J Comput Aided Mol Des (2012) 26:319–327
DOI 10.1007/s10822-012-9558-x

A reverse combination of structure-based and ligand-based strategies for virtual screening

Álvaro Cortés-Cabrera · Federico Gago · Antonio Morreale

Received: 18 November 2011 / Accepted: 24 February 2012 / Published online: 7 March 2012

Abstract A new approach is presented that combines structure- and ligand-based virtual screening in a reverse way. Opposite to the majority of the methods, a docking protocol is first employed to prioritize small ligands ("fragments") that are subsequently used as queries to search for similar larger ligands in a database. For a given chemical library, a three-step strategy is followed consisting of (1) *contraction* into a representative, non-redundant, set of fragments, (2) *selection* of the three best-scoring fragments docking into a given macromolecular target site, and (3) *expansion* of the fragments' structures back into ligands by using them as queries to search the library by means of fingerprint descriptions and similarity criteria. We tested the performance of this approach on a collection of fragments and ligands found in the ZINC database and the directory of useful decoys, and compared the results with those obtained using a standard docking protocol. The new method provided better overall results and was several times faster. We also studied the chemical diversity that both methods cover using an in-house compound library and concluded that the novel approach performs similarly but at a much smaller computational cost.

Keywords Fragment screening · Structure-based virtual screening · Ligand-based virtual screening · Docking · Drug design

Introduction

The publication of Abbott's seminal paper describing the SAR (structure–activity relationships) by nuclear magnetic resonance (NMR) method [1] introduced the *fragment* as a new concept within the field of drug discovery, shifting the emphasis from the paradigmatic more conventional ligands (in terms of size and affinity) to smaller pieces. These "fragments" are usually endowed with reduced affinities but are better suited for chemical modifications aimed at producing novel drug candidates. Since its onset, the technique has experienced a soaring success in large pharma, small biotechs, and academia [2]. Fragments can sample chemical space more effectively than regular ligands do [3] and fragment docking clearly outperforms traditional high-throughput screening in terms of hit rates [4, 5]. Parallel to the purposeful deployment of customized software, some other computational techniques have been adapted to handle fragments, in particular those that attempt to yield ligands by starting off from these building blocks. Fragments can evolve virtually (by adding different chemical decorations), be linked (to join two or more fragments that occupy different regions of the binding site), self-assemble (through direct bond formation between different reacting fragments), and/or be optimized to better fulfill drug-like properties. Fragments are usually docked and scored, but due to the fact that their volumes are smaller than the binding site cavity erroneous binding modes can be obtained. Besides, scoring functions need to be fine-tuned as they are parameterized for much larger

Electronic supplementary material The online version of this article (doi:10.1007/s10822-012-9558-x) contains supplementary material, which is available to authorized users.

Á. Cortés-Cabrera · F. Gago
Departamento de Farmacología, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain

Á. Cortés-Cabrera · A. Morreale (✉)
Unidad de Bioinformática, Centro de Biología Molecular Severo Ochoa (CSIC/UAM), Campus UAM, c/Nicolás Cabrera 1, 28049 Madrid, Spain
e-mail: amorreale@cbm.uam.es

molecular entities. Nonetheless, despite these deficiencies, much progress has been made in the field and fragment-based ligand design (FBLD) has become a routine tool nowadays. Fragments can be designed de novo or obtained from ligand databases by chemical dissociation (i.e. fragmentation).

Virtual screening (VS) techniques that rely on the structure of either the macromolecular receptor (SBVS) or a known ligand (LBVS) use chemical libraries to search for hits that can then be transformed into leads. But due to the huge amount of ligands that can be found in today's databases (e.g. 13 million in ZINC [6]), it is impractical, as well as inefficient, to perform lengthy full docking studies. To speed up the process, it is customary to employ a series of computational cost-effective filters to narrow down the number of molecules that will be subjected to the demanding tasks of docking and scoring. Lipinki's rule of five [7] and/or other physico-chemical-based principles can be used as filters, as well as pharmacophoric hypotheses. LBVS can also be employed by taking some known ligands as templates although the outcome may be devoid of novelty because the resulting molecules tend to resemble the original queries.

In order to overcome this drawback, we designed a new protocol that combines SBVS and LBVS in a reverse order, that is, fragment SBVS serves as a previous filter to LBVS. This greatly reduces the computing time while maintaining the chemical diversity that is contained in the original library. For a given compound collection the new three-step strategy performs the following tasks: (1) *contracts* the database into a representative, non-redundant, set of fragments that are used for docking against the target of interest, (2) *selects* the three best-scoring fragments; and (3) *expands* the structure of the fragments back into ligands by employing the fragments as queries to search the database using a fingerprints description and a similarity criterion.

To test this new approach we first demonstrated that our docking tool is able to reproduce the experimental poses for a set of receptor-bound fragments in complexes of known 3D structure. Next, we applied the three-step procedure to the "fragment-like subset" in the ZINC database [6] using the 40 macromolecular targets contained in the directory of useful decoys (DUD) [8] and our CGRID/CDOCK docking tool [9]. The performance of the model was assessed by means of the area under the curve (AUC) of the generated receiver operating characteristics (ROC) plots. Finally, a comparison was made between the chemical space covered by the hits obtained from a standard VS protocol based on small-molecule docking and that provided by the top-ranking docked fragments generated from these same molecules.

Methods

Fragment definition

A fragment is defined as a molecular entity endowed with the following properties: (a) log $P \leq 2.5$, (b) molecular weight $\leq$250 Da, and (c) <6 rotatable bonds.

Fragment docking

Thirty-four complexes [10] from the ASTEX diverse set [11] in which the ligands fulfill the above fragment definition criteria were used as a test set. The fragments were extracted from the complexes and converted to SMILES [12] strings using OpenBABEL [13]. Then, our standard docking workflow was followed:

1. For the ligands: (a) conversion from SMILES to 3D MOL2 using CORINA [14], (b) atomic charge calculations with MOPAC [15] (AM1 ESP method) on every single structure provided by CORINA; and c) atom type assignment according to the AMBER force field [16] and conformational analysis using ALFA [17]. The protonation and tautomeric states for some of the ligands were manually adjusted (Fig. S1, Supplementary Information).
2. The receptors, including those water molecules and metal ions essential for ligand binding (Table S1, Supplementary Information), were prepared using pdb2pqr [18] and adapted to the AMBER force field by using 250 steps of steepest descent followed by 2,000 steps of Polak-Ribiere conjugate gradient energy minimization.
3. Docking of ligands and fragments was performed with our in-house CGRID/CDOCK tool: (a) definition of the binding site as the space delimited by the axis-parallel box containing the co-crystallized ligand, augmented by 5 Å in each axis direction, (b) CGRID calculation of protein interaction fields (a 12–6 Lennard–Jones term and an electrostatic term modeled with a sigmoidal dielectric screening function) covering the binding site (0.5 Å spacing in all directions) using common atom probes (C, N, O, S, P, H, F, Cl, Br, and I), (d) exhaustive exploration by CDOCK of the location and orientation of each fragment within the binding site by positioning their centers of mass on grid points and performing discrete rotations of 27° on each axis, (e) energy evaluation of each pose by the molecular mechanics force-field scoring function, as implemented in CDOCK, and (f) selection of the best-scoring pose for each fragment as the docking solution.

The standard criterion to validate the ability to predict the native pose was the root-mean-square deviation (rmsd) of the heavy atoms between the docking solution and the

native conformation for each fragment. In molecular docking, those poses within 2 Å from the experimental structure are usually considered as correct solutions. In the case of fragments, due to their reduced size, this cutoff is customarily decreased to 1.5 Å. Each docking experiment was run 20 times, and the reported rmsd corresponds to the average value. The success rate was defined as the percentage of poses having an average rmsd below 1.5 Å.

Fragment screening

The fragments used here were those belonging to the "fragment-like subset" in the ZINC database. Due to different conversion problems, not all the SMILES codes were able to produce the corresponding 3D structures, and for this reason only 6,183 were employed out of the original 7,106 strings. As receptors, we used the 40 structures comprising the DUD dataset (Table 1) and these were prepared for docking as explained before for the ASTEX test set.

Initially, all the fragments were docked (SBVS) using CGRID/CDOCK on each DUD target and only the best-scoring pose for each one was retained. Then, the three higher-ranking fragments for each target were selected and converted to MACCS fingerprints [19]. The LBVS protocol consisted of using these MACCS as queries to screen the DUD sets of real binders and decoys using the Tanimoto coefficient (Tc) index as the score. Performance is reported as the AUCs corresponding to the best experiment out of the three performed for each target (one for each of the three best fragments/target). The ligand list retrieved from each of these three fragments, at least in theory, should belong to different regions of chemical space. Thus, repeating the experiments 3 times is aimed at increasing the chemical diversity of the results. Test calculations with more than three fragments did not result in better coverage. In a real-world situation the top-scoring compounds for each of the three lists should be selected for testing, so as to improve the likelihood of finding new hits.

Comparative test

Fifty-two thousand two hundred and thirty-one molecules from an in-house chemical library that were ranked as possible hits in a standard SBVS campaign were decomposed into fragments with our tailor-made program, based on the chemistry development kit (CDK) [20], that makes use of the *exhaustive fragmentator* tool to break all rotatable bonds and generate fragments with at least 5 heavy atoms. A total of 1,137,482 fragments were thus extracted and then clustered using the stochastic clustering algorithm that is implemented in the SUBSET program [21]. MACCS fingerprints represented the fragments and a maximum Tc similarity index of 0.6 was used as a cutoff. This procedure, which is the same that was used in the ZINC database to obtain the so-called "fragment-like subset", yielded 2,540 non-redundant fragments. The new protocol then proceeded as follows:

1. Fragment docking and scoring with CGRID/CDOCK;
2. Selection of the best-scoring pose for each fragment and of the three best-scoring fragments;
3. Comparison between the chemical spaces covered by the fragments and by the parent compounds. The overlap (O, Eq. 1) between both spaces was obtained by:
   - an all-versus-all comparison between the 2,540 non-redundant fragments and the 52,231 parent compounds, and calculation of the Tcs among them as a control test;
   - selecting the top 1% compounds from both protocols and calculating the Tcs among them; and finally,
   - selecting the top 1% compounds from the standard protocol and the three best-scoring fragments, and calculating the Tcs among them.

$$O = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathrm{Tc}\ (i,j) \qquad (1)$$

where O, the overlap, is obtained as the averaged sum of the Tcs; m is the total number of ligands (520 [1% of 52,231]) and n the total number of fragments (either 24 [1% of 2,540] or only 3), and i and j are indices. To perform these comparisons ligands and fragments were represented using chemically advanced template search (CATS) descriptors [22].

Results and discussion

Fragment docking versus ligand docking

We first performed a "self-docking experiment" to test the accuracy of our CGRID/CDOCK docking engine when working with fragments. On average, for regular ligands (Fig. 1a), CGRID/CDOCK was able to reproduce the pose found in the X-ray crystal structure of the complex within an rmsd of 2.0 Å with a success rate of 75–80% (unpublished results). For fragments, and considering 1.5 Å as the cutoff value, we obtained a success rate of 80% with an average rmsd value of 0.88, in good consonance with recent studies [23, 24].

Fragment screening using ZINC fragments and DUD targets

Next, we tested the *contraction/selection/expansion* approach (Fig. 1b):

Table 1 AUCs values corresponding to each individual target depending on the method used: FBP (the fragment-based protocol presented here) and SP (standard ligand docking protocol)

Target[a]	FBP[b]	SP[c]	Target[a]	FBP[b]	SP[c]
ACE	0.79	0.63	HIVRT	0.53	0.61
AChE	0.45	0.73	HMGR	0.11	0.57
ADA	0.60	0.65	HSP90	0.83	0.69
ALR2	0.57	0.57	INHA	0.57	0.40
AMPC	0.36	0.58	MR	0.82	0.78
AR	0.76	0.65	NA	0.77	0.75
CDK2	0.68	0.52	P38	0.75	0.50
COMT	0.71	0.87	PARP	0.69	0.65
COX1	0.24	0.53	PDE5	0.67	0.78
COX2	0.51	0.67	PDGFRB	0.58	0.24
DHFR	0.67	0.49	PNP	0.53	0.60
EGFR	0.77	0.52	PPARγ	0.90	0.43
ER_{ago}	0.97	0.64	PR	0.76	0.49
ER_{antago}	0.75	0.81	RXRα	0.96	0.92
FGFR1	0.63	0.31	SAHH	0.51	0.82
FXa	0.61	0.55	SRC	0.53	0.48
GART	0.70	0.55	Thr	0.60	0.67
GPB	0.69	0.83	TK	0.68	0.59
GR	0.78	0.61	Trypsin	0.63	0.66
HIVPR	0.28	0.33	VEGFR2	0.71	0.41
					Averages
				SP	0.60
				FB[d]	0.54
				FB[e]	0.64

The average values appear in the last three rows

[a] *ACE* angiotensin-converting enzyme, *AChE* acetylcholinesterase, *ADA* adenosine deaminase, *ALR2* aldose reductase, *AmpC*, AmpC β-lactamase, *AR* androgen receptor, *CDK2* cyclin-dependent kinase 2, *COMT* catechol *O*-methyltransferase, *COX-1* cyclooxygenase-1, *COX-2*, cyclooxygenase-2, *DHFR* dihydrofolate reductase, *EGFR* epidermal growth factor receptor, ER_{ago} estrogen receptor (agonist-bound conformation), ER_{antago} estrogen receptor (antagonist-bound conformation), *FGFR1* fibroblast growth factor receptor kinase, *FXa* factor Xa, *GART* glycinamide ribonucleotide transformylase, *GPβ* glycogen phosphorylase β, *GR* glucocorticoid receptor, *HIVPR* HIV protease, *HIVRT* HIV reverse transcriptase, *HMGR* hydroxymethylglutaryl-CoA reductase, *HSP90* human heat shock protein 90, *INHA* enoyl ACP reductase, *MR* mineralocorticoid receptor, *NA* neuraminidase, *P38 MAP* P38 mitogen activated protein, *PARP* poly(ADP-ribose) polymerase, *PDE5* phosphodiesterase 5, *PDGFRB* platelet derived growth factor receptor kinase, *PNP* purine nucleoside phosphorylase, *PPARγ* peroxisome proliferator activated receptor γ, *PR* progesterone receptor, *RXRα* retinoic X receptor α, *SAHH* S-adenosyl-homocysteine hydrolase, *SRC* tyrosine kinase SRC, *Thr* thrombin, *TK* thymidine kinase, *VEGFR2* vascular endothelial growth factor receptor

[b] Fragment-based protocol presented in this paper

[c] Standard protocol

[d] The average was calculated with the original collection of fragments (not properly representing NHR binders' fragments)

[e] The average was calculated with the appropriate fragments to represent NHR binders (see Fig. 4)

(a) *Contract* the ligand library into a representative, non-redundant set of fragments that are used for docking;
(b) *select* the three best-scoring fragments; and
(c) *expand* these fragments by searching the database for ligands containing similar substructures.

For comparison purposes, SBVS was also carried out with the original ligands (true binders + decoys). Despite the fact that the total number of ligands was less than the number of fragments, the time and computational resources required for this procedure were generally larger than in the new fragment-based approach (see below). Overall, the average AUC for the two protocols was very similar ($\sim$0.6, Table 1) because, although the new method was superior for 62% of the targets (26 out of 42), it proved inferior in a few others. To analyze in more detail these differences in performance the targets were grouped into families (Fig. 2).

This allowed us to see that the new methodology outperformed the standard method when kinases and folate

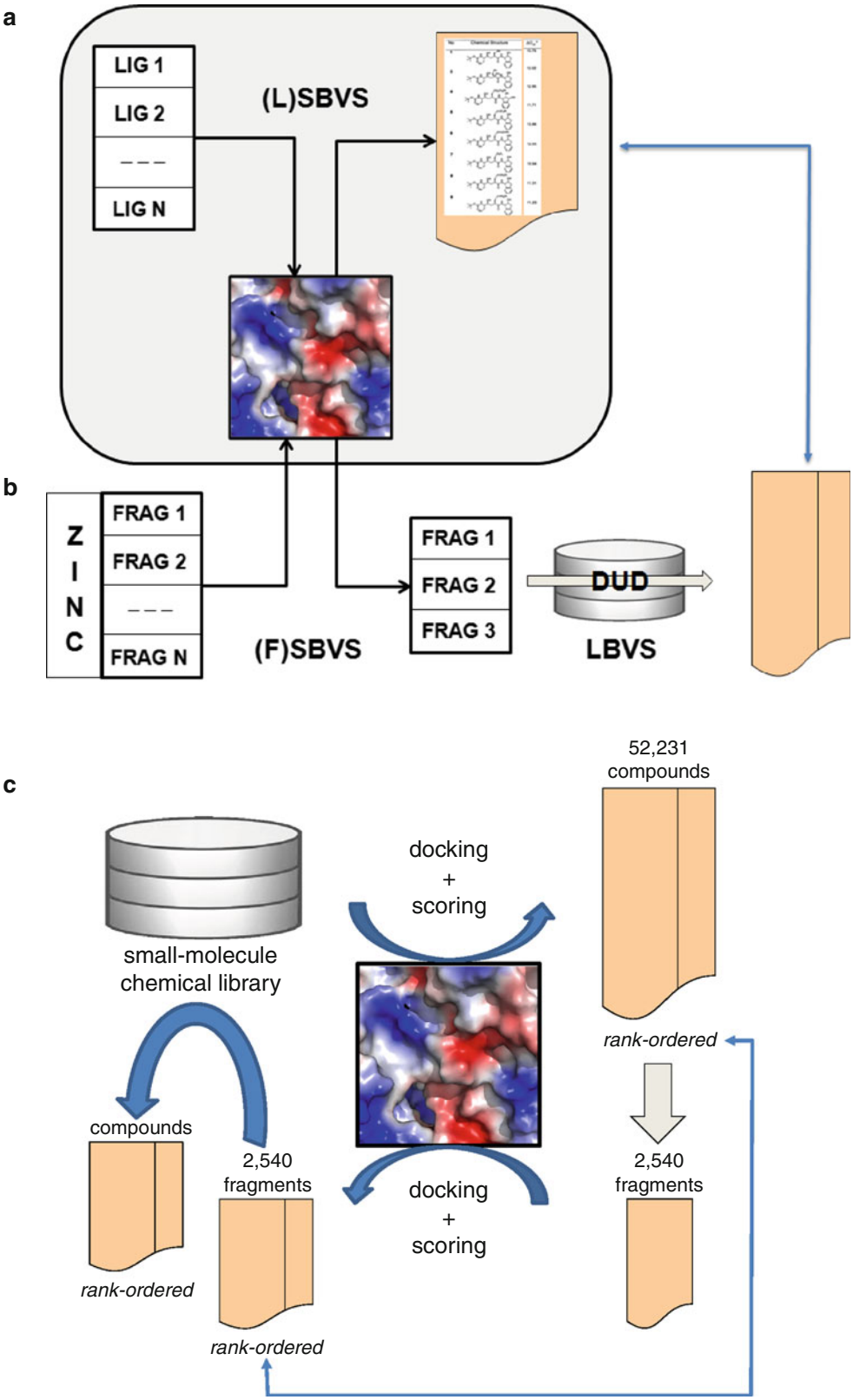

Fig. 1 Schematic representation of the classical ligand—based (**a**) and fragment-based (**b**) SBVS procedures and the novel *contraction/selection/expansion* approach (**c**). The *blue double arrowheaded lines* stand for comparisons of the results between two given methods

enzymes were used as targets whereas true binders were recovered at similar rates in metalloenzymes and serine proteases. Strikingly, the worst performance for the new method was observed in the case of nuclear hormone receptors (NHR). When we analyzed the distribution of similarity indices between the ZINC fragments and the true

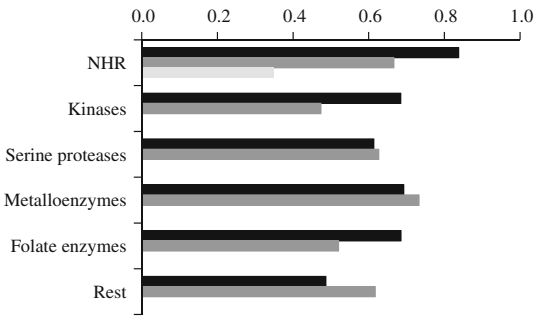

Fig. 2 AUCs obtained for the targets grouped into families. NHR: AR, ER_{ago}, ER_{antago}, GR, MR, PPARγ, PR, and RXRα; kinases: CDK2, EGFR, FGFR1, HSP90, P38 MAP, PDGFRB, SRC, TK, and VEGFR2; serine proteases: FXa, Thr, and trypsin; metalloenzymes: ACE, ADA, COMT, and PDE5; folate enzymes: DHFR and GART; and the rest: AChE, ALR2, AmpC, COX-1, COX-2, GPB, HIVPR, HIVRT, HMGR, INHA, NA, PARP, PNP, and SAHH (for the abbreviations see Table 1). *Black* and *grey bars* correspond to the fragment-based protocol presented here and the standard SBVS protocol with full ligands, respectively. An additional *bar* (*light gray*) was included in the case of NHR targets to highlight the performance of the new approach when the original collection of fragments was used

binders for these targets we realized that the latter were not properly represented using the original fragment collection, as illustrated in Fig. 3 for two prototypical targets of this class: MR, and RXRα.

Upon inclusion of appropriate representative ad hoc fragments (Fig. 4) we obtained AUC values of 0.82 and 0.96, respectively, for these two NHR that compared very favorably to the previous 0.13 and 0.24 that were achieved when these fragments were not included, and also to the 0.78 and 0.92 obtained with the standard protocol. In fact, the overall performance for the whole NHR family improved notably, as seen in Fig. 2 and also in Table 1, which displays the results upon incorporation of these new fragments. In view of this result, we tested the performance of fragments specifically derived from ligands bound to representative DUD targets (1/family) and found AUC values (0.78 for HSP90, 0.68 for ACE, 0.73 for AChE, and 0.74 for fXa) similar to those reported in Table 1.

These findings highlight the fact that as long as the fragment database is able to adequately represent the chemical diversity of the true binders, the present method can clearly outperform the standard classical SBVS procedure employing whole ligands. To better appreciate the structural similarities between true binders and decoys Fig. 5 shows the chemical structure of a query molecule (at the center), three true binders retrieved with high Tcs (at the bottom) and three decoys with low Tcs (at the top). The true binders's substructures that resemble the query fragment are highlighted in blue.

A fragment-based approach displaying similarities with our own has been published recently [25], although it appears to be more focused on the optimization of query molecules because the ligand database is decomposed into fragments that are evaluated for binding affinity using docking and scoring. Thereafter, those fragments exhibiting the lowest affinities are replaced by new ones and the affinity is re-calculated. The outcome is an optimized ligand made up of the best fragments. A more elaborate approach [26, 27] extends the query beyond the fragment itself by considering its microenvironment, which includes the relevant interacting protein residues. By compiling a diverse set of micro-environments (e.g. from the PDB) it is possible to optimize already known structures and/or suggest novel and improved compounds.

Comparative test of chemical diversity

Finally, to test the extent to which the chemical space represented by a set of docked compounds is covered by their corresponding fragments once they have been docked

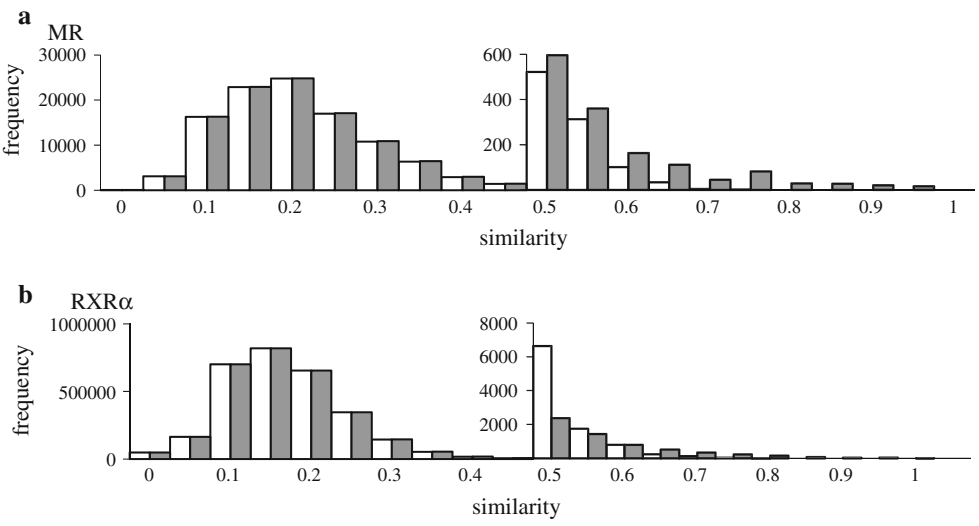

Fig. 3 *Histograms* showing the distribution of similarity indices between the ZINC fragments and the ligands (true binders + decoys) for two NHR targets (for the abbreviations see Table 1), MR (**a**) and RXRα (**b**), before (*white bars*) and after (*grey bars*) incorporation of the new fragments. Note that for Tc values (*x*-axes) higher than 0.5 the *y*-axes have been *scaled* up to *highlight* that the main differences occur in this region

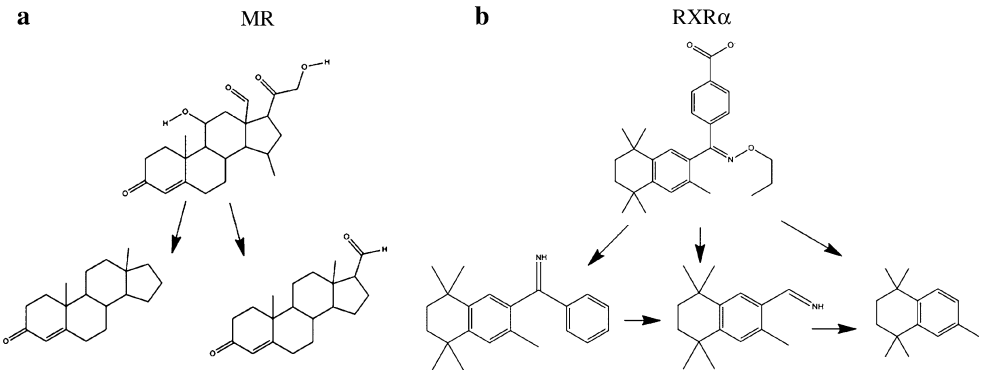

Fig. 4 MR (**a**) and RXRα (**b**) true binders and their appropriate fragment decomposition. These fragments were not generated in the default procedure that yielded the original collection of fragments

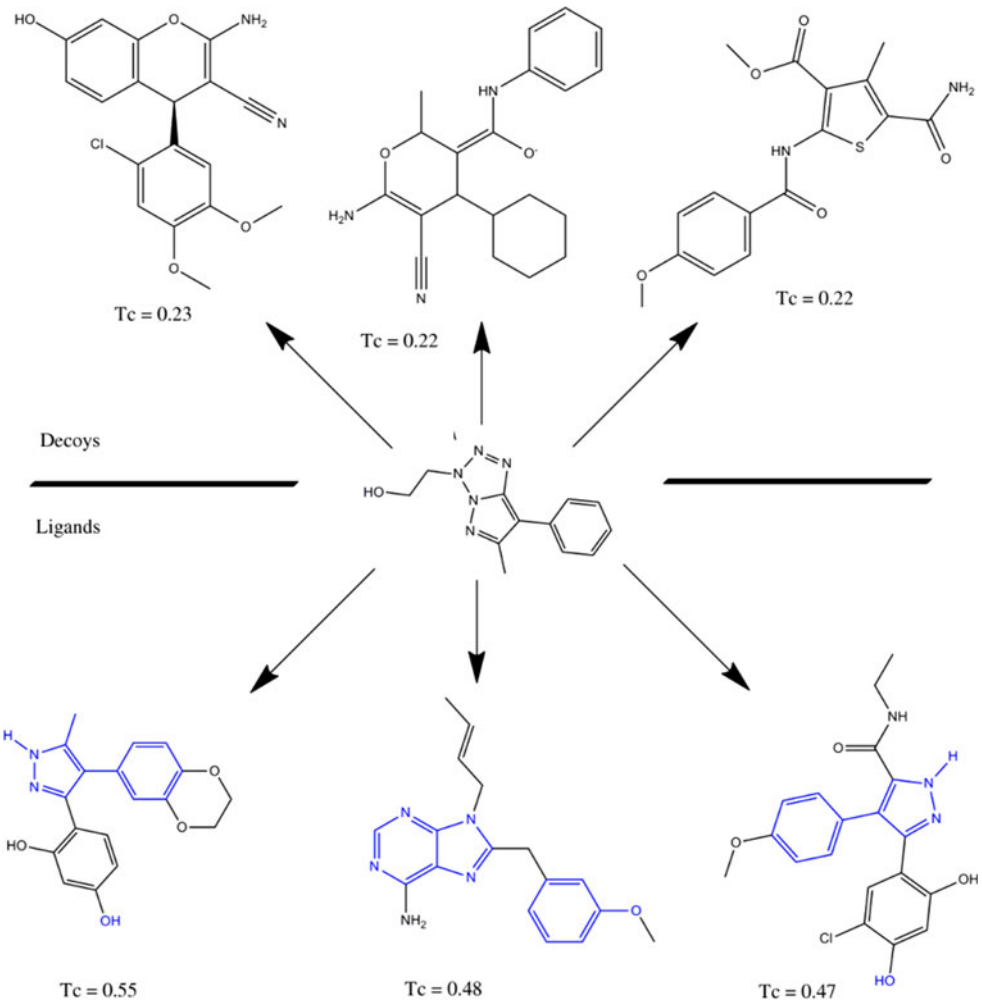

Fig. 5 Structural similarities between a query molecule (at the *center*), three true HSP90 binders (at the *bottom*), and three decoys (at the *top*). The true binders' substructures contained in the query are *highlighted* in *blue*

into the same target, we took 52,231 molecules from an in-house chemical library that had been subjected to SBVS and ranked by docking scores. The all-versus-all comparison between the 2,540 non-redundant fragments and the 52,231 parent compounds afforded a modest value (50%) for the overlap (*O*). This percentage reflects that considerable noise is introduced when the comparison is made without a previous filtering step, that is, the overlap one might expect when a brute force approach is employed. This value, however, is increased to 72% when the top 1% compounds from both rank-ordered lists are considered, which suggests that a greater success can be achieved when compound selection is made on the basis of a more robust criterion such as the complementarity between the fragments and the binding site, as represented here by the docking step. Furthermore, when only the 3 best-scoring fragments were selected for the comparison, an overlap of 82% was obtained. This indicates, again, that careful

fragment selection is likely to result in rather successful solutions that efficiently cover the chemical space explored by a traditional SBVS protocol. Therefore we believe that further improvements on this novel approach can be expected by refining the fragment generation procedure and by fine-tuning the similarity searches.

Benchmarking

Our in-house VS platform VSDMIP [28] was used for all the calculations. The input is a SMILES string representation for each ligand or fragment. These SMILES are inserted into the VSDMIP database by processing them as outlined before (conversion to 3D, charges and radii assignment, and conformational analysis). On average, using either a 32-bit 3.2 GHz PIV or a 32-bit 3.06 GHz Xeon processor, VSDMIP is able to insert into the database 2,600 fragments per CPU and day. This is roughly the same number of ligands that can be typically processed. Taking into account that fragments, by definition, are smaller in size than ligands, the equivalence in these insertion times reflects the structural complexity of the fragments. In fact, they are not fragments in the strictest sense as they have not been generated from splitting ligands into smaller pieces. Rather, they are small ligands that fulfill the properties arbitrarily defined for fragments. On the other hand, in terms of docking, and due to their reduced size compared to typical ligands, processing times are greatly shortened (248 ligands vs. 553 fragments docked per CPU and day, using the same CPUs as above). Finally, VSDMIP can perform 2.4×10^9 comparisons/day and CPU using 2D fingerprints. With all these numbers in hand, it can be stated that they are affordable on almost any small- to medium-size cluster ($\approx$25–100 processors) and would allow users to perform several complete protocols (as the one described here)/day. In this way, many tests can be easily conducted, which increases the likelihood of obtaining new promising hits.

Conclusions

We have presented a new way to combine SBVS and LBVS strategies that consists of three steps: (1) organize and reduce a ligand database into a non-redundant fragment library (*contraction* step), (2) dock all of these fragments into the binding pocket of the macromolecular target and select the three best-scoring ones (*selection* step, SBVS), and (3) perform a similarity analysis using the selected fragments to interrogate the database and select the most similar ligands (*expansion* step, LBVS). Compared to a typical SBVS campaign, computer running times are considerably reduced as a consequence of the fragments' smaller size relative to typical ligands and their non-redundancy. On the other hand, chemical space coverage is not critically compromised. This protocol allows the user to focus on specific regions within the chemical space present in the ligand database by enabling: (1) more efficient VS runs, (2) development of focused virtual libraries, and (3) scaffold hopping for chemical modification. Provided that the fragments have been previously obtained (e.g. using CDK as mentioned above), all the operations can be performed within the VSDMIP graphical environment, which was implemented as a PyMOL plugin [28]. Finally, it should be mentioned that the performance of the method is expected to be highly dependent on the ligand database that is employed to obtain the fragments: the greater the diversity, the better the coverage of chemical space.

Acknowledgments This work was supported by grants from Ministerio de Ciencia e Innovación (MICINN) BIO2008-04384 (to Antonio Morreale) and SAF2009-13914-C02-02 (to Federico Gago), and Comunidad Autónoma de Madrid (CAM) S-BIO-0214-2006 (BIPEDD) and S2010-BMD-2457 (BIPEDD-2). Antonio Morreale acknowledges CAM for financial support to the Fundación Severo Ochoa through the AMAROUTO program. Álvaro Cortés-Cabrera thanks Ministerio de Educación for the FPU Grant AP2009-0203. We are grateful to OpenEye Scientific Software, Inc. for providing us with an academic license for their software. The technical support and advice from the Bioinformatics team at CBMSO is gratefully acknowledged.

References

1. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274(5292):1531–1534
2. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6(3):211–219. doi:10.1038/nrd2220
3. Lipinski C, Hopkins A (2004) Navigating chemical space for biology and medicine. Nature 432(7019):855–861. doi:10.1038/nature03193
4. Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. J Chem Inf Comput Sci 41(3):856–864. doi:10.1021/ci000403i
5. Schuffenhauer A, Ruedisser S, Marzinzik AL, Jahnke W, Blommers M, Selzer P, Jacoby E (2005) Library design for fragment based screening. Curr Top Med Chem 5(8):751–762
6. Irwin JJ, Shoichet BK (2005) ZINC—a free database of commercially available compounds for virtual screening. J Chem Inf Model 45(1):177–182. doi:10.1021/ci049714+
7. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46(1–3):3–26
8. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. J Med Chem 49(23):6789–6801. doi:10.1021/jm0608356
9. Perez C, Ortiz AR (2001) Evaluation of docking functions for protein-ligand docking. J Med Chem 44(23):3768–3785

10. PDB IDs 1GPK, 1HWW, 1IA1, 1JD0, 1J3J, 1HNN, 1HQ2, 1IG3, 1K3V, 1LRH, 1N1M, 1N2V, 1OF1, 1OF6, 1OWE, 1P2Y, 1P62, 1Q1G, 1Q41, 1Q4G, 1R9O, 1SG0, 1SQW, 1TOW, 1TT1, 1TZ8, 1UIC, 1U4D, 1UOU, 1W1P, 1W2G, 1X8X, and 1XM6
11. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WT, Mortenson PN, Murray CW (2007) Diverse, high-quality test set for the validation of protein-ligand docking performance. J Med Chem 50(4):726–741. doi:10.1021/jm061277y
12. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28(1):31–36. doi:10.1021/ci00057a005
13. Open Babel (2011) The open source chemistry toolbox. http://openbabel.org/wiki/Main_Page. Accessed 01 March 2012
14. Corina (2000) Computerchemie Langemarckplatz 1 E, Germany, MNG
15. Stewart JJP (1990) MOPAC: a semiempirical molecular orbital program. J Comput Aided Mol Des 4(1):1–103. doi:10.1007/bf00128336
16. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688. doi:10.1002/jcc.20290
17. Gil-Redondo R (2006) Master thesis UNED, Madrid
18. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res 32(Web Server issue):W665–W667. doi:10.1093/nar/gkh381
19. Murray CW, Baxter CA, Frenkel AD (1999) The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. J Comput Aided Mol Des 13(6):547–562
20. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E (2003) The chemistry development kit (CDK): an open-source Java library for chemo- and bio-informatics. J Chem Inf Comput Sci 43(2):493–500. doi:10.1021/ci025584y
21. Voigt JH, Bienfait B, Wang S, Nicklaus MC (2001) Comparison of the NCI Open Database with Seven Large Chemical Structural Databases. J Chem Inf Comput Sci 41(3):702–712. doi:10.1021/ci000150t
22. Schneider G, Neidhart W, Giller T, Schmid G (1999) "Scaffold–Hopping" by topological pharmacophore search: a contribution to virtual screening. Angew Chem Int Ed Engl 38(19):2894–2896
23. Sandor M, Kiss R, Keseru GM (2010) Virtual fragment docking by Glide: a validation study on 190 protein-fragment complexes. J Chem Inf Model 50(6):1165–1172. doi:10.1021/ci1000407
24. Verdonk ML, Giangreco I, Hall RJ, Korb O, Mortenson PN, Murray CW (2011) Docking performance of fragments and druglike compounds. J Med Chem 54(15):5422–5431. doi:10.1021/jm200558u
25. Lin FY, Tseng YJ (2011) Structure-based fragment hopping for lead optimization using predocked fragment database. J Chem Inf Model 51(7):1703–1715. doi:10.1021/ci200136j
26. Moriaud F, Doppelt-Azeroual O, Martin L, Oguievetskaia K, Koch K, Vorotyntsev A, Adcock SA, Delfaud F (2009) Computational fragment-based approach at PDB scale by protein local similarity. J Chem Inf Model 49(2):280–294. doi:10.1021/ci8003094
27. Durrant JD, Friedman AJ, McCammon JA (2011) CrystalDock: a novel approach to fragment-based drug design. J Chem Inf Model 51(10):2573–2580. doi:10.1021/ci200357y
28. Cabrera AC, Gil-Redondo R, Perona A, Gago F, Morreale A (2011) VSDMIP 1.5: an automated structure- and ligand-based virtual screening platform with a PyMOL graphical user interface. J Comput Aided Mol Des 25(9):813–824. doi:10.1007/s10822-011-9465-6