J Comput Aided Mol Des (2015) 29:713–723
DOI 10.1007/s10822-015-9856-1

Ligand- and receptor-based docking with LiBELa

Heloisa dos Santos Muniz[1] · **Alessandro S. Nascimento**[1]

Received: 28 January 2015 / Accepted: 25 June 2015 / Published online: 4 July 2015

Abstract Methodologies on molecular docking are constantly improving. The problem consists on finding an optimal interplay between the computational cost and a satisfactory physical description of ligand-receptor interaction. In pursuit of an advance in current methods we developed a mixed docking approach combining ligand- and receptor-based strategies in a docking engine, where tridimensional descriptors for shape and charge distribution of a reference ligand guide the initial placement of the docking molecule and an interaction energy-based global minimization follows. This hybrid docking was evaluated with soft-core and force field potentials taking into account ligand pose and scoring. Our approach was found to be competitive to a purely receptor-based dock resulting in improved logAUC values when evaluated with DUD and DUD-E. Furthermore, the smoothed potential as evaluated here, was not advantageous when ligand binding poses were compared to experimentally determined conformations. In conclusion we show that a combination of ligand- and receptor-based strategy docking with a force field energy model results in good reproduction of binding poses and enrichment of active molecules against decoys. This strategy is implemented in our tool, LiBELa, available to the scientific community.

Keywords Docking · Ligand similarity · Hybrid docking · Scoring function · LiBELa

Introduction

The proper evaluation of molecular interactions is still an important challenge in molecular modeling [1, 2]. This evaluation is typically tackled by methods that can generally be classified among three major groups. The most rigorous group of methods involves an alchemical transformation of the system from an initial state (e.g., unbound) to a second stage (bound) in very small steps. Following Zwanzig's relationship [3], binding free energies can be computed if the energies are collected in 'windows' during the slow interconversion of the system [4]. These alchemical methods, such as free energy perturbation and thermodynamic integration, are computationally intensive and usually involve a number of molecular dynamics simulations. On the other hand, in good cases, binding free energies can be computed with accuracies of $\sim$1 kcal/mol or even better [5]. The second group of methods estimates the binding energy by evaluating the energies, as computed by molecular mechanics methods, in the endpoints of the thermodynamic cycle for binding. These methods, such as MM-GBSA or MM-PBSA, are less computationally intensive than the alchemy methods but still involve molecular dynamics or Monte Carlo simulation of the system in the bound and unbound conformation. A binding energy is then estimated based on small differences subtracted from large energies computed at the endpoints [6, 7]. Finally, the docking methods provide a rapid estimate of the binding energy by sampling a limited number of ligand configurations in the receptor active site and scoring them typically using a force field

Electronic supplementary material The online version of this article (doi:10.1007/s10822-015-9856-1) contains supplementary material, which is available to authorized users.

✉ Alessandro S. Nascimento
asnascimento@ifsc.usp.br

[1] Instituto de Física de São Carlos, Av. Trabalhador são-carlense, 400. Centro, São Carlos, SP 13566-590, Brazil

description of the binding energy [8]. These methods are much less intensive and definitely less accurate. They are certainly not adequate for precise estimates of binding free energies but were proven very useful for the rapid screening of compound libraries looking for binders of specific targets (e.g. [9]). Docking methods are now routinely used as important tools in most pharmaceutical companies and on most drug development projects in academia.

Whatever the chosen method, scoring the molecular interaction is a difficult task. Typically, small energy differences are computed from large energies, reducing the accuracy in the determination of the binding energy. Additionally, the binding energy landscape resembles the typical rugged folding landscape, with many local minima [10, 11], making the achievement of a correct low-energy binding pose even trickier. These points are especially important in the context of docking methods, where sampling is limited and quickness matters.

In principle, an increased convergence towards the global minimum in the binding energy landscape can be reached by smoothing the energy surface through a soft-core potential, for example. This approach could also be of advantage for modeling local receptor flexibility or adaptation upon ligand binding. Ferrari and coworkers evaluated the effect of smoothing the Lennard-Jones potential by reducing the repulsive term dependence from $1/r^{12}$ to $1/r^{9}$ [12]. The authors found a better enrichment for two specific molecular targets (T4 lysozyme and aldose reductase), although the increase in the enrichment is clearly overcome by the use of multiple receptor structures. The results suggest promising results for ligand docking, especially for single receptor conformation, but it is important to highlight that the outcomes are based on a limited number of targets.

Alternatively, an increased convergence towards a binding minimum can be reached by imposing valuable structural restraints that could act as driving forces leading to a faster achievement of a proper low-energy binding pose. Hawkins and coworkers analyzed HYBRID, a docking tool based on FRED, but that combines ligand structural information in the docking process, and found out that the introduction of this additional information results in significantly improved docking poses as compared to a purely receptor-based approach [13]. The enrichment against docking benchmarks were also improved [14], suggesting that a proper combination of ligand- and receptor-based strategies can be a feasible approach to tackle the docking problem.

Given the relevance of docking in the pipeline of drug discovery, we propose the combined usage of ligand structural information and receptor structure to increase the convergence towards a global minimum in the binding energy landscape. In addition, two different energy models, AMBER force field and a smoothed variation of AMBER force field, were evaluated. The results suggest that smoothed potentials can, at best, reproduce the enrichments observed for typical force fields (hardcore) enrichments. In terms of ligand pose, the smoothed potential allows greater plasticity increasing the average RMSDs and reducing the success rate of finding near native conformations in self-docking and cross-docking experiments. The ligand- and receptor-based docking tool LiBELa used all through this work is available to the academic community as a ligand- and receptor-based docking tool.

Methods

Hybrid docking algorithm design and energy models

The motivation for LiBELa's development is based on the results observed with our ligand similarity engine *MolShaCS*. This tool uses the overlap of molecular shape and charge distribution to measure the degree of similarity between two small molecules [15]. Since these descriptors are tridimensional, the molecules have to be properly aligned in the space prior to similarity computation and good alignments can be obtained by maximizing the degree of overlap in Cartesian space. Even topologically different molecules can be correctly and quickly placed in most cases using this algorithm. These results inspired the development of a docking algorithm that initially places a *search ligand* based on the match of physicochemical properties with a *reference ligand* already placed in the receptor binding site, and then optimizes this initial ligand pose towards a minimum in the interaction energy.

The overlap between two molecules, A and B, can be described by:

$$V_{AB} = \sum_{i \in A} \sum_{j \in B} \int dr \rho_i(r) \rho_j(r) \tag{1}$$

where

$$\rho(r) = p_i \, exp \left\{ \left[-\pi \left(\frac{3p_i}{4\pi\sigma_i^3} \right)^{\frac{2}{3}} \right] (r - r_i)^2 \right\} \tag{2}$$

Here, p_i is a constant ($2\sqrt{2}$), r_i is the distance to the *ith* atom and σ_i is the atomic van der Waals radius taken from AMBER FF99SB or GAFF force fields [16–18]. The best overlap is defined by the orientation of the search molecule that maximizes V_{AB}, when compared to a reference molecule. As previously described [15], the final objective function also incorporates the charge distribution in a similar way as described for atomic volume distribution.

Ligand flexibility is taken into account by generating multiple low-energy conformers of the search ligand. This

is achieved by the genetic algorithm implemented in the OpenBabel API [19]. Unless otherwise stated, 20 conformers were generated and combined with the initial (input) conformation for the evaluations described in the sections below. Each conformer is overlaid on the reference ligand and the best conformer (as judged by degree of overlap or by the initial interaction energy) is then used for the second stage of the docking procedure. Afterwards, this initial ligand pose is optimized in Cartesian space as a rigid body, using the interaction energy as the objective function. The two optimization procedures are achieved with the augmented lagrangians [20–22] and the dividing rectangles [23] algorithms, respectively, as implemented in the library NLOPT [24]. The dividing rectangles algorithm is considered a global optimization algorithm, in the sense that it probes a large phase space by subdividing it in smaller fractions for systematic searches.

Two energy models were initially tested here. The first model is a smooth variation of the all-atom AMBER force field non-covalent interaction energy:

$$E_{bind}^{SC} = \sum_{i \neq j} \frac{q_i q_j}{D(d_{ij} + \delta_{ele})} + \sum_{i \neq j} \frac{A_{ij}}{(d_{ij}^6 + \delta_{VDW}^6)^2} - \frac{B_{ij}}{\left(d_{ij}^6 + \delta_{VDW}^6\right)} \quad (3)$$

These terms account for polar interaction, in the form of a modified Coulomb potential and the van der Waals repulsive and attractive terms, in the form of a modified Lennard-Jones potential. Here, q_i is the charge of atom i, d_{ij} is the distance between atoms i and j, A_{ij} and B_{ij} are Lennard-Jones repulsive and attractive coefficients, described below; D accounts for the dielectric constant and is set as the interatomic distance d_{ij}. δ_{ele} and δ_{vdw} are smoothing parameters for electrostatic and Lennard-Jones potentials, respectively.

The second model is a pure all-atom AMBER force field for nonbonded interactions:

$$E_{bind}^{AMBER} = \sum_{i \neq j} \frac{q_i q_j}{D d_{ij}} + \sum_{i \neq j} \frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^6} \quad (4)$$

It is also interesting to note that these energy models have just a few adjustable parameters, namely the smoothing parameter δ_{ele} and δ_{vdw}. For all the tests reported here, δ_{ele} and δ_{vdw} were set as 1.0 and 1.5 Å, respectively, according to the results of preliminary analyses made in the group.

The computation of binding energies can be done directly or through pre-computed energy grids [25, 26]. In case where grids are used to speed up the calculations, a cubic box with 30 Å in each direction and centered on the center of mass of the reference ligand was defined and grid points were distributed equally in the box using a spacing of 0.3 Å. The computation of the energy grid is done 'on-the-fly' in LiBELa, using the geometric mean approximation [25, 27]:

$$A_{ij} = \sqrt{A_{ii}}\sqrt{A_{jj}} \quad (5)$$

and

$$B_{ij} = \sqrt{B_{ii}}\sqrt{B_{jj}} \quad (6)$$

where $A_{ii} = \varepsilon_i (2r_0)^{12}$ and $B_{ii} = 2\varepsilon_i (2r_0)^6$, with r_0 taken as the atomic radius and ε taken as the well depth parameter from Amber force field.

LiBELa was developed using C/C++ and also brings a MPI implementation for parallel docking of screening libraries. A graphical user interface (GUI) written using Qt libraries is provided, facilitating its usage. Instructions on how to obtain, compile and execute LiBELa are available in the page http://www.biotechmol.ifsc.usp.br.

Binding pose test

The correctness of the binding poses was evaluated by comparing the root mean square deviation for the test set SB2012 [28, 29], used as provided by Rizzo's group. This set contains 1,043 receptor-ligand complexes in SYBYL MOL2 files already prepared for docking. Briefly, the preparations was done by adding hydrogen atoms to the receptors and assignment of atomic charges using AMBER force field, both with *tleap* tool. The ligands were prepared using MOE, where the net charge was assigned and, afterwards, AM1-BCC charges were computed by antechamber [28]. In this *self-docking* test, the achievement of low RMSDs as compared to the experimental pose (crystallographic) were analyzed for non-hydrogen atoms.

Cross-docking

In a more stringent test, the ability of the hybrid docking algorithm to reproduce ligand poses in different receptor structures (cross-docking) was evaluated using the Astex non-native test set [30]. From the 58 provided structures with analogous complexes, four were left beside due to missing force field parameters. The remaining set resulted in 860 ligands for 54 receptor structures. The receptors were prepared using the DockPrep tool as available in UCSF Chimera [31]. Briefly, the atomic charges were assigned using AMBER FF99SB force field and atom types were assigned as SYBYL atom types. For ligands, the program ANTECHAMBER was used to compute AM1-BCC charges and assign SYBYL atom types. After docking, the RMSDs were computed using non-hydrogen atoms using the experimental pose as reference. The Astex test set was used in two experiments. First, a self-docking experiment was performed using different reference ligands. In

the second experiment, the ligands were docked in a different receptor structure, in a typical cross-docking evaluation.

Enrichment test

The ability to discriminate actual binders from decoy molecules was evaluated using the test sets DUD [32] and DUD-E [33]. In all cases, the receptors were automatically prepared using UCSF Chimera DockPrep tool [31]. Binders and decoys were used as provided with atomic charges assigned according to the typical protocol defined in ZINC [34, 35]. The enrichment was quantified by the area under the curve on a semi-log scale (logAUC). The area was then corrected by subtracting the area expected for a random enrichment, resulting the adjusted logAUC measure [36]. In all cases, twenty conformers and the original conformation ($n = 21$) of the search ligand were initially placed over the reference ligand provided together with the test sets using the match in their physicochemical properties (atomic volume and charge distribution) as the guiding parameter, as described by Eq. 1. The conformer with lowest interaction energy was then re-optimized in Cartesian space as a rigid body using the interaction energy as the objective function. The parameters for the optimization procedures were the same as described in the previous section. The ROC curves are represented as the fraction of active molecules versus the fraction of decoys (on log scale).

The same test sets were also evaluated with DOCK6.6 [37] in order to compare the results of a structure-based tool with our hybrid proposal. The grid score, which is similar in spirit to our AMBER force field was chosen as the scoring function in DOCK6. For grid computation in DOCK6, a 6–12 Lennard-Jones potential was used and the dielectric factor was set to 1.0. The grid spacing was set to 0.3 Å in both programs. During the docking, the maximal number of orientations was set to 500 using the chemical matching algorithm. After matching, the ligand structures were energy minimized with up to 100 cycles of simplex minimization with score and cycle convergence set to 0.1 and 1.0, respectively. Ligand flexibility was taken into account using the anchor-and-grow algorithm implemented in DOCK6.

Results and discussion

All the docking simulations were run on a computer cluster running Linux Operational System and equipped with two Intel Xeon E5645 2.4 GHz processors per node. The DUD enrichment test, involving over 116,800 molecules (ligands and decoys) typically took 8300 min for execution in total, corresponding to 14 molecules being processed per minute. For the sake of comparison, the program DOCK6 takes processes about 11 molecules per minute in the same benchmark, showing that the current implementation of the hybrid docking is competitive in terms of execution time. In this test, each target was assigned to a computed node for both tools and the time spent in grid computation was taken out of equation. These results show that our implementation of the hybrid approach in molecular docking is efficient in terms of execution time and probably feasible for virtual screening experiments.

We then assessed the correctness of energy evaluation in LiBELa by comparing the binding energies as computed by the grid evaluation and the Grid Score as computed by DOCK6. Since both tools use Amber force field for binding energy evaluation, the binding energies computed by these programs are expected to be similar, or at least correlated. Here, the SB2012 test set was used only to score the provided ligand orientations using both DOCK6 and LiBELa, i.e., with no ligand docking. We started by testing the correlation of the computed energies and, as shown in Fig. 1, the binding energies as computed by DOCK6 and LiBELa are highly correlated, indicating an efficient evaluation of the force field parameters in the grid implementation for LiBELa, as compared to the well-validated DOCK6. The Pearson correlation coefficient computed for DOCK6 grid score and LiBELa Amber energies is 0.45. However, after removal of four clear outlier points, the correlation coefficient reaches 0.88, indicating that the current implementation of energy computation in LiBELa is efficient, as compared to the well-known implementation of Amber binding energy available in DOCK. It is import to highlight two important differences in energy evaluation

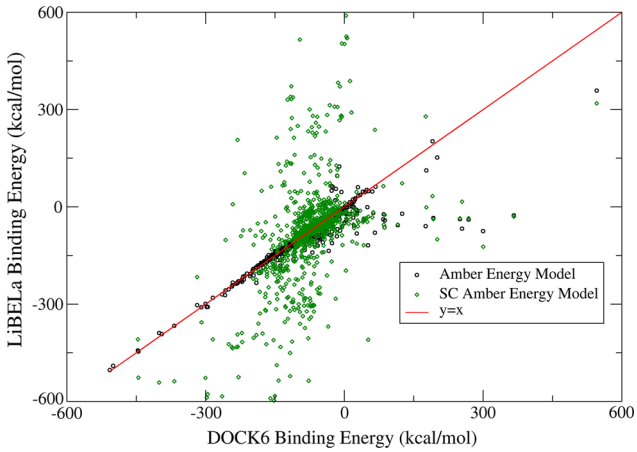

Fig. 1 Correlation among LiBELa binding energies (*vertical axis*) and DOCK6 Grid Score (*horizontal axis*) for SB2012 complexes. Energies were evaluated using both AMBER energy model (*black circles*) and soft-core energy model (*green diamonds*). The *red line* shows y = x

between DOCK6 and LiBELa. First, DOCK6 uses the dielectric 'constant' as $4r_{ij}$, where r_{ij} is the interatomic distance in the Coulomb potential, while in LiBELa we use r_{ij} only as the dielectric constant. Second, for atoms lying outside the grid box, DOCK6 computes the energy by a direct evaluation, as far as we are aware. In LiBELa, those atoms are just assigned with a very high and constant energy (10^6 kca/mol), indicating to the optimizer that a non-viable pose was achieved. The role of these differences can be observed in the right region of Fig. 1, where unfavorable (positive) energies are computed and the divergence between the tools increase.

We then set out to analyze the binding mode predictions on the test set SB2012 in a self-docking experiment. Some of the SB2012 complexes could not be used by LiBELa, due to missing force field parameters and, for some complexes, DOCK6 could not complete the growth, resulting in 953 complexes docked by both tools. The average RMSD values computed for the SB2012 complexes are shown in Table 1. The AMBER force field energy model resulted in better RMSDs, with an average of 1.01 Å and median in 0.49 Å (standard deviation 1.79 Å). The soft-core AMBER energy model resulted in an average RMSD of 2.01 Å, with median in 0.55 Å and standard deviation of 3.99 Å. A simple *t* test assuming that the variances are not equivalent reveals that there is a statistically significant difference between these two energy models in finding the best conformation for the ligand within the binding site ($p < 0.001$). Since the ligand is being docked against its own receptor, protein flexibility is not expected to play a role in this context. For the AMBER energy model, 91 % of the targets resulted in RMSD below 2.0 Å and 89 % were found below 1.5 Å, while for the soft-core energy model, 84 % are found below 2.0 Å and 82 % below 1.5 Å. Also, a rough comparison with a pure receptor-based tool shows that LiBELa works well on pose prediction in that scenario (Table 1). The average RMSD observed after ligand docking with DOCK6 was 5.6 Å with median in 4.86 Å and standard deviation of 4.74 Å. Correcting DOCK6 RMSD values using the closest heavy atom (RMSDm) results in an average RMSD of 2.09 Å with median in 0.78 Å and standard deviation of 2.70 Å. The success in achieving a low RMSD conformation (<2.0 Å for non-hydrogen atoms) increases from 36 to 68 % after this correction of DOCK6 poses. It is interesting to note that this success rate agrees well with the most recent publications from DOCK6 developers [29]. The results from the self-docking with SB2012 reveal that the introduction of a reference ligand in the hybrid strategy can be useful for guiding the docking process towards a final correct pose in most cases. The smoothed potential, which allows the sampling of an increased number of non-physical poses, results in increased RMSDs for the poses and reduced success rate. Finally, the global optimization of the final docking pose using the binding energy as the objective function retains the ligand within its correct conformation for the great majority of the targets evaluated in this test.

A deeper analysis of the failures for docking could involve a classification of the failures in sampling failure or scoring failure. Most of the binding energies computed with LiBELa after docking were more favorable than the initial energy. For the Amber FF model specifically, from the 953 complexes docked, 87 (9.1 %) had RMSD values >2.0 Å, characterizing a docking failure. However, for 85 complexes, the binding energy computed after docking was more favorable than the initial energy, suggesting that those failures (8.9 %) are more likely to be scoring failures, rather than sampling failures (Table 1). Two complexes (3HAD and 1BIR) had energies less favorable with RMSDs over 7 Å, suggesting that the correct conformation could not be achieved within the optimization time limit (30 s). The comparison of these results with the results obtained with Smoothed Amber energy model or with

Table 1 RMSD computed for docked compounds of the SB2012 dataset with two different scoring functions for LiBELA and the Grid Score for DOCK6

	LiBELa		DOCK6	DOCK6-RMSDm
	AMBER FF	Smoothed AMBER	Grid Score	Grid Score
Average (Å)	1.01	2.01	5.60	2.09
Median (Å)	0.49	0.55	4.86	0.78
Standard deviation (Å)	1.79	3.99	4.74	2.70
<3.0 Å	93.5 %	85.7 %	41.4 %	74.2 %
<2.0 Å	90.9 %	83.9 %	36.4 %	68.1 %
<1.5 Å	88.7 %	81.7 %	30.6 %	63.3 %
Sampling failure	0.2 %	5.7 %	23.7 %	12.9 %
Scoring failure	8.9 %	10.4 %	39.9 %	19.9 %
Docking failure	9.1 %	16.1 %	63.6 %	31.9 %

N = 953 and RMSDm is the minimum-distance heavy atom RMSD, as computed by DOCK6

DOCK6 shows that scoring failure plays is the common cause for failure, rather than sampling. This result highlights the need of better energy models for docking, that could take into account the role of the solvent [38], for example. In addition, the reason for the worse results observed with the smoothed potential is now evident: more time is spent in the sampling of non-viable conformations and less time is spent in the sampling of poses closer to the experimental ones.

In SB2012 test set the results are somewhat facilitated since the ligand is initially docked on itself in the hybrid approach. A more stringent test would be a cross-docking experiment, where the search ligand is docked against a different crystal structure and using a different reference ligand. We tested two different approaches in this cross-docking test using the Astex non-native dataset. In this dataset, a collection of *j* different ligands are provided linked to a reference (*i*) ligand or complex. The coordinates of the receptor *i* (R_i) were downloaded directly from the protein data bank and prepared using the DockPrep tool available in UCSF Chimera. The ligand *i* (L_i) was prepared using Chimera with AM1 atomic charges computed with antechamber. Similarly, each ligand *j* was prepared with Babel [19] and antechamber, using AM1 as the atomic charge model. The first approach consisted in the docking of the ligand (L_i) in its own receptor (R_i) using every ligand *j* as the reference ligand in LiBELa and taking the average RMSD over all *j* ligands. This test evaluates the ability of LiBELa to achieve a correct binding pose in self-docking even using a dissimilar ligand as the reference ligand. In a second and more stringent test, every ligand *j* was docked against receptor R_i using ligand L_i as a reference in LiBELa and the deviation of the final pose was compared to the experimentally determined conformation.

The results for the first experiment are given in Table 2. The average RMSD for each target (R_i) was computed (Supplementary Material) and the average over the entire set is shown in the table. For the Amber FF energy model, an average RMSD of 3.41 Å was observed (median of 1.00 Å), as compared to the average RMSD of 3.75 Å observed for the Smoothed Amber energy model (median of 1.40 Å). The results show that LiBELa can reach an experimental-like conformation (RMSD < 2.0 Å) in at least 60 % of cases even with dissimilar reference ligand. For the sake of comparison, we tested the ability of our Ligand-Based only algorithm to reproduce the ligand poses in this test. Interestingly, the ability to recover low-RMSD binding poses was reduced to 55 %, indicating that the energy optimization plays an important role in the achievement of the correct binding pose in LiBELa.

The results from the second experiment, summarized in Table 3, show that, averaging over the 530 ligand poses that were actually scored, an average RMSD of 4.28 Å was observed for the Amber FF energy model (median in 2.96 Å). Using the Smoothed Amber potential resulted in a slight improvement of the results, as indicated by the reduced median (2.51 Å), although the average RMSD was 4.37 Å, suggesting that a smoothed potential can actually compensate for some of the minor adaptations in receptor structure. Removing four targets that had average RMSD values over 7 Å, what can indicate deeper problems in the automatic receptor parametrization, for example, results in RMSD (averaged over the remaining set) of 3.85 Å with median of 2.66 Å for Amber FF energy model. In summary, the cross-docking test revealed that LiBELa can achieve correct poses even when dissimilar molecules are provided as a reference ligand and, most importantly, it is also efficient for the recovery of binding poses even when docking is done using different receptor structures. The comparison with the well-validated and receptor-based tool DOCK6 shows that, although the sampling of ligand conformation is reduced in LiBELa, as compared to DOCK6, the usage of a reference ligand in the ligand-based step of the hybrid approach drives the search to a sub-region of the search space that is close to the correct ligand binding site, resulting in good binding modes observed in different tests but without increasing the time spent on an exhaustive sampling, for example.

We analyzed the profile of the complexes docked in this cross-docking experiment trying to understand the main reasons for failures. In the AMBER FF model, from the 530 complexes, 313 resulted in the RMSDs above 2.0 Å, characterizing a docking failure (59.1 %). From those, 294 complexes, or 93.9 % showed binding energies below the binding energies of the original (already docked) poses, suggesting that a scoring failure plays a major role in these cases. The same analysis with the Smoothed potential revealed that scoring error seem to play a role in 85.5 % of the docking errors, i.e., in the poses where the RMSD is above 2.0 Å.

In order to be useful for typical screening campaigns, it is crucial for docking engines to score binding molecular interactions well enough to discriminate actual binders from non-binders. This scoring challenge can be evaluated in enrichment experiments, where a set of actual binders and decoy molecules are docked against a target and the enrichment of the binders is evaluated in receiver operating characteristic (ROC) curves. The selection of the decoys takes into account the charge distribution, the molecular weight, the logP and the solvent accessible area, to mention some of the parameters [32, 33].

Thirty-seven targets of the DUD dataset were used for enrichment analysis of binders against decoys using the two energy models (Table 4). Interestingly, the results observed with both energy models are indistinguishable when the average adjusted logAUC is examined. For the

Table 2 RMSD values for the self-docking experiment using the Astex non-native dataset

	LiBELa		LiBELa
	AMBER FF	Smoothed AMBER	Ligand-Based
Average (Å)	3.41	3.75	3.25
Median (Å)	1.00	1.40	1.65
Standard deviation (Å)	5.34	5.65	4.57
<3.0 Å	67 %	68 %	65 %
<2.0 Å	61 %	63 %	55 %
<1.5 Å	58 %	53 %	47 %

The RMSD values are averaged over the 530 docked conformations for the two energy models under evaluation and for the Ligand-Based only strategy

Table 3 RMSD values for the cross-docking experiment using the Astex non-native dataset

	LiBELa		DOCK6	DOCK6-RMSDm	DOCK6-RMSDh
	AMBER FF	Smoothed AMBER	Grid Score	Grid Score	Grid Score
Average (Å)	4.28	4.37	7.65	3.37	6.80
Median (Å)	2.96	2.51	6.83	1.71	5.87
Standard deviation (Å)	4.85	5.44	5.82	4.49	5.44
<3.0 Å	50.6 %	54.0 %	19.2 %	63.3 %	26.4 %
<2.0 Å	40.9 %	42.0 %	14.0 %	53.2 %	18.1 %
<1.5 Å	34.5 %	32.3 %	11.9 %	46.2 %	13.9 %

The RMSD values were averaged over the 530 docked conformations. The same complexes were docked with DOCK6. Again, heavy-atom RMSDm values are shown together with RMSDh (Hungarian symmetry-corrected heavy-atom RMSD) values for DOCK6

AMBER energy model, an average adjusted logAUC of 20.8 was observed, with median in 21.0 and standard deviation of 13.7 ($N = 37$). For the soft-core energy model, an average adjusted logAUC of 20.5 was found, with median in 22.0 and standard deviation in 14.4. For the sake of comparison, the average adjusted logAUC observed with DOCK6 grid score was 17.3 with median in 19.2 and standard deviation of 17.6, revealing that both energy models result in good enrichments. Is it worth to point out that the logAUC weights the enrichments spanning three decades (0.1–1.0 %, 1–10 %, 10–100 %) equally, so that the early enrichment is relevant to good logAUC values [36]. The enrichments for individual targets are shown in Table 3 and the curves are shown in the supplementary material.

We also evaluated the enrichment on a set of DUD-Enhanced (DUD-E) targets. Thirteen targets were selected from the database. Again, the differences between the AMBER energy model and the soft-core were negligible (Table 5). For the former, an average adjusted logAUC of 8.1 was achieved, with median in 5.5 and standard deviation of 9.6, while for the later an average adjusted logAUC of 8.5 with median in 5.6 and deviation of 9.1 were observed. Here, as in the DUD benchmark, there is a small advantage for soft-core energy model as compared to the AMBER energy model, that can be observed both in the averages as in the medians, but this small difference is unlikely to be significant. Using the same targets of DUD-E with DOCK6 resulted in an average adjusted logAUC of 5.9 with median in 8.4 and standard deviation of 12.3 (individual curves are shown in the supplementary material). Taken together the results provide here show that both energy models can be efficient in recovering actual binders, even when compared with an outstanding tool such as DOCK6.

Figure 2 shows the comparison among the enrichments observed with DOCK6 and LiBELa with AMBER and AMBER Soft-core energy models. Some interesting differences can be observed here. First, there is some complementarity in the DOCK6 results and LiBELa results. The targets AR, COX2 and HMGA on DUD, KITH on DUD-E and MR on both sets have poor enrichments with DOCK6, leading to negative adjusted logAUC values (enrichment worse than random) while using the hybrid docking embedding ligand structural information results in important improvement on enrichments. Since both DOCK6 and LiBELa use similar energy scores, the improvement is likely to be due to better search method. Since in LiBELa searching is driven by the match of structural features between the reference and the search ligand, we conclude that adding this information to docking helps to rescue actual binders in some scenarios. Of course,

Table 4 Enrichments for 37 DUD targets using AMBER and soft-core energy models, as implemented in LiBELa and DOCK6 Grid Score

#	Targets	AMBER		Soft-core		DOCK 6.6	
		Adjusted logAUC	$EF_{1\%}$ (%)	Adjusted logAUC	$EF_{1\%}$ (%)	Adjusted logAUC	$EF_{1\%}$ (%)
1	ACE	3.12	6.12	−1.80	2.04	2.41	8.16
2	ACHE	4.56	1.87	4.58	0.93	12.55	1.87
3	ADA	26.38	15.38	22.00	10.26	30.47	20.51
4	AMPC	24.73	42.86	26.25	42.86	17.63	19.05
5	AR	24.49	10.13	23.20	15.19	−4.17	3.80
6	CDK2	31.19	18.06	34.27	18.06	27.02	23.61
7	COMT	24.52	18.18	23.86	18.18	57.72	45.45
8	COX1	6.48	0.00	5.44	0.00	−2.89	0.00
9	COX2	29.32	27.93	29.46	29.81	−3.33	3.29
10	EGFR	25.41	26.11	27.22	28.00	13.11	21.68
11	ER AGONIST	19.08	8.96	14.67	8.96	16.86	22.39
12	ER ANTAGONIST	13.83	5.13	16.43	2.56	19.24	15.38
13	FGFR1	39.99	42.50	40.28	40.83	42.21	48.33
14	FXA	33.05	12.33	32.82	17.81	26.67	18.49
15	GPB	9.09	1.92	10.67	5.77	54.21	59.62
16	GR	13.20	10.26	13.04	11.54	−6.89	2.56
17	HIVPR	5.80	11.29	3.65	11.29	−6.07	3.23
18	HIVRT	7.44	9.30	10.29	11.63	19.72	18.60
19	HMGA	40.56	40.00	42.08	40.00	−5.10	8.57
20	HSP90	4.52	2.70	3.85	2.70	−5.38	2.70
21	INHA	7.43	9.30	4.75	9.30	2.33	1.16
22	MR	45.37	60.00	42.83	60.00	−1.74	0.00
23	NA	37.24	22.45	35.87	28.57	36.87	30.61
24	P38	20.96	18.72	22.81	19.16	15.18	20.26
25	PARP	19.13	8.57	16.17	5.71	21.00	14.29
26	PDE5	13.83	9.09	11.88	9.09	2.91	11.36
27	PDGFRB	0.93	4.12	−0.69	1.18	4.92	11.18
28	PNP	10.85	12.00	8.40	14.00	23.98	8.00
29	PPAR GAMMA	−1.23	4.76	−2.46	1.90	22.62	5.88
30	PR	2.78	7.41	2.63	7.41	−5.24	3.70
31	RXR ALPHA	25.17	20.00	28.50	20.00	23.25	20.00
32	SAHH	34.89	33.33	32.59	30.30	26.01	21.21
33	SRC	35.03	34.59	38.28	36.48	27.72	33.96
34	THROMBIN	48.45	43.06	50.08	43.06	34.95	19.44
35	TK	17.49	9.09	17.77	18.18	44.09	40.91
36	TRYPSIN	33.64	20.41	31.84	20.41	34.81	32.65
37	VEGFR2	30.81	29.55	34.22	30.68	20.80	23.86
	Average	20.8	17.8	20.5	18.2	17.3	17.5
	Median	21.0	12.0	22.0	15.2	19.2	18.5
	Standard deviation	13.7	14.4	14.4	14.7	17.6	14.6

LogAUC values are shown in the table together with enrichment factors at 1 % of decoys ($EF_{1\%}$). $EF_{1\%}$ is defined as the fraction of actives found at 1 % of decoys

the searching is actually being restricted to smaller fraction of the phase space that could be accessible in principle, when the match in similarity is used. A consequence is that for a number of targets, hybrid docking can result in worsening of the enrichments. This is observed for the targets GPB and PDGFRB in DUD, for example (Fig. 2), where DOCK6 results in better enrichment than LiBELa. It is also interesting to note that two of the targets that take

Table 5 Enrichments for selected DUD-E targets using AMBER and Soft-core energy models in LiBELa and DOCK6 Grid Score

#	Targets	AMBER		Soft-core		DOCK 6.6	
		Adjusted logAUC	$EF_{1\%}$ (%)	Adjusted logAUC	$EF_{1\%}$ (%)	Adjusted logAUC	$EF_{1\%}$ (%)
1	CP2C9	−0.83	1.09	−0.69	0.00	−3.87	0.55
2	CXCR4	−0.22	0.82	0.63	0.82	−5.25	0.00
3	GRIK1	8.69	3.29	8.99	1.32	9.97	4.61
4	MK10	1.35	2.15	2.35	0.54	−3.06	0.54
5	XIAP	7.56	4.65	6.72	2.33	9.88	6.20
6	MCR	3.96	6.22	5.64	7.77	−10.95	0.00
7	THB	8.25	3.59	12.08	4.79	–	–
8	HIVINT	3.66	2.37	3.70	1.42	6.94	6.16
9	KITH	3.23	0.00	3.88	0.00	−7.69	0.00
10	PUR2	33.42	7.96	32.04	3.98	30.79	1.49
11	LKHA4	8.73	2.46	8.89	2.05	17.06	6.97
12	PPARD	5.51	3.47	4.25	2.78	10.69	2.08
13	DYR	22.11	24.38	21.78	23.85	16.10	23.85
	Average	8.1	4.8	8.5	4.0	5.9	4.4
	Median	5.5	3.3	5.6	2.1	8.4	1.8
	Standard deviation	9.6	6.3	9.1	6.4	12.3	6.7

LogAUC and $EF_{1\%}$ are shown for each target

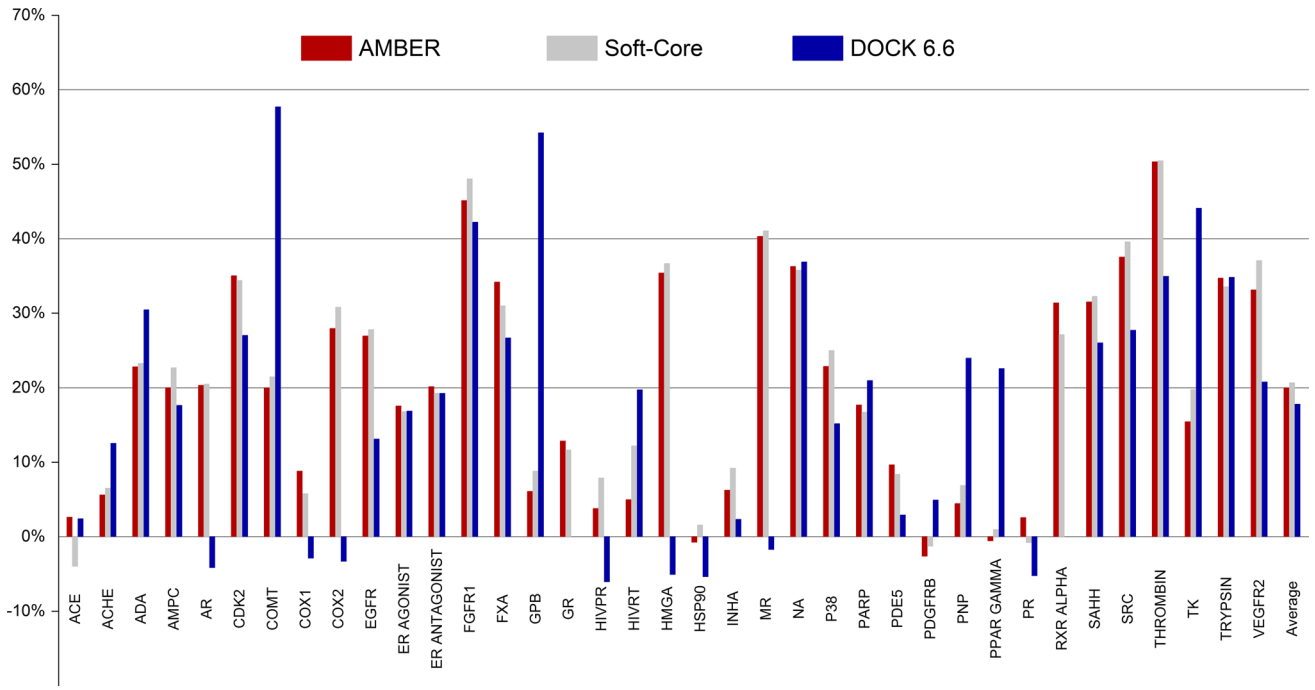

Fig. 2 Enrichments observed with LiBELa energy models and DOCK6

benefit of the hybrid approach are steroid receptors. These ligands recognize steroids with very high affinity in their buried pocket and have most of their known binders as steroid derivatives. This observation leaded us to hypothesize that many targets would be favored due to the higher similarity between binders and reference ligand. To assess the relevance of ligand similarity on the enrichment, we recomputed a ROC curve for the DUD dataset by computing an Hodgkin-like similarity index (SI) for each search molecule against the reference molecule, where

$SI = 2V_{AB}/(V_{AA} + V_{BB})$, with V computed as described in Eq. 1 [39–41]. The similarities are used to rank the database and no interaction energy is computed. Adjusted logAUC values were computed and resulted in an average logAUC of 8.3 with median in 9.3 and standard deviation of 9.4, revealing that the interaction energies, rather than the similarity among the ligands, seems to be the most relevant criteria for scoring in LiBELa. Similar results were also observed for DUD-E (data not shown).

Is there any advantage on introducing ligand similarity in the docking process? Our results show that the results obtained with our tool, LiBELa, are slightly better than the results obtained using the tool DOCK6, receptor pocket shape descriptors are used to guide the ligand docking. However, the differences are small and the results advantages are inverted for some targets, as shown in Tables 4 and 5. Still, as evidenced in Fig. 2, the results are rather complementary. In cases where pocket shape descriptors are not sufficient to guide the ligand placements, the introduction of ligand similarity seems to be introduce relevant structural information to achieve a global minimum in the binding surface. Our results then suggest that the introduction of ligand similarity into the docking pipeline can be a useful and interesting complementary approach to the traditional receptor-only—based approach for ligand docking. The implementation of this hybrid docking approach in LiBELa was found to be efficient both in terms of ligand processivity, i.e., ligands screened per minute, and in terms of reproduction of near native structures. Worthy of note, McGann [14] also observed, using a different algorithm, that the introduction of ligand structural information into HYBRID resulted in a slight improvement of average AUC in comparison with FRED, a similar algorithm that does not use the hybrid approach. In line with our observations, these results observed with different datasets suggest that ligand structural information can be of advantage in molecular docking.

For the energy models evaluated, interesting conclusions can be drawn. First, using smoothed potentials resulted in larger deviations from native structures in the binding pose evaluation (Table 1), as expected, since a larger fraction of the phase space is energetically allowed by the smoothening. On the other hand, no clear advantage could be observed in terms of enrichment of actual binders against non-binders or in terms of computation speed. Therefore, different from what observed previously with different approaches for soft-docking, our results do not suggest any relevant advantage for using a smoothed potential, even for single receptor conformations, as used in this work. In the cross-docking scenario, probably the most challenging scenario, the smoothed potential resulted in reduced RMSD values in the median. However, the overall success rate is smaller than the rate observed in the pure FF model. The reason for the failure of smoothed potentials in leading to increased convergence towards a global minimum in the binding surface is understood by the lack of selectivity for actual binders. The smoothed potential facilitates the achievement of a minimum in the binding energy for ligands but decoys also get more permissive binding poses that are not clearly discriminated in this energy model. Hence, after fine-tuning of smoothening parameters, the results obtained with soft-core potential at best can reproduce the results observed with force field parameters, as shown in Tables 4 and 5.

One can argue whether the results observed with this hybrid tool is mainly due to a similarity screening with a 'flavor' of a structure-based approach. In order to assess the influence of ligand similarity on the observed enrichments, we repeated the docking using similar parameters but without using the receptor structure and the binding energy and ranking the screened ligands by their similarity computed against the reference ligand in each target. The results observed for DUD showed that the average enrichment is reduced in 50 % as compared to the hybrid approach, revealing that the combined strategy, rather than the similarity itself, plays the important role in our analysis. The energy optimization was also shown to be outstanding for ligand pose, as shown in Table 2.

Broccatelli and Brown, observed that in their combination of a ligand-based strategy with a receptor-based strategy for ligand screening, there was payoff between exploration and exploitation, meaning that the docking success is increased for ligands with higher Tanimoto similarity [42]. Our previous analysis of the ligand similarity using three-dimensional descriptors showed that higher success rate in the enrichment of DUD binders using a ligand-based approach only was not correlated with average Tanimoto similarities [15], revealing that three-dimensional descriptors can capture structural features beyond topology or atomic connectivity. So, we would not expect hybrid-docking to be importantly limited to scenarios where screened and reference ligands are topologically similar, although the bias in this scenario is obvious. Also, it is also important to take into account that the receptor structure itself carry a structural bias to the ligand structure it was crystallized with.

In line with our observations, Svensson and coworkers reported the results of typical virtual screening campaigns with a subset of DUD targets using different fusion strategies for data coming from structure-based (docking) and ligand-based strategies [43]. The authors observed that data fusion consistently resulted in better enrichments than any strategy alone.

In conclusion, we show here that the combination of ligand- and receptor-based strategies can be efficient in ligand docking, driving the initial ligand conformation

towards a neat native conformation. Scoring these conformations is still an important challenge and the results evidence the lack of any clear advantage of soft-docking against the typically used force field energy model. The introduction of corrections to model the solvent effects are expected to result in even better enrichments bringing the energy model closer to reality.

Acknowledgments The authors thank Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), and to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial support through grants 2014/06565-2 and 485950/2013-8. HSM also thanks CNPq for her PhD stipend. We also thank Dr. Paul Hawkings and João Victor Cunha for the fruitful discussions.

References

1. Leach AR (2001) Molecular modelling: principles and applications, 2nd edn. Prentice Hall, Upper Saddle River
2. Leach A, Shoichet B, Peishoff C (2006) J Med Chem 49(20):5851
3. Zwanzig R (1954) J Chem Phys 22(8):1420
4. Kollman P (1993) Chem Rev 93(7):2395
5. Jorgensen W, Thomas L (2008) J Chem Theory Comput 4(6):869
6. Kollman P, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case D, Cheatham T (2000) Acc Chem Res 33(12):889
7. Kuhn B, Kollman P (2000) J Med Chem 43(20):3786
8. Kuntz ID, Meng EC, Shoichet BK (1994) Acc Chem Res 27(5):117
9. Doman T, McGovern S, Witherbee B, Kasten T, Kurumbail R, Stallings W, Connolly D, Shoichet B (2002) J Med Chem 45(11):2213
10. Verkhivker G, Bouzida D, Gehlhaar D, Rejto P, Freer S, Rose P (2002) Curr Opin Struct Biol 12(2):197
11. Wang J, Verkhivker GM (2003) Phys Rev Lett 90(18). doi:10.1103/PhysRevLett.90.188101
12. Ferrari A, Wei B, Costantino L, Shoichet B (2004) J Med Chem 47(21):5076
13. Hawkins P, Kelley B, Warren G (2014) J Chem Inf Model 54(5):1339
14. McGann M (2012) J Comput Aided Mol Des 26(8):897
15. VazdeLima LAC, Nascimento AS (2013) Eur J Med Chem 59:296
16. Grant JA, Pickup BT (1995) J Phys Chem 99(11):3503
17. Grant JA, Gallardo MA, Pickup BT (1996) J Comput Chem 17(14):1653
18. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25(9):1157
19. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) J Cheminform 3:33
20. Andreani R, Birgin EG, Martinez JM, Schuverdt ML (2008) Math Program 111(1–2):5
21. Birgin EG, Martinez JM (2008) Optim Methods Softw 23(2):177
22. Andreani R, Birgin EG, Martinez JM, Schuverdt ML (2007) SIAM J Optim 18(4):1286
23. Jones DR, Perttunen CD, Stuckman BE (1993) J Optim Theory Appl 79(1):157
24. Jhonson SG (2015) The NLopt nonlinear-optimization package. http://abinitio.mit.edu/nlopt
25. Meng EC, Shoichet BK, Kuntz ID (1992) J Comput Chem 13(4):505
26. Luty BA, Wasserman ZR, Stouten PFW, Hodge CN, Zacharias M, McCammon JA (1995) J Comput Chem 16(4):454
27. Pattabiraman N, Levitt M, Ferrin T, Langridge R (1985) J Comput Chem 6(5):432
28. Mukherjee S, Balius T, Rizzo R (2010) J Chem Inf Model 50(11):1986
29. Allen WJ, Balius TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, Case DA, Kuntz ID, Rizzo RC (2015) J Comput Chem 36(15):1132
30. Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW (2008) J Chem Inf Model 48(11):2214
31. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) J Comput Chem 25(13):1605
32. Huang N, Shoichet BK, Irwin JJ (2006) J Med Chem 49(23):6789
33. Mysinger M, Carchia M, Irwin J, Shoichet B (2012) J Med Chem 55(14):6582
34. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) J Chem Inf Model 52(7):1757
35. Irwin JJ, Shoichet BK (2005) J Chem Inf Model 45:177
36. Mysinger MM, Shoichet BK (2010) J Chem Inf Model 50(9):1561
37. Brozell S, Mukherjee S, Balius T, Roe D, Case D, Rizzo R (2012) J Comput Aided Mol Des 26(6):749
38. Shoichet BK, Leach AR, Kuntz ID (1999) Proteins Struct Funct Genet 34(1):4
39. Good AC, Hodgkin EE, Richards WG (1992) J Comput Aided Mol Des 6(5):513
40. Good AC, Hodgkin EE, Richards WG (1992) J Chem Inf Comput Sci 32(3):188
41. Hodgkin EE, Richards WG (1987) Int J Quantum Chem 32:105
42. Broccatelli F, Brown N (2014) J Chem Inf Model 54(6):1634
43. Svensson F, Karlen A, Skold C (2012) J Chem Inf Model 52(1):225