S.; Sanschagrin, P.C.; Greenwood, J.R.; Repasky, M.P.; Sherman, W.; Farid, R. Improving Database Enrichment through Ensemble Docking. *J. Comput. Aided Mol. Des.* **2008**, *22*, 621-627.
27. Rueda, M.; Bottegoni, G.; Abagyan, R. Recipes for the Selection of Experimental Protein Conformations for Virtual Screening. *J. Chem. Inf. Model.* **2010**, *50*, 186-193.
28. Bottegoni, G.; Kufareva, I.; Totrov, M.; Abagyan, R. Four-Dimensional Docking: A Fast and Accurate Account of Discrete Receptor Flexibility in Ligand Docking. *J. Med. Chem.* **2009**, *52*, 397-406.
29. Nichols, S.E.; Baron, R.; Ivetac, A.; McCammon, J.A. Predictive Power of Molecular Dynamics Receptor Structures in Virtual Screening. *J. Chem. Inf. Model.* **2011**, *51*, 1439-1446.
30. Bolstad, E.S.; Anderson, A.C. In Pursuit of Virtual Lead Optimization: The Role of the Receptor Structure and Ensembles in Accurate Docking. *Proteins* **2008**, *73*, 566-580.
31. Sperandio, O.; Mouawad, L.; Pinto, E.; Villoutreix, B.O.; Perahia, D.; Miteva, M.A. How to Choose Relevant Multiple Receptor Conformations for Virtual Screening: A Test Case of Cdk2 and Normal Mode Analysis. *Eur. Biophys. J.* **2010**, *39*, 1365-1372.
32. Michino, M.; Abola, E.; Brooks, C.L., 3rd; Dixon, J.S.; Moulton, J.; Stevens, R.C. Community-Wide Assessment of GPCR Structure Modelling and Ligand Docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.* **2009**, *8*, 455-463.
33. Birch, L.; Murray, C.W.; Hartshorn, M.J.; Tickle, I.J.; Verdonk, M.L. Sensitivity of Molecular Docking to Induced Fit Effects in Influenza Virus Neuraminidase. *J. Comput. Aided Mol. Des.* **2002**, *16*, 855-869.
34. Verdonk, M.L.; Mortenson, P.N.; Hall, R.J.; Hartshorn, M.J.; Murray, C.W. Protein-Ligand Docking against Non-Native Protein Conformers. *J. Chem. Inf. Model.* **2008**, *48*, 2214-2225.
35. Korb, O.; Olsson, T.S.; Bowden, S.J.; Hall, R.J.; Verdonk, M.L.; Liebeschutz, J.W.; Cole, J.C. Potential and Limitations of Ensemble Docking. *J. Chem. Inf. Model.* **2012**, *52*, 1262-1274.
36. Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J.S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands and Proteins (Flap): Theory and Application. *J. Chem. Inf. Model.* **2007**, *47*, 279-294.
37. Cross, S.; Baroni, M.; Carosati, E.; Benedetti, P.; Clementi, S. Flap: Grid Molecular Interaction Fields in Virtual Screening. Validation Using the DUD Data Set. *J. Chem. Inf. Model.* **2010**, *50*, 1442-1450.
38. Decherchi, S.; Berteotti, A.; Bottegoni, G.; Rocchia, W.; Cavalli, A. The Ligand Binding Mechanism to Purine Nucleoside Phosphorylase Elucidated Via Molecular Dynamics and Machine Learning. *Nat. Commun.* **2015**, *6*, 6155.

39. Ho, M.C.; Shi, W.; Rinaldo-Matthis, A.; Tyler, P.C.; Evans, G.B.; Clinch, K.; Almo, S.C.; Schramm, V.L. Four Generations of Transition-State Analogues for Human Purine Nucleoside Phosphorylase. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4805-4812.
40. Schramm, V.L. Enzymatic Transition States and Transition State Analogues. *Curr Opin. Struct. Biol.* **2005**, *15*, 604-613.
41. Giblett, E.R.; Ammann, A.J.; Wara, D.W.; Sandman, R.; Diamond, L.K. Nucleoside-Phosphorylase Deficiency in a Child with Severely Defective T-Cell Immunity and Normal B-Cell Immunity. *Lancet* **1975**, *1*, 1010-1013.
42. Ravandi, F.; Gandhi, V. Novel Purine Nucleoside Analogues for T-Cell-Lineage Acute Lymphoblastic Leukaemia and Lymphoma. *Expert. Opin. Investig. Drugs* **2006**, *15*, 1601-1613.
43. Robak, T.; Lech-Maranda, E.; Korycka, A.; Robak, E. Purine Nucleoside Analogs as Immunosuppressive and Antineoplastic Agents: Mechanism of Action and Clinical Activity. *Curr. Med. Chem.* **2006**, *13*, 3165-3189.
44. Clinch, K.; Evans, G.B.; Frohlich, R.F.; Furneaux, R.H.; Kelly, P.M.; Legentil, L.; Murkin, A.S.; Li, L.; Schramm, V.L.; Tyler, P.C.; Woolhouse, A.D. Third-Generation Immucillins: Syntheses and Bioactivities of Acyclic Immucillin Inhibitors of Human Purine Nucleoside Phosphorylase. *J. Med. Chem.* **2009**, *52*, 1126-1143.
45. Evans, G.B.; Furneaux, R.H.; Lewandowicz, A.; Schramm, V.L.; Tyler, P.C. Synthesis of Second-Generation Transition State Analogues of Human Purine Nucleoside Phosphorylase. *J. Med. Chem.* **2003**, *46*, 5271-5276.
46. Kline, P.C.; Schramm, V.L. Purine Nucleoside Phosphorylase. Catalytic Mechanism and Transition-State Analysis of the Arsenolysis Reaction. *Biochemistry* **1993**, *32*, 13212-13219.
47. Hirschi, J.S.; Arora, K.; Brooks, C.L., 3rd; Schramm, V.L. Conformational Dynamics in Human Purine Nucleoside Phosphorylase with Reactants and Transition-State Analogues. *J. Phys. Chem. B* **2010**, *114*, 16263-16272.
48. de Azevedo, W.F., Jr.; Canduri, F.; dos Santos, D.M.; Silva, R.G.; de Oliveira, J.S.; de Carvalho, L.P.; Basso, L.A.; Mendes, M.A.; Palma, M.S.; Santos, D.S. Crystal Structure of Human Purine Nucleoside Phosphorylase at 2.3 Å Resolution. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 545-552.
49. Stahl, M.; Mauser, H. Database Clustering with a Combination of Fingerprint and Maximum Common Substructure Methods. *J. Chem. Inf. Model.* **2005**, *45*, 542-548.
50. Canduri, F.; Silva, R.G.; dos Santos, D.M.; Palma, M.S.; Basso, L.A.; Santos, D.S.; de Azevedo, W.F., Jr. Structure of Human Pnp Complexed with Ligands. *Acta Crystallogr. D Biol. Crystallogr.* **2005**, *61*, 856-862.
51. dos Santos, D.M.; Canduri, F.; Pereira, J.H.; Vinicius Bertacine Dias, M.; Silva, R.G.; Mendes, M.A.; Palma, M.S.; Basso, L.A.; de Azevedo, W.F., Jr.; Santos, D.S. Crystal Structure of Human Purine Nucleoside Phosphorylase Complexed with Acyclovir. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 553-559.
52. Timmers, L.F.; Caceres, R.A.; Vivan, A.L.; Gava, L.M.; Dias, R.; Ducati, R.G.; Basso, L.A.; Santos, D.S.; de Azevedo, W.F., Jr. Structural Studies of Human Purine Nucleoside Phosphorylase: Towards a New Specific Empirical Scoring Function. *Arch. Biochem. Biophys.* **2008**, *479*, 28-38.
53. Bosshard, H.R. Molecular Recognition by Induced Fit: How Fit Is the Concept? *News Physiol. Sci.* **2001**, *16*, 171-173.
54. Beauchamp, L.M.; Tuttle, J.V.; Rodriguez, M.E.; Sznajdman, M.L. Guanine, Pyrazolo[3,4-D]Pyrimidine, and Triazolo[4,5-D]Pyrimidine (8-Azaguanine) Phosphonate Acyclic Derivatives as Inhibitors of Purine Nucleoside Phosphorylase. *J. Med. Chem.* **1996**, *39*, 949-956.
55. Ealick, S.E.; Babu, Y.S.; Bugg, C.E.; Erion, M.D.; Guida, W.C.; Montgomery, J.A.; Secrist, J.A., 3rd Application of Crystallographic and Modeling Methods in the Design of Purine Nucleoside Phosphorylase Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 11540-11544.

56. Tuttle, J.V.; Krenitsky, T.A. Effects of Acyclovir and Its Metabolites on Purine Nucleoside Phosphorylase. *J. Biol. Chem.* **1984**, *259*, 4065-4069.
57. D'Muniz Pereira, H.; Oliva, G.; Garratt, R.C. Purine Nucleoside Phosphorylase from *Schistosoma Mansoni* in Complex with Ribose-1-Phosphate. *J. Synchrotron Radiat.* **2011**, *18*, 62-65.
58. Luic, M.; Koellner, G.; Yokomatsu, T.; Shibuya, S.; Bzowska, A. Calf Spleen Purine-Nucleoside Phosphorylase: Crystal Structure of the Binary Complex with a Potent Multisubstrate Analogue Inhibitor. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1417-1424.
59. Hantschel, O.; Superti-Furga, G. Regulation of the C-Abl and Bcr-Abl Tyrosine Kinases. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 33-44.
60. Lovera, S.; Sutto, L.; Boubeva, R.; Scapozza, L.; Dolker, N.; Gervasio, F.L. The Different Flexibility of C-Src and C-Abl Kinases Regulates the Accessibility of a Druggable Inactive Conformation. *J. Am. Chem. Soc.* **2012**, *134*, 2496-2499.
61. Cowan-Jacob, S.W.; Fendrich, G.; Floersheimer, A.; Furet, P.; Liebetanz, J.; Rummel, G.; Rheinberger, P.; Centeleghe, M.; Fabbro, D.; Manley, P.W. Structural Biology Contributions to the Discovery of Drugs to Treat Chronic Myelogenous Leukaemia. *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63*, 80-93.
62. Nagar, B.; Bornmann, W.G.; Pellicena, P.; Schindler, T.; Veach, D.R.; Miller, W.T.; Clarkson, B.; Kuriyan, J. Crystal Structures of the Kinase Domain of C-Abl in Complex with the Small Molecule Inhibitors Pd173955 and Imatinib (Sti-571). *Cancer Res.* **2002**, *62*, 4236-4243.
63. Panjarian, S.; Iacob, R.E.; Chen, S.; Engen, J.R.; Smithgall, T.E. Structure and Dynamic Regulation of Abl Kinases. *J. Biol. Chem.* **2013**, *288*, 5443-5450.
64. Mollica, L.; Decherchi, S.; Zia, S.R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of Protein-Ligand Unbinding Via Smoothed Potential Molecular Dynamics Simulations. *Sci. Rep.* **2015**, *5*, 11539.
65. Cristalli, G.; Lambertucci, C.; Marucci, G.; Volpini, R.; Dal Ben, D. A2a Adenosine Receptor and Its Modulators: Overview on a Druggable GPCR and on Structure-Activity Relationship Analysis and Binding Requirements of Agonists and Antagonists. *Curr. Pharm. Des.* **2008**, *14*, 1525-1552.
66. Ng, H.W.; Laughton, C.A.; Doughty, S.W. Molecular Dynamics Simulations of the Adenosine A2a Receptor in POPC and POPG Lipid Bilayers: Effects of Membrane on Protein Behavior. *J. Chem. Inf. Model.* **2014**, *54*, 573-581.
67. Lai, P.C.; Crasto, C.J. Beyond Modeling: All-Atom Olfactory Receptor Model Simulations. *Front. Genet.* **2012**, *3*, 61.
68. Zare, B.; Madadkar-Sobhani, A.; Dastmalchi, S.; Mahmoudian, M. Prediction of the Human Ep1 Receptor Binding Site by Homology Modeling and Molecular Dynamics Simulation. *Sci. Pharm.* **2011**, *79*, 793-816.
69. Shan, J.; Khelashvili, G.; Mondal, S.; Mehler, E.L.; Weinstein, H. Ligand-Dependent Conformations and Dynamics of the Serotonin 5-HT_{2A} Receptor Determine Its Activation and Membrane-Driven Oligomerization Properties. *PLoS Comput. Biol.* **2012**, *8*, e1002473.
70. Congreve, M.; Andrews, S.P.; Dore, A.S.; Hollenstein, K.; Hurrell, E.; Langmead, C.J.; Mason, J.S.; Ng, I.W.; Tehan, B.; Zhukov, A.; Weir, M.; Marshall, F.H. Discovery of 1,2,4-Triazine Derivatives as Adenosine a(2a) Antagonists Using Structure Based Drug Design. *J. Med. Chem.* **2012**, *55*, 1898-1903.
71. Liu, W.; Chun, E.; Thompson, A.A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G.W.; Roth, C.B.; Heitman, L.H.; AP, I.J.; Cherezov, V.; Stevens, R.C. Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions. *Science* **2012**, *337*, 232-236.
72. Jaakola, V.P.; Griffith, M.T.; Hanson, M.A.; Cherezov, V.; Chien, E.Y.; Lane, J.R.; Ijzerman, A.P.; Stevens, R.C. The 2.6 Angstrom Crystal Structure of a Human A2a Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322*, 1211-1217.
73. Bortolato, A.; Tehan, B.G.; Bodnarchuk, M.S.; Essex, J.W.; Mason, J.S. Water Network Perturbation in Ligand Binding: Adenosine a(2a) Antagonists as a Case Study. *J. Chem. Inf. Model.* **2013**, *53*, 1700-1713.

74. Case, D.; Darden, T.; Cheatham, T.I.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; et al., *Amber 12*. University of California: San Francisco, CA, 2012.
75. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
76. Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
77. Kabsch, W. A Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallogr. A* **1976**, *A32*, 922-923.
78. Tribello, G.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. Plumed 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2013**, *185*, 604-613.
79. Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2007**, *2*, e880.
80. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. Charmm-Gui: A Web-Based Graphical User Interface for Charmm. *J. Comput. Chem.* **2008**, *29*, 1859-1865.
81. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with Namd. *J. Comput. Chem.* **2005**, *26*, 1781-1802.
82. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of Useful Decoys, Enhanced (Dud-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582-6594.
83. Huang, N.; Shoichet, B.K.; Irwin, J.J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789-6801.
84. Cross, S.; Ortuso, F.; Baroni, M.; Costa, G.; Distinto, S.; Moraca, F.; Alcaro, S.; Cruciani, G. Grid-Based Three-Dimensional Pharmacophores II: Pharmbench, a Benchmark Data Set for Evaluating Pharmacophore Elucidation Methods. *J. Chem. Inf. Model.* **2012**, *52*, 2599-2608.
85. Muratore, G.; Goracci, L.; Mercorelli, B.; Foeglein, A.; Digard, P.; Cruciani, G.; Palu, G.; Loregian, A. Small Molecule Inhibitors of Influenza A and B Viruses That Act by Disrupting Subunit Interactions of the Viral Polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 6247-6252.
86. Spyraakis, F.; Cellini, B.; Bruno, S.; Benedetti, P.; Carosati, E.; Cruciani, G.; Micheli, F.; Felici, A.; Cozzini, P.; Kellogg, G.E.; Voltattorni, C.B.; Mozzarelli, A. Targeting Cystalyisin, a Virulence Factor of *Treponema Denticola*-Supported Periodontitis. *ChemMedChem* **2014**, *9*, 1501-1511.
87. Spyraakis, F.; Felici, P.; Bayden, A.S.; Salsi, E.; Miggiano, R.; Kellogg, G.E.; Cozzini, P.; Cook, P.F.; Mozzarelli, A.; Campanini, B. Fine Tuning of the Active Site Modulates Specificity in the Interaction of O-Acetylserine Sulfhydrylase Isozymes with Serine Acetyltransferase. *Biochim. Biophys. Acta* **2013**, *1834*, 169-181.
88. Spyraakis, F.; Singh, R.; Cozzini, P.; Campanini, B.; Salsi, E.; Felici, P.; Raboni, S.; Benedetti, P.; Cruciani, G.; Kellogg, G.E.; Cook, P.F.; Mozzarelli, A. Isozyme-Specific Ligands for O-Acetylserine Sulfhydrylase, a Novel Antibiotic Target. *PLoS One* **2013**, *8*, e77558.
89. Carosati, E.; Sciabola, S.; Cruciani, G. Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the Grid Force Field. *J. Med. Chem.* **2004**, *47*, 5114-5125.
90. Goodford, P.J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849-857.
91. Siragusa, L.; Cross, S.; Baroni, M.; Goracci, L.; Cruciani, G. Biogps: Navigating Biological Space to Predict Polypharmacology, Off-Targeting, and Selectivity. *Proteins* **2015**, *83*, 517-532.
92. Trott, O.; Olson, A.J. Autodock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455-461.

93. Sastry, G.M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221-234.
94. Schrödinger Suite 2014-1 Protein Preparation Wizard; Epik Version 2.7, Schrödinger, Llc, New York, Ny, 2013; Impact Version 6.2, Schrödinger, Llc, New York, Ny, 2014; Prime Version 3.5, Schrödinger, Llc, New York, Ny, 2014.
95. Truchon, J.F.; Bayly, C.I. Evaluating Virtual Screening Methods: Good and Bad Metrics for The "Early Recognition" Problem. *J. Chem. Inf. Model.* **2007**, *47*, 488-508.

“For Table of Contents use only”

A pipeline to enhance ligand virtual screening: integrating Molecular Dynamics and FLAP

Francesca Spyraakis ^{1*}, Paolo Benedetti ², Sergio Decherchi ^{3,4}, Walter Rocchia ³,
Andrea Cavalli ^{5,6}, Stefano Alcaro ⁷, Francesco Ortuso ⁷, Massimo Baroni ⁸,
Gabriele Cruciani ^{2*}.

