J Comput Aided Mol Des (2014) 28:363–373
DOI 10.1007/s10822-013-9702-2

Combining in silico and in cerebro approaches for virtual screening and pose prediction in SAMPL4

Arnout R. D. Voet · Ashutosh Kumar · Francois Berenger · Kam Y. J. Zhang

Received: 8 November 2013 / Accepted: 17 December 2013 / Published online: 21 January 2014

Abstract The SAMPL challenges provide an ideal opportunity for unbiased evaluation and comparison of different approaches used in computational drug design. During the fourth round of this SAMPL challenge, we participated in the virtual screening and binding pose prediction on inhibitors targeting the HIV-1 integrase enzyme. For virtual screening, we used well known and widely used in silico methods combined with personal in cerebro insights and experience. Regular docking only performed slightly better than random selection, but the performance was significantly improved upon incorporation of additional filters based on pharmacophore queries and electrostatic similarities. The best performance was achieved when logical selection was added. For the pose prediction, we utilized a similar consensus approach that amalgamated the results of the Glide-XP docking with structural knowledge and rescoring. The pose prediction results revealed that docking displayed reasonable performance in predicting the binding poses. However, prediction performance can be improved utilizing scientific experience and rescoring approaches. In both the virtual screening and pose prediction challenges, the top performance was achieved by our approaches. Here we describe the methods and strategies used in our approaches and discuss the rationale of their performances.

Keywords Virtual screening · Molecular docking · Pose prediction · Pharmacophore modeling · Electrostatic similarity · EleKit · Computer aided drug design

Introduction

Computational approaches are an indispensable component of drug discovery complementing experimental approaches. Virtual screening, pose prediction and binding energy calculation are often regarded as three separate stages of lead discovery. Virtual screening, also known as in silico screening is a procedure to identify drug-like compounds for a given target using computational methods [1]. These computational methods typically include molecular docking simulations, pharmacophore-based screening, similarity searches and more complex procedures involving artificial intelligence [2]. Pose prediction refers to the simulation of how known active compounds bind to their protein target. Binding energy calculation involves the estimation of the free energy of binding of known actives at a given pose. However, these three stages are often intertwined. For example, the binding poses are predicted in order to identify active compounds in the virtual screening stage, whereas the binding free energies maybe calculated in order to predict the binding poses. With a variety of methods, software and user experiences of computational approaches, their performances are influenced by many variables and their outcome may be dramatically different. Moreover, the error in the computational approaches used in drug discovery tends to be larger than experimental methods in general. Retrospective evaluation of computational methods sometimes does not translate into actual performance in real-life use cases. Prospective

Arnout R. D. Voet and Ashutosh Kumar have contributed equally to this work.

A. R. D. Voet · A. Kumar · F. Berenger · K. Y. J. Zhang (✉)
Zhang Initiative Research Unit, Institute Laboratories, RIKEN, 2-1 Hirosawa, Wakō, Saitama 351-0198, Japan
e-mail: kamzhang@riken.jp

assessment of these methods using blinded experimental data is required to evaluate their accuracy and to improve their reliability in the future.

The SAMPL challenges provide an important platform to evaluate various in silico methods [3–6]. The fourth round of this SAMPL challenge (SAMPL4) selected the HIV-1 integrase (HIV-1 IN) protein as the target for the virtual screening, pose prediction and binding free energy calculation categories. HIV-1 IN is one of the three HIV enzymes and is responsible for the integration of the viral cDNA into the host cell genome. It consists of three domains of which the catalytic core domain (CCD) is the most interesting for drug development [7]. To carry out its role, HIV-1 IN carries out two consecutive enzymatic steps (3′ processing and strand transfer) and participates in a number of interactions of interest as drug target. The first approved integrase strand transfer inhibitors (INSTIs) target the enzymatic strand transfer binding site, but emerging resistance has driven the research field to focus on novel binding sites [8]. One of the alternative strategies targeting HIV-1 IN is by interfering with the binding site of a human cofactor, LEDGF/p75, which is an essential PPI interaction to complete the viral lifecycle [9, 10]. In recent years, the first inhibitors targeting this interaction (LEDGINs) have been reported [11–13]. In one of the reports, the researchers followed a fragment-based drug design approach [14] where small molecule fragments were first identified and later grown into full inhibitors [8]. Next to the discovery of inhibitors targeting the LEDGF/p75 binding site, inhibitors bound to allosteric sites were also identified [8, 15]. The SAMPL4 challenge builds on this report: novel, previously unreleased structures and activities of compounds were made available to the organizers of SAMPL4 [16].

For SAMPL4, participants could compete in three different categories employing the HIV-1 IN test case. For the virtual screening category, participants were given a list of molecules containing experimentally validated active and inactive compounds. The binding site (and mechanism of action) was not revealed and could be anywhere on the HIV-1 IN CCD protein surface. The second category involved binding mode prediction, and participants were given a list of active compounds. Again the exact binding site was not revealed, but three possible binding sites were suggested. For the third assessment, participants had to predict the affinity of the active compounds for the target.

Our team participated in the virtual screening and the binding mode prediction challenges. For both experiments we were ranked first amongst all participants. In this manuscript we give an overview of our methods and discuss the rationale and performance.

Materials and methods

Virtual screening methods

The first step in the virtual screening experiments was to collect all structures of the HIV-1 IN with inhibitors bound to the LEDGF/p75 binding site (PDB entries 2b4j, 3lpu, 3lpt, 3nf6, 3nf7, 3nf8, 3nf9, 3zsq, 3zsr, 3zsv, 3zsw, 3zsx, 3zsy, 3zsz, 3zt0, 3zt1, 3zt2, 3zt13, 3zt4, 4dmn, 4e1m and 4e1n). Since the inhibitory mechanisms of the other binding sites reported by Peat et al. [8] remains uncertain, our virtual screening only focused on the LEDGIN site [15].

All structures were superimposed on the crystal structure of the HIV-1 IN CCD in complex with the integrase binding domain (IBD) of LEDGF/p75 [17]. Using LigandScout [18], the common interactions of the inhibitors were merged into a common feature pharmacophore query (Fig. 1). This simple query was used to screen a conformational database of 321 small molecules from SAMPL4. The small molecule conformational database was generated using the LigandScout implementation of the Omega algorithm [19–21]. This filtered out 32 compounds. In our final ranking of compounds, these compounds were ranked last with the highest confidence of inactivity.

Further attempts to use pharmacophore queries to identify molecules that were closer towards the previously reported inhibitors, and thereby removing more false positives, failed [8, 15]. More complex pharmacophore queries were not able to identify more than 10 potential hit molecules. However, it is clear from visual inspection that many compounds from the SAMPL4 virtual screening set are in fact close homologues of active compounds and should have hit the pharmacophore query in the correct conformation. We reasoned that the conformation generation tool failed to generate conformations close enough to the bio-active conformation.

The next step in the virtual screening protocol was to employ molecular docking simulations. For the virtual screening challenge, we opted for GOLD as it was successfully applied in the identification of the first LEDGINs, and the predicted binding modes were validated using crystallography [13, 22]. Docking was performed using GOLD on the receptor protein with a maximum of 10 poses per compound restrained to the CCD at the IBD interface. PLP was chosen as the scoring function [23].

Following the docking simulations, the results were post-filtered using the pharmacophore query to remove compounds with incompatible binding modes. MOE was used as the screening tool to accomplish this, in "use absolute positions" mode and with partial matching disallowed [24]. In the final ranking of the compounds, these discarded compounds were ranked just before the

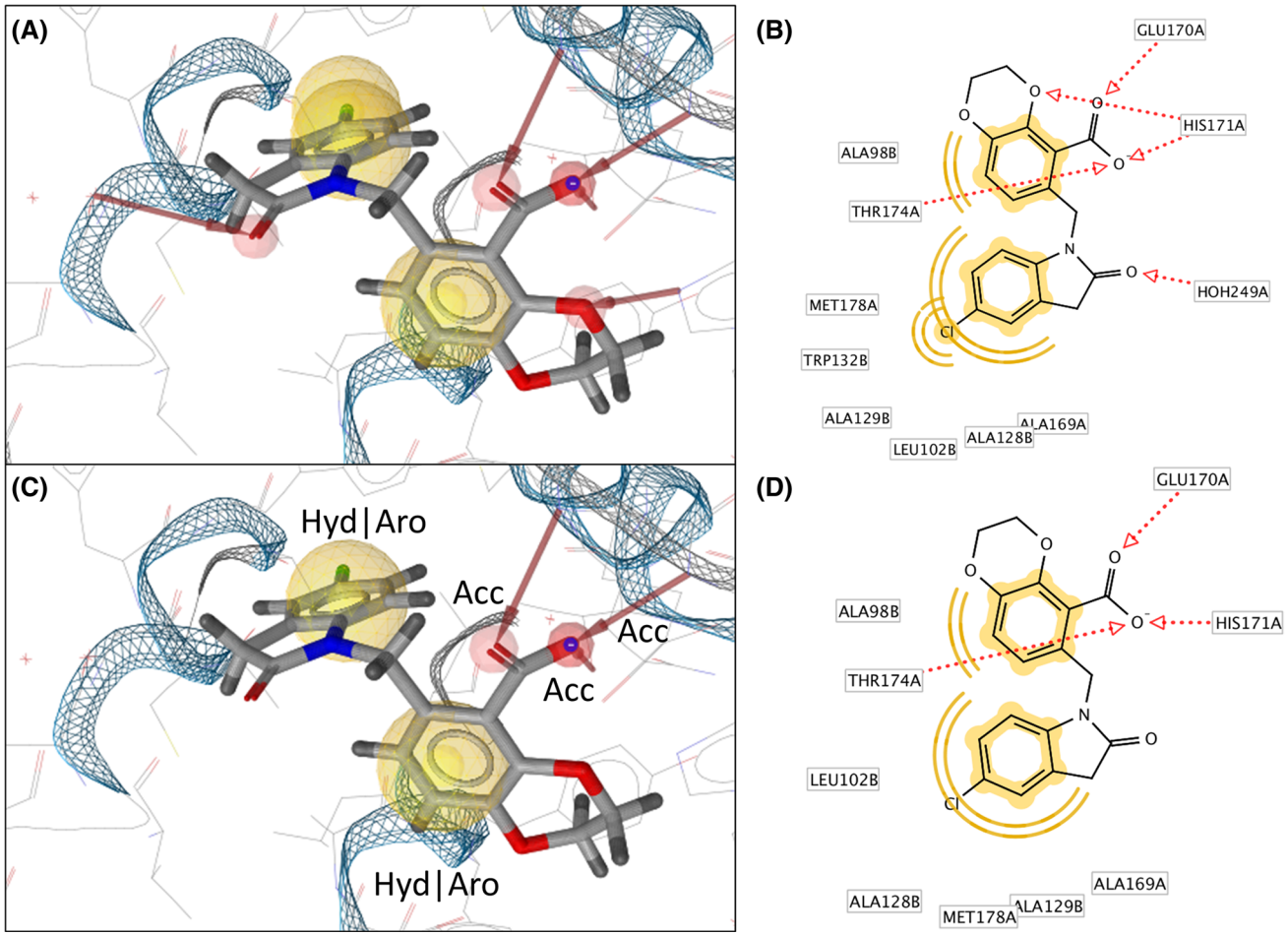

Fig. 1 Pharmacophore query employed during the virtual screen. LigandScout was used to create a receptor based pharmacophore of PDB 3nf8 which was designated as the receptor with the SAMPL4 virtual screening dataset (*panels* **a** and **b**: 3D and 2D representation). The pharmacophore query was also created for all other complex structures with molecules binding to the same binding pocket. Only the key interactions conserved in all inhibitors were retained in the final pharmacophore query that was used at step one in the virtual screen and post filtering of the docking results (*panels* **c** and **d**: 3D and 2D representation). The final query consisted out of 3 "hydrogen bond acceptor" features on the carboxylate function and 2 "hydrophobic or aromatic" features

previously discarded (not fulfilling the pharmacophore query) compounds according to their respective PLP docking scores.

Subsequently, we assessed the electrostatic potential field similarity of the inhibitory compounds with the LEDGF/p75 IBD protein ligand using EleKit [25], removing compounds that were in dissimilar binding modes. EleKit calculates the similarity of the Poisson–Boltzmann electrostatic potential at the interface of the interaction using a Spearman rank correlation coefficient (ρ). It was observed that existing small molecule protein–protein interaction inhibitors (SMPPIIs) have similar potential fields compared to the protein–ligands they prevent to bind. For the electrostatic similarity with the IBD, PDB entry 2b4j was superposed on entry 3nf8 and the default parameters of EleKit was used [25]. The cut-off similarity value was $\rho > 0.2$. In the final ranking of the compounds, these discarded compounds were combined together with the compounds that did not pass the pharmacophore based docking post-filter. They were sorted according to their respective PLP docking scores and ranked before the compounds discarded by the pharmacophore-based initial filter.

For the final ranking, all 289 compounds that passed the initial pharmacophore query, we inspected and analyzed in cerebro the detailed binding modes and chemical structures in correspondence with the previously reported inhibitors. When different stereo-isomers were still present in the dataset, all compounds with different stereo-chemistry compared to previously reported inhibitors were moved backwards in the rank list. For the remaining compounds, a visual inspection similarity approach was used to identify the molecules having chemical functionalities that are bio-isosteres or matched molecular pairs with the previously

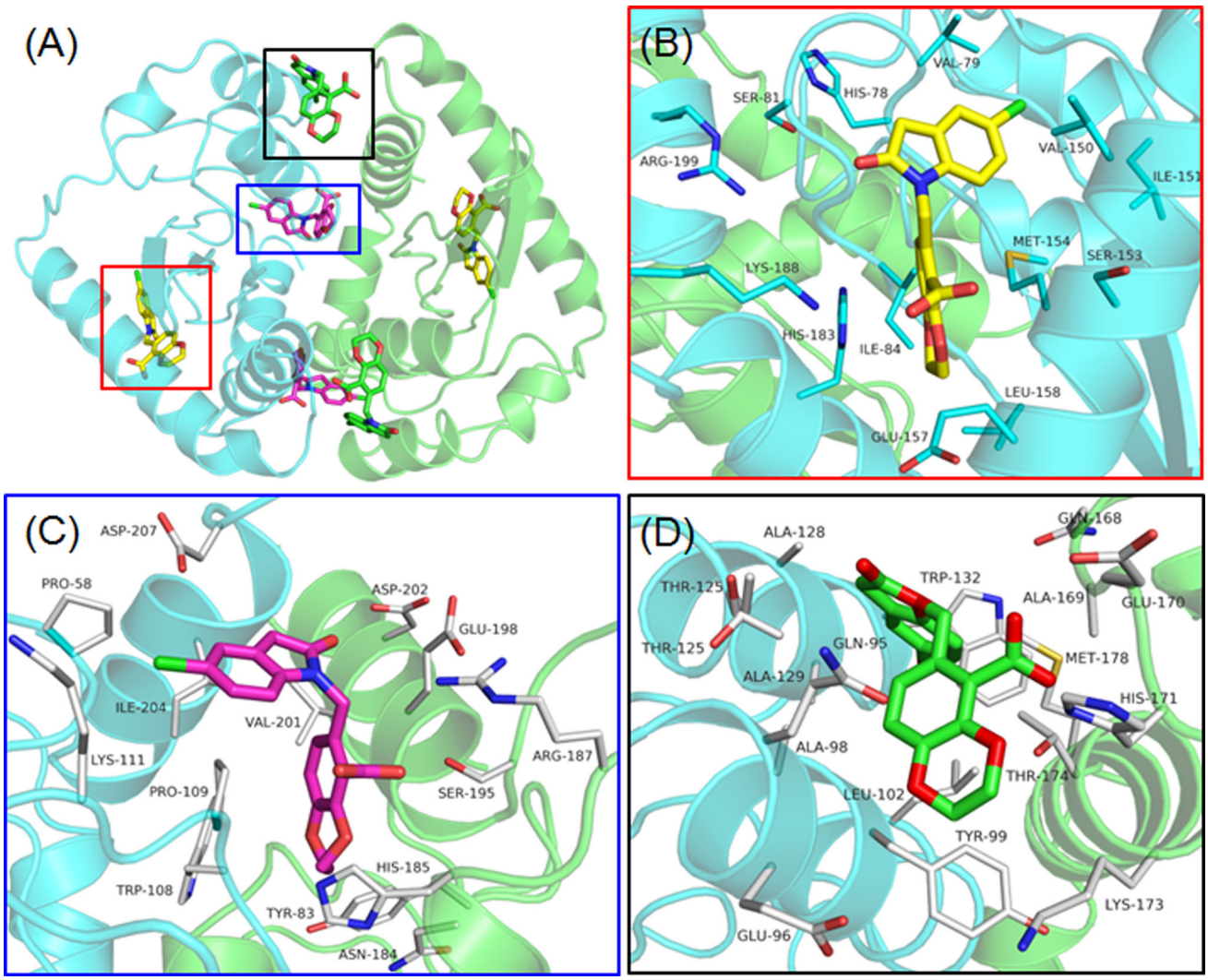

Fig. 2 Overview of the possible binding sites in the HIV-1 IN CCD dimer. **a** The crystal structure of the HIV-1 IN CCD dimer (PDB 3nf8) represented by green and blue cartoons, bound with the SAMPL4 pose prediction challenge reference ligand CDQ at the allosteric site (**b**), the fragment site (**c**) and the LEDGF/p75 site (**d**)

reported inhibitors or amino acid side chains observed at the PPI interface.

Pose prediction methods

The binding poses of ligands in the SAMPL4 challenge were predicted using the molecular docking approach as implemented in the Glide program [26–29]. Glide was used instead of GOLD for docking due to the personal experience of the researcher performing the pose prediction experiments. There are six ligand binding sites on a HIV-1 IN dimer and those sites form three equivalent pairs. To predict the binding poses, all SAMPL4 challenge compounds were docked to the three HIV-1 IN CCD ligand binding sites on one HIV-1 IN monomer as suggested by the SAMPL4 organizers, (1) Site1: the allosteric site for CDQ257 binding in PDB code 3nf8, (2) Site2: the fragment site for CDQ277 binding in PDB code 3nf8 and (3) Site3: the LEDGF/p75 site for CDQ267 binding in PDB code 3nf8 (Fig. 2). The preparation of the protein structures for the docking simulations was accomplished by using the protein preparation utility of Maestro (Schrödinger) [30]. The HIV-1 IN CCD crystal structure (3nf8) was used for the molecular docking [15]. To prepare the receptor structures, hydrogen atoms were added, bond orders were assigned and all the water molecules except those within 4.5 Å of bound ligand (CDQ257, CDQ277 and CDQ267 for allosteric, fragment and LEDGF/p75 site respectively) were removed. The protonation state of the charged residues was determined using Maestro. The SAMPL4 challenge compounds were prepared for docking using LigPrep [31] by adding hydrogen atoms and generating ionization

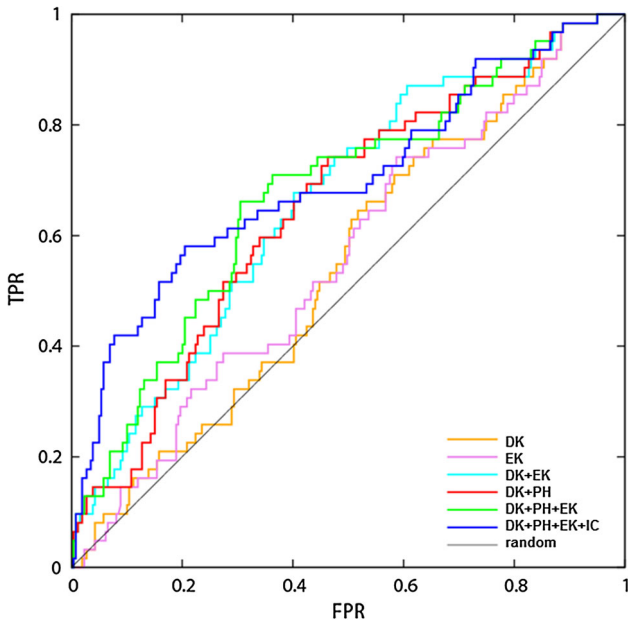

Fig. 3 ROC plot for the individual and combined steps of our virtual screen. DK represents the performance of the PLP score during the docking simulations. EK represents the electrostatic similarity with the IBD as calculated with EleKit based on the docked poses. DK + EK represents the performance of the docking simulations after post-filtering the results based on the electrostatic similarity. DK + PH represents the performance of the docking simulations after post-filtering the results based on the pharmacophore query. DK + PH + EK represents the performance of the docking simulations after post-filtering the results based on the pharmacophore query and the electrostatic similarity. DK + PH + EK + IC represents the performance of the docking simulations after post-filtering the results based on the pharmacophore query and the electrostatic similarity and reordering the ranking based on in cerebro assessment of the results. For more details see Table 1

states. All structures were subjected to minimization using the OPLS-2005 force field [32]. All of the molecular docking calculations were performed using Glide Version 5.7 in extra precision mode (Glide-XP) [26–29]. The grids for molecular docking simulations were generated by using sites defined by the centroid of the bound ligand (CDQ257, CDQ277 and CDQ267 for allosteric, fragment and LEDGF/p75 site respectively). The docking results were analyzed using Glide-XP visualizer. The rescoring was performed using Prime MM-GBSA [33–35] and DrugScore [36]. The graphics for pose prediction challenge were prepared using PyMOL Version 1.4 [37] and Gnuplot Version 4.6 [38].

Hardware

This work was carried out using Dell Precision T5400 workstations with 2.0 GHz Intel Xeon CPU and a 97.4TFLOPS Intel Xeon 5570 based massively parallel PC cluster of RIKEN integrated cluster of clusters (RICC).

Table 1 The performance of the individual and a combination of several virtual screening methods

Methods	AUC	ER (10 %)
DK	0.546	1
EK	0.558	1.2
DK + PH	0.652	1.5
DK + EK	0.659	1.8
DK + PH + EK	0.676	2.2
DK + PH + EK + IC	0.699	3

For a graphical representation see Fig. 2

DK docking, *PH* pharmacophores, *EK* EleKit, *IC* in cerebro

Results and discussion

Analysis of the virtual screening performance

Our combined screening approach uses different methodologies. To assess the individual contribution of each method, we retrospectively analyzed the different results. Figure 3 shows the receiver operating characteristic (ROC) curves and Table 1 lists the performance and enrichment data [area under the curve value (AUC) on the ROC-plot and enrichment factor (ER)].

The first step involved the removal of 32 compounds, since they could not fulfill the pharmacophore query. Of these compounds, two compounds were active (pC2A03 and AVX17631). In case of a random selection of 32 compounds, it would contain six actives indicating that this pharmacophore query is efficient at removing inactive compounds. Furthermore, we only focused on the IBD binding site in the HIV-1 IN, and those compounds are not bound in this site. We believe that this simple pharmacophore query was not able to remove more compounds, especially false positives, because of the similarity of the active with the inactive compounds including the high number of possible stereoisomers. Since our pharmacophore filter was based on a simple "yes" or "no" criterion, ranking is impossible and ROC, AUC and ER data are not available.

Since it is commonly observed that SMPPIIs mimic the key interactions at the interface, pharmacophore features can be derived from that interface [39–41]. In general all compounds reported here fulfill the pharmacophore query that describes the key interactions formed by the natural partner proteins, further demonstrating the usefulness of pharmacophores in SMPPII discovery.

The second step of our virtual screening experiment was to dock all the compounds into the receptor structure provided by the SAMPL4 organizers. Analyzing the performance of the PLP scoring function for docking the compounds in the receptor with GOLD shows us that we

reach an AUC of 0.546 and an ER of 1 at 10 % of the ranking which is only slightly better than a random selection of compounds on AUC and an early enrichment equal to random. These values are, however, in good agreement with the performance of the other submissions in this challenge that employed docking simulations as well [42].

For the third step in our virtual screening protocol, we post-filtered the docking results based on the pharmacophore query. All 10 different poses were kept and analyzed during all post-filtering steps, in contrast to the general practice of rejecting poses with poorer scores. Compounds that did not fulfill the query were re-ranked at the end of the list according to their docking scores. At this step we significantly improved our performance and reached an AUC of 0.652 with an ER of 1.5. These values are already significantly better than the majority of the other submissions in this challenge. A fundamental conclusion from our analysis is that the majority of the computational docking poses are not in full agreement with the key interactions required for optimal binding, and these compounds are therefore unlikely to bind the target. Our work highlights the usefulness of pharmacophore post-filtering the docking results to identify compounds able to form similar interactions as known active compounds with the target. Significant enrichment of the docking results can be achieved by discarding compounds that do not match the pharmacophore.

In the fourth step we used electrostatic similarity to the protein ligand as a post-filter of the docked solutions. When analyzing the similarity metric alone as a scoring function we observe an AUC of 0.558 and an ER of 1.2, which is only slightly better than random or the docking scoring function. When analyzing the results of the electrostatic similarity metric as a post-filter directly on the docking data, by ranking the docked compounds that are not similar to the protein–ligand at the end of the list, the performance is highly improved with an AUC of 0.659 and an ER of 1.8. This indicates that SMPPIIs mimic the interactions at the PPI interface and computational drug design of SMPPIIs would benefit from incorporating this mimicry effect. In our virtual screening setup, however, we applied the electrostatic similarity filter after the pharmacophore filter and the combined performance is only marginal improved compared to the individual performances (AUC: 0.676, ER: 2.2).

The last step of our approach was the personal in cerebro assessment of the ranking using common chemical sense and experience with this target. As many molecules were present as different stereoisomers, we analyzed all available crystal structures and only considered the stereoisomer to be active if a similar stereochemistry was found in one of the existing crystal structures on which our pharmacophore query was built initially. This human operator based reordering boosted the performance of the virtual screen to an AUC of 0.699 and an ER of 3, which makes it (SAMPL4_ID 164) the top performer among all the submissions participated in HIV-1 IN virtual screening (Fig. 4). Our work underlines the importance of experience in drug discovery, which has yet to be captured by any combination of algorithms. While the AUC does not significantly improve at this step of our procedure, the early enrichment is substantially better, demonstrating that AUC alone is not an optimal parameter to assess the performance of virtual screening; only the top-ranked molecules are evaluated in experimental situations. In fact, in cerebro selections are frequently performed after in silico screening experiments [43, 44], but this is the first time that the contribution of human experience to improvement of virtual screening analysis has been demonstrated in an unbiased blind challenge.

Pose prediction challenge overview and results

The aim of the SAMPL4 pose prediction challenge was to evaluate the ability of methods to predict accurately the binding pose of HIV-1 IN binding molecules. For this purpose, a set of 58 previously unpublished HIV-1 IN binders developed by CSIRO and Avexa Ltd. [16] were released to the SAMPL4 participants. These compounds were optimized from fragment screening hits and are known to bind to one or more sites in HIV-1 IN CCD.

To predict the poses of the 58 SAMPL4 ligands, we employed a strategy that amalgamated the state-of-art docking tool Glide-XP [26–29] with the knowledge of interaction requirements for HIV-1 IN inhibition. Furthermore, force-field based and knowledge-based scoring functions, Prime MM-GBSA score [33–35] and DrugScore [36], were also used during the binding prediction. Our pose prediction protocol was based on the assumption that a ligand should have the highest affinity for its real binding site when docked to all the three binding sites. Therefore, Glide-XP docking score was the primary component during the pose prediction by docking. Additionally, to strengthen confidence in selected poses, Glide-XP generated poses were rescored using Prime MM-GBSA score and DrugScore.

Three sets of docking simulations were performed focusing on the allosteric site, the fragment site and the LEDGF/p75 binding site (Fig. 2). The ligands and receptor structure for docking were prepared as described above. Although some of the SAMPL4 ligands were known to display multiple binding modes, only a single binding mode for each compound was predicted due to requirement of binding site specification for molecular docking. The site with the highest Glide-XP docking score for a particular ligand was regarded as its binding site and the best ranked

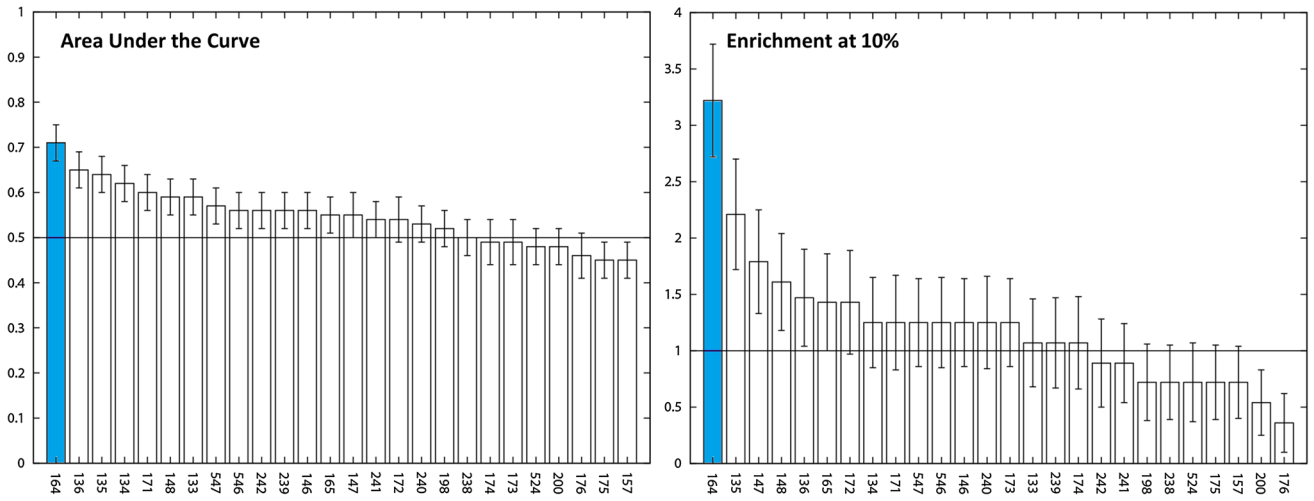

Fig. 4 Our performance and enrichment in the SAMPL4 virtual screening competition. Our submission (SAMPL4_ID 164) is represented in *blue* while the other participants are represented in *white bars*. On the *left*, the AUC of the ROC plot is given, while on the *right* the enrichment at 10 % of the ranking is given. The *horizontal line* indicates the random level. According to both metrics our approach was superior

pose for that particular ligand was selected. A threshold score difference of 1 kcal/mol was chosen, so that docking poses were only selected if the Glide-XP score of the best pose differed from poses at the other two sites by at least 1 kcal/mol. This threshold was chosen based on the docking scores obtained for all ligands at all three binding sites. Structural analysis of the LEDGIN type inhibitors had previously revealed the importance of the carboxylate functionality [8, 13, 15]. In fact this moiety overlaid well when the ligand-bound crystal structures of HIV-1 IN were superimposed. Therefore, the correct placement of this moiety with respect to its position in the reference crystal structure (PDB code 3nf8) was important in the binding of these known ligands. We could use this structural knowledge in SAMPL4 pose prediction challenge, since the reference crystal structure used for docking contains ligands with acid functionalities bound at all the three sites. This structure-based insight was used when Glide-XP score was unable to discriminate between binding poses for three binding sites. A root mean square deviation (RMSD) value of 1 Å between the acid moieties in SAMPL4 ligand and the corresponding crystal structure ligand was used as a cut-off. Moreover, docking poses generated by Glide-XP were rescored by using two different scoring functions: Prime MM-GBSA score [33–35] and DrugScore [36]. These scores were used when the usage of Glide-XP score and placement of acid functionality was not sufficient to discriminate the poses at three binding sites. Utilizing the above mentioned docking strategy the binding poses for 58 SAMPL4 challenge ligands were selected and submitted for evaluations.

In the SAMPL4 challenge, various methodologies were evaluated for identifying the correct binding site for the 55 SAMPL4 ligands (three ligands were excluded from analysis) as well as the accuracy of binding pose with respect to crystal structure conformation [42]. The SAMPL4 organizers used RMSD, AUC, RMS Tanimoto and binding fingerprints Tanimoto to assess the results from various submissions. The visual analysis of ligand bound crystal structures obtained from organizers revealed that 48 out of 55 ligands bind to the LEDGF/p75 site only. Three compounds displayed multiple binding modes and bound to the allosteric site (AVX101140 and AVX17257) and the fragment site (AVX17679) in addition to the LEDGF/p75 site. AVX15988, AVX17389 and AVX17631 bound only to the fragment site while ligand pC2-A03 was observed to bind at another allosteric site (site for CDQ247 in PDB code 3nf8) in HIV-1 IN dimer. Our docking strategy performed consistently well among different metrics used for analysis [42]. Our own analysis also suggests slightly better performance of our approach over other submissions.

To analyze the performance of our docking strategy in SAMPL4 pose prediction challenge, we first checked for differences in the affinities of SAMPL4 ligands predicted by Glide-XP scoring function for three binding sites of HIV-1 IN. The comparison revealed that only 52.7 % ligands displayed a clear preference for the LEDGF/p75 site over the allosteric and fragment sites, as judged from the Glide-XP score alone (Fig. 5). In the other ligands, either the Glide-XP scores were too similar (within 1 kcal/mol) to discriminate between the three binding sites, or an incorrect binding site scored best (AVX17560, AVX38781). Among the ligands displaying multiple binding modes, docking of AVX101140 resulted in very similar Glide-XP scores for the allosteric site and the LEDGF/p75 site (−6.37 and −6.27 for allosteric site and

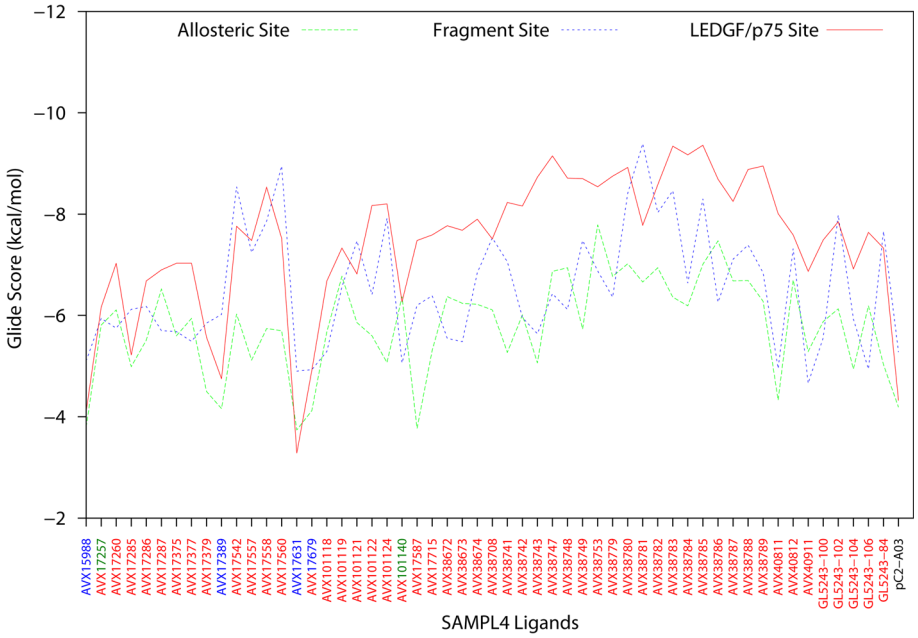

Fig. 5 Comparison of the Glide-XP docking scores. The *graph* shows the Glide-XP scores obtained for the SAMPL4 pose prediction challenge ligands for the allosteric site (*green*), the fragment site (*blue*) and the LEDGF/p75 site (*red*). The *color* of *X* axis label indicates the location of ligands observed in crystal structures. Mixed *color* labels show multiple binding modes while the label in *black color* shows no binding at site considered for docking

LEDGF/p75 site, respectively). Docking scores of −4.93 and −4.92 were obtained for the fragment site and LEDGF/p75 site respectively with AVX17679. This shows that the Glide-XP score alone is insufficient to discriminate between the binding poses of the SAMPL4 ligands. The confidence in the prediction can be improved by incorporating protein–ligand interaction knowledge. Further, scoring functions such as Prime MM-GBSA and DrugScore were also helpful in building confidence for selecting the binding poses.

The SAMPL4 organizers have made 55 ligand bound crystal structures available after the submission. These crystal structures were then used to evaluate the performance of our docking strategy in picking up the correct binding site as well as pose accuracy. The performance was measured by the RMSD of ligand docking poses with the corresponding crystal structure. The heavy atoms of the ligands were used to calculate the RMSD values using the *rmsd.py* script from Schrodinger [30]. Figure 6 reports the pose recovery expressed in terms of the percentage of ligands docked into the correct binding site. It shows the performance of our method that combines Glide-XP with rescoring (SAMPL4_ID 177). For comparison, the result obtained from using Glide-XP alone (represented by a pseudo ID 177*) in our submission is also plotted to demonstrate the contribution of rescoring in our strategy: (1) Glide-XP (177*): Only the docking scores were used to pick the correct binding site and ligand binding poses. This was performed in a completely automated way and no other information was used; (2) Glide-XP + Rescoring (SAMPL4_ID 177): Other information such as protein ligand interaction requirements, scores from Prime MM-GBSA and DrugScore were used to select binding poses. As shown in Fig. 6, Glide-XP alone performed well as correct site was picked for 85.45 % of the ligands. However, there were a few cases where Glide-XP scores alone discriminated poorly and additional knowledge helped to pick right binding site. The rescoring of Glide-XP generated poses using the interaction information and other scoring functions slightly improved the pose prediction as the correct site was picked for 89.09 % of the ligands while missing only 10.91 % of the ligands (Fig. 6). Out of the six

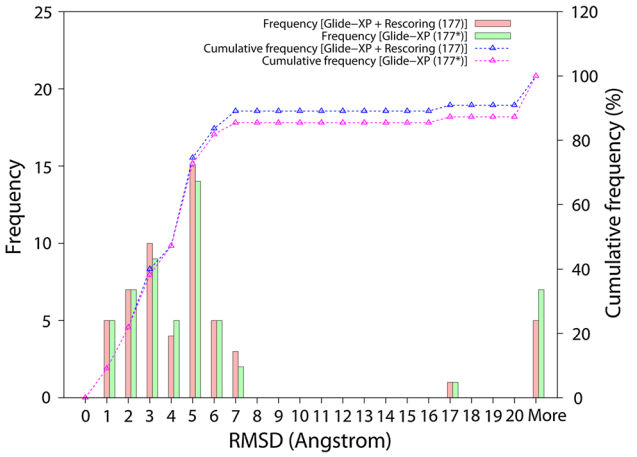

Fig. 6 Frequency histogram and cumulative frequency of docking pose recovery using Glide-XP docking and rescoring approach. SAMPL4 ID of our submission is 177 which represents "Glide-XP with a rescoring" approach. For comparison, the RMSDs obtained by only using Glide-XP scores for prediction were *plotted* as 177*. The performance improvement of 177 over 177* indicates the contribution of the rescoring strategy

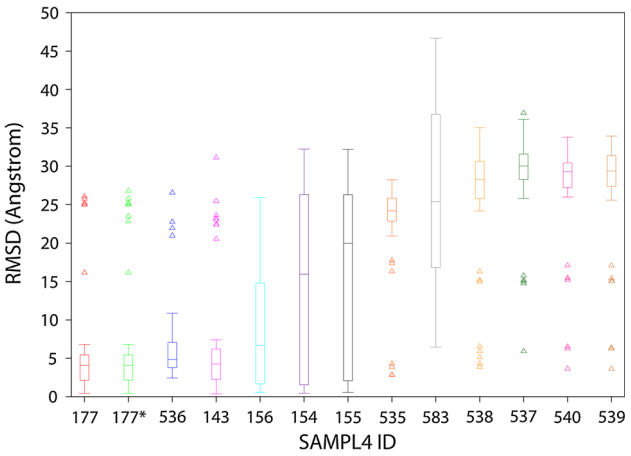

Fig. 7 Comparison of the results of various submissions for SAMPL4 pose prediction challenge utilizing RMSD. SAMPL4 ID of our submission is 177 which represents "Glide-XP with a rescoring" approach. For comparison, the RMSDs obtained by only using Glide-XP scores for prediction were *plotted* as 177*

ligands (AVX17379, AVX17542, AVX17560, GL5243-84, GL5243-102 and pC2-A03) where our docking strategy failed to pick the correct site, pC2-A03 does not bind to any of the three sites analyzed.

The SAMPL4 organizers also shared data from various participants of the pose prediction challenge to facilitate comparison between submissions. Our docking strategy (SAMPL4_ID 177) performed fairly well when compared with other submissions (Fig. 7). Other comparable predictions were observed from SAMPL4_ID 143 and SAMPL4_ID 536 (Fig. 7). SAMPL4_ID 143 used Autodock-Vina coupled with visual inspection [45] to predict binding poses while DOCK3.7 automated docking [46] was utilized by SAMPL4_ID 536. Although a reasonable docking performance was achieved when only the Glide-XP scores (177*) were used for predicting the binding poses, the prediction was improved when Glide-XP scores were complemented with other information such as interaction requirement, rescoring using Prime MM-GBSA and DrugScore (SAMPL4_ID 177). The RMSDs for SAMPL4_ID 177 span from 0.45 to 6.8 for 89.09 % of the ligands. A very high RMSD was observed for six compounds (depicted as outliers in Fig. 7) where our approach failed to pick the correct binding site. Around 21 % of the ligands displayed a RMSD of <2 Å while a RMSD of <3 Å was observed for 40 % of the ligands. Although achieving a success rate of 40 % (when using 3 Å RMSD as the success criterion) is a reasonable performance for our protocol, it is far from ideal. For the ligands where the correct binding site was picked, our docking approach fails to predict accurate poses for 27 ligands. The RMSD for these failure cases ranged from 3.56 to 6.8 Å. Overlay of the predicted poses of a few failure cases with the crystal structure revealed that the docking simulation could accurately predict the position of the acid group containing a benzodiazole or a benzodioxin ring with an RMSD of <1 Å for most of the ligands (Fig. 8). However, the placement of the functional groups adjoining a central secondary and tertiary amine was badly predicted (Fig. 8). This largely contributed to the higher RMSD for 21 ligands (AVX-17558, AVX38672-AVX38674, AVX38708, AVX 38753, AVX38779-AVX38789 and AVX101118-AVX 101122). Some of these compounds are shown in Fig. 8.

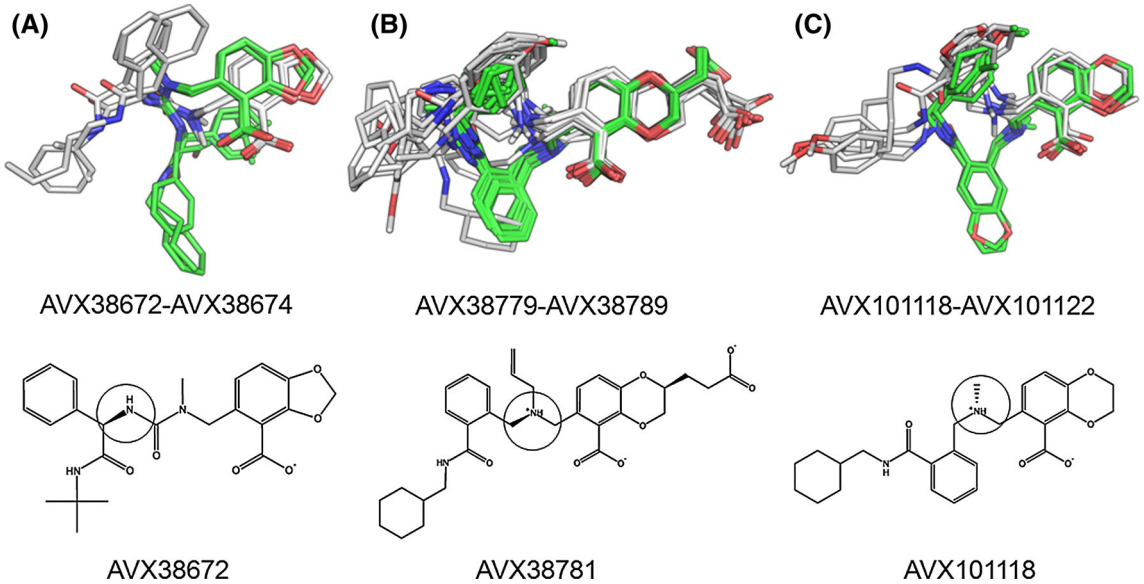

Fig. 8 Overlay of predicted poses of a few failure cases with crystal structure conformations. For clarity structurally similar compounds are shown separately as **a**, **b** and **c**. A representative molecule from each class is shown below highlighting the nitrogen stereo center

These compounds contain a nitrogen stereo center that is difficult to predict by the current docking algorithms. These compounds require special treatment, which we missed in the SAMPL4 pose prediction challenge. We have extensively enumerated all possible combinations of stereoisomers, which may not be a good idea to predict the poses of these ligands. In such cases, restricting the conformational search by providing bias towards absolute stereochemistry will avoid oversampling and may help improve the docking results.

Conclusion

For our virtual screening results, the success can largely be attributed to the use of appropriate filters on the results of the docking simulations. With significant understanding of active molecules it makes good sense to exploit this information using pharmacophore post-filters or protein ligand interaction fingerprints. Furthermore, since SMPPIIs often mimic the interactions at the PPI interface, this mimicry can be used as a post-filter as well. Therefore, virtual screening should rely not only on docking scores and rescoring, but also on all available experimental data and scientific experience. For the pose prediction challenge, the docking approach achieved reasonable performance in predicting the poses when the correct binding site is known. However, compounds with a stereo center, especially a tertiary amine are difficult to predict. The performance of the docking based pose prediction can be improved further by utilizing knowledge-based constraints and results from different, complementary scoring functions. Furthermore, docking simulation methods with special consideration for stereochemistry are clearly desirable.

Acknowledgments We thank RIKEN Integrated Cluster of Clusters (RICC) at RIKEN for the supercomputing resources used for this study. We are grateful to Dr. David Mobley and SAMPL4 participants for sharing the SAMPL4 challenge analyses and prediction data prior to publication. We acknowledge the Initiative Research Unit program from RIKEN, Japan for funding. AV acknowledges the JSPS for a postdoctoral fellowship. Prof. Jeremy Tame is acknowledged for proofreading the manuscript.

References

1. Heikamp K, Bajorath J (2013) The future of virtual compound screening. Chem Biol Drug Des 81(1):33–40
2. Lill M (2013) Virtual screening in drug design. Methods Mol Biol 993:1–12
3. Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) The SAMPL2 blind prediction challenge: introduction and overview. J Comput Aided Mol Des 24(4):259–279
4. Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. J Phys Chem B 113(14):4501–4507
5. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS (2008) Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. J Med Chem 51(4):769–779
6. Skillman AG (2012) SAMPL3: blinded prediction of host-guest binding affinities, hydration free energies, and trypsin inhibitors. J Comput Aided Mol Des 26(5):473–474
7. Voet AR, Maeyer MD, Debyser Z, Christ F (2009) In search of second-generation HIV integrase inhibitors: targeting integration beyond strand transfer. Future Med Chem 1(7):1259–1274
8. Peat TS, Rhodes DI, Vandegraaff N, Le G, Smith JA, Clark LJ, Jones ED, Coates JA, Thienthong N, Newman J, Dolezal O, Mulder R, Ryan JH, Savage GP, Francis CL, Deadman JJ (2012) Small molecule inhibitors of the LEDGF site of human immunodeficiency virus integrase identified by fragment screening and structure based design. PLoS ONE 7(7):e40147
9. Busschots K, Voet A, De Maeyer M, Rain JC, Emiliani S, Benarous R, Desender L, Debyser Z, Christ F (2007) Identification of the LEDGF/p75 binding site in HIV-1 integrase. J Mol Biol 365(5):1480–1492
10. Hombrouck A, De Rijck J, Hendrix J, Vandekerckhove L, Voet A, De Maeyer M, Witvrouw M, Engelborghs Y, Christ F, Gijsbers R, Debyser Z (2007) Virus evolution reveals an exclusive role for LEDGF/p75 in chromosomal tethering of HIV. PLoS Pathog 3(3):e47
11. Cavalluzzo C, Christ F, Voet A, Sharma A, Singh BK, Zhang KY, Lescrinier E, De Maeyer M, Debyser Z, Van der Eycken E (2013) Identification of small peptides inhibiting the integrase-LEDGF/p75 interaction through targeting the cellular co-factor. J Pept Sci 19(10):651–658
12. Cavalluzzo C, Voet A, Christ F, Singh BK, Sharma A, Debyser Z, Maeyer MD, Van der Eycken E (2012) De novo design of small molecule inhibitors targeting the LEDGF/p75-HIV integrase interaction. RSC Adv 2(3):974
13. Christ F, Voet A, Marchand A, Nicolet S, Desimmie BA, Marchand D, Bardiot D, Van der Veken NJ, Van Remoortel B, Strelkov SV, De Maeyer M, Chaltin P, Debyser Z (2010) Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat Chem Biol 6(6):442–448
14. Kumar A, Voet A, Zhang KY (2012) Fragment based drug design: from experimental to computational approaches. Curr Med Chem 19(30):5128–5147
15. Rhodes DI, Peat TS, Vandegraaff N, Jeevarajah D, Le G, Jones ED, Smith JA, Coates JA, Winfield LJ, Thienthong N, Newman J, Lucent D, Ryan JH, Savage GP, Francis CL, Deadman JJ (2011) Structural basis for a new mechanism of inhibition of HIV-1 integrase identified by fragment screening and structure-based design. Antivir Chem Chemother 21(4):155–168
16. Peat TS, Dolezal O, Newman J, Mobley D, Deadman JJ (2014) Interrogating HIV integrase for compounds that bind—a SAMPL challenge. J Comput Aided Mol Des (this issue)
17. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci USA 102(48):17308–17313
18. Wolber G, Langer T (2005) LigandScout: 3-D pharmacophores derived from protein–bound ligands and their use as virtual screening filters. J Chem Inf Model 45(1):160–169
19. OMEGA, version 2.4.6 OpenEye Scientific Software, Inc., Santa Fe, NM, USA. www.eyesopen.com, 2012
20. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank

and cambridge structural database. J Chem Inf Model 50(4):572–584
21. Hawkins PC, Nicholls A (2012) Conformer generation with OMEGA: learning from the data set and the analysis of failures. J Chem Inf Model 52(11):2919–2936
22. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267(3):727–748
23. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol 2(5):317–324
24. Molecular Operating Environment (MOE), version 2011.10; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2010
25. Voet A, Berenger F, Zhang KYJ (2013) Electrostatic similarities between protein and small molecule ligands facilitate the design of protein–protein interaction inhibitors. PLoS ONE 8(10):e75762
26. Glide, version 5.7, Schrödinger, LLC, New York, NY, 2011
27. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47(7):1739–1749
28. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. J Med Chem 49(21):6177–6196
29. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 47(7):1750–1759
30. Maestro, version 9.2, Schrödinger, LLC, New York, NY, 2011
31. LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 2011
32. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236
33. Prime, version 3.0, Schrödinger, LLC, New York, NY, 2011
34. Bashford D, Case DA (2000) Generalized Born models of macromolecular solvation effects. Annu Rev Phys Chem 51(1): 129–152
35. Tsui V, Case DA (2000) Theory and applications of the generalized born solvation model in macromolecular simulations. Biopolymers 56(4):275–291
36. Velec HFG, Gohlke H, Klebe G (2005) DrugScore^{CSD}-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. J Med Chem 48(20):6296–6303
37. The PyMOL Molecular Graphics System, version 1.5.0.4 Schrödinger, LLC
38. Williams T, Kelley C (2012) Gnuplot 4.6: an interactive plotting program
39. Fry DC (2012) Small-molecule inhibitors of protein–protein interactions: how to mimic a protein partner. Curr Pharm Des 18(30):4679–4684
40. Voet A, Banwell EF, Sahu KK, Heddle JG, Zhang KY (2013) Protein interface pharmacophore mapping tools for small molecule protein: protein interaction inhibitor discovery. Curr Top Med Chem 13(9):989–1001
41. Voet A, Zhang KY (2012) Pharmacophore modelling as a virtual screening tool for the discovery of small molecule protein–protein interaction inhibitors. Curr Pharm Des 18(30):4586–4598
42. Mobley D, Liu S, Lim N, Deng N, Branson K, Perryman A, Forli S, Levy R, Gallicchio E, Olson A (2014) Blind prediction of HIV integrase binding from the SAMPL4 challenge. J Comput Aided Mol Des (this issue)
43. Voet A, Callewaert L, Ulens T, Vanderkelen L, Vanherreweghe JM, Michiels CW, De Maeyer M (2011) Structure based discovery of small molecule suppressors targeting bacterial lysozyme inhibitors. Biochem Biophys Res Commun 405(4):527–532
44. Voet A, Helsen C, Zhang KYJ, Claessens F (2013) The discovery of novel human androgen receptor antagonist chemotypes using a combined pharmacophore screening procedure. Chem Med Chem 8(4):644–651
45. Perryman AL, Santiago DN, Forli S, Olson AJ (2013) Virtual Screening with AutoDock Vina and the Common Pharmacophore Engine of a low diversity library of fragments and hits against the three allosteric sites of HIV integrase: participation in the SAMPL4 protein–ligand binding challenge. J Comput Aided Mol Des. doi:10.1007/s1082201497093
46. Coleman RG, Sterling T, Weiss DR (2013) SAMPL4 & DOCK3.7: Lessons for automated docking procedures. J Comput Aided Mol Des (in press)