J Comput Aided Mol Des (2009) 23:143–152
DOI 10.1007/s10822-008-9246-z

An improved scoring function for suboptimal polar ligand complexes

Giovanni Cincilla · David Vidal · Miquel Pons

Received: 19 December 2007 / Accepted: 13 September 2008 / Published online: 9 October 2008

Abstract Learning strategies can be used to improve the efficiency of virtual screening of very large databases. In these strategies new compounds to be screened are selected on the basis of the results obtained in previous stages, even if truly good ligands have not yet been identified. This approach requires that the scoring function used correctly predicts the energy and geometry of suboptimal complexes, i.e. weak complexes that are not the final solution of the screening but help direct the search toward the most productive regions of chemical space. We show that a small modification in the treatment of the solvation of polar atoms corrects the tendency of the original Autodock 3.0 scoring function to bury ligand polar atoms away from solvent, even if no complementary groups are present in the target and improves the performance of Autodock 3.0 and 4.0 in reproducing the experimental docking energies of weak complexes, resembling the suboptimal complexes encountered in the intermediate stages of virtual screening.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-008-9246-z) contains supplementary material, which is available to authorized users.

G. Cincilla · D. Vidal · M. Pons
Laboratory of Biomolecular NMR, Institute for Research in Biomedicine, Science Research Park, Josep Samitier 1-5, 08028 Barcelona, Spain

G. Cincilla · D. Vidal · M. Pons (✉)
Departament de Química Orgànica, Universitat de Barcelona, Martí i Franquès 1-11, 08028 Barcelona, Spain
e-mail: mpons@ub.edu

Present Address:
D. Vidal
CHEMOTARGETS S.L., Av. Dr. Aiguader 88, 4a Planta, 08003 Barcelona, Spain

Keywords Docking · Drug design · Virtual screening · Scoring function · Solvation

Introduction

Virtual screening has the power to substantially improve the efficiency of experimental screening by identifying sets of candidates that have a higher probability than average to bind a target [1–4]. Docking algorithms offer reasonable estimates of the binding energy to targets of known structure. However, with the widespread use of combinatorial chemistry and virtual chemistry tools, the size of chemical libraries is growing exponentially to a point that an exhaustive exploration is no longer possible even with the most efficient docking algorithms presently available. The use of heuristic search methods exploring only a portion of the total chemical library is a realistic alternative. Genetic algorithms focus on the evolution of a population in which specific features, encoded as genes, are combined and selectively accumulated according to some fitness rules that determine which individuals are maintained or eliminated and how to select new individuals [5].

In this context, we have recently described the massive processing algorithm (MPA), a high throughput virtual screening algorithm [6]. MPA is an evolutionary based algorithm that combines receptor based scoring with similarity based search methods [7, 8] using a learning strategy approach [3, 9]. In particular MPA uses Autodock 3.0 [10], the most cited docking software [11], and text based (LINGO) similarity searches [12]. Autodock 3.0 had been previously modified in our group by adapting it to parallel processing and by including an enhanced genetic algorithm [13]. In the MPA approach, the original Autodock's estimated binding free energy of the best pose is used as fitness

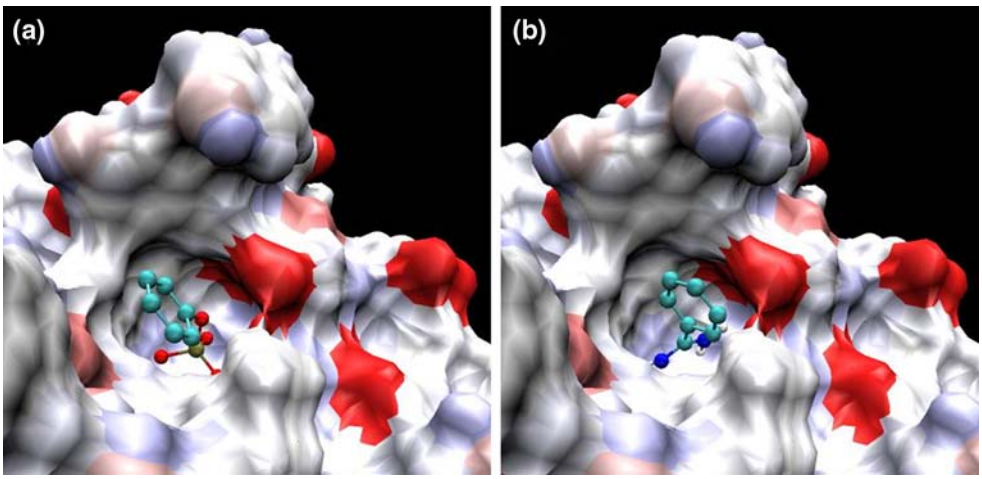

Fig. 1 Orientation of phosphoric acid phenyl ester (**a**) ($\Delta G_{bind} = -7.58$ kcal/mol) and benzamidine (**b**) ($\Delta G_{bind} = -5.36$ kcal/mol) complexed with *low molecular weight protein tyrosine phosphatase* calculated with the original scoring function of Autodock 3.0

function to rank each individual in a population of molecules. The ranking is then used to select the next generation of molecules by considering the similarity of all the molecules in the database to the "best" individuals in the present generation.

However, during the initial stages of an MPA search applied to an unbiased population, even the best individuals in each population are unlikely to be optimal ligands for the target of interest and the performance of the scoring function for suboptimal ligands may be a matter of concern. Standard scoring functions have been developed to reproduce the binding energy and geometry of optimal complexes found in crystal structures and suboptimal complexes found during an MPA search may contain unusual features, not present in the complexes with optimal ligands, but that represent useful guides in the search of better ligands in the "chemical space".

The analysis of recent virtual screening projects carried out in our group using the standard Autodock 3.0 scoring function as implemented in our MPA methodology has uncovered some intrinsic Autodock 3.0 scoring function limitations. Some of them have already been reported [14], and have been improved in the recently released Autodock 4.0 version [15]. In particular, Autodock 3.0 shows a tendency to overestimate predicted affinities of weak complexes. This is the result of unrealistic binding modes, i.e. burying polar groups in hydrophobic or non-complementary polar sites resulting in the unnatural pairing of two hydrogen bond acceptors or donors.

The latter situation can be exemplified by the docking of benzamidine to a bovine *low molecular weight protein tyrosine phosphatase* (*lmwPTP* 1DG9). The *lmwPTP* active site has a marked H-bond donor capability able to complex a phosphate group. Consistently, a ligand such as phosphoric acid phenyl ester is correctly predicted to form complexes with the phosphate group inside the active site, forming H-bonds with the H-bond donor groups of the protein. In contrast, although no H-bond is expected between the binding pocket groups and the polar group of benzamidine, the standard Autodock 3.0 scoring function also predicts that benzamidine is docked with its amidine group inside the pocket (see Fig. 1). In the absence of H-bonds with the protein, one would expect the amidine group to be oriented outside the pocket forming H-bonds with the solvent.

The complementary situation is observed in the binding of benzamidine and benzoic acid to *factor Xa*, with H-bond acceptor groups in the binding site. The benzamidine polar group is experimentally and computationally found forming H-bonds with the site of *factor Xa*. However, the polar portion of the benzoate anion, with a H-bond acceptor character is expected to remain solvated outside the binding pocket not interacting with factor Xa active site. However, the lowest energy pose of benzoate in the Autodock 3.0 scoring function presents a desolvated carboxilate group buried inside the protein cativity although there is not any compensating interaction with the protein polar groups (see Fig. 2).

These examples show that the standard Autodock 3.0 scoring function, among other related problems, underestimates the cost of desolvation of polar groups in the ligand, at least when they are not compensated by the formation of hydrogen bond with the target. The problem can be traced to the treatment of solvation in Autodock 3.0 that includes a constant term added to each point in the grid of atoms able to form H-bonds, to model destabilization due to desolvation of polar atoms. This approach does not take into account the relative positions of the polar atoms in the ligand and the protein and the possibility that a polar group remains solvated if it is not in direct contact with the

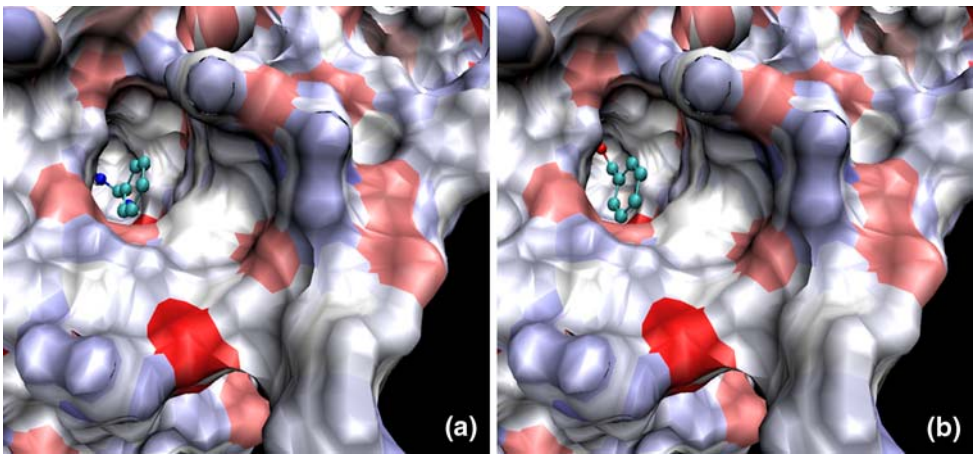

Fig. 2 Orientation of benzamidine (**a**) ($\Delta G_{bind} = -7.06$ kcal/mol) and benzoate (**b**) ($\Delta G_{bind} = -5.13$ kcal/mol) complexed with Factor Xa calculated with the original scoring function of Autodock 3.0

protein. The problem is minimized in cases of protein–ligand complementarity, typical of the best ligands but it can be a real problem when evaluating the relative affinity of suboptimal complexes leading to false positives that would direct the search algorithm towards unproductive regions of the chemical space.

In addition, the standard 3.0 scoring function does not differentiate between "polar" heteroatoms with "acceptor" or "acceptor/donor" H-bond character and "non-polar" heteroatoms (including those that have exclusively H-bond donor character or no tendency to participate in H-bonds). In order to overcome these limitations we have modified the solvation treatment by:

(i) Removing the constant E_{hb} term associated to the ligand polar atoms.
(ii) Differentiating between "polar" heteroatoms (that we denote by N, O, S) and "non-polar" heteroatoms (represented by n, c, o) on the basis of atom hybridization and connectivity.
(iii) Including a ΔG^{P}_{sol} term describing the desolvation of polar heteroatoms using the Stouten method [16], consistent with the treatment used by Autodock 3.0 for non-polar groups. This method takes into account the position of the ligand polar groups with respect to the protein and its associated hindrance.

For the implementation of these modifications we performed a minimal reparametrization of the standard scoring function using 189 protein–ligand complexes from the AffinDB database [17–19]. We show that these modifications indeed provide the expected orientation of ligand polar groups that do not form H-bonds with the target while preserving the geometry and binding affinity of the complexes used in the calibration of the standard Autodock 3.0 scoring function. In addition we have compared the ability of the modified and original scoring functions to reproduce experimental binding energies in a wide range of affinities and we show that the modified treatment provides a more faithful representation of intermediate affinity ligands, that can be taken as a proxy of the suboptimal ligands that guide the MPA virtual screening process.

Methods

The original and modified free energy equations of Autodock 3.0 are shown below together with the one of Autodock 4.0:

$$\begin{aligned}\Delta G_{bind} = {} & C_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)\\ & + C_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}+E_{hbond}\right)\\ & + C_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}}+C_{tor}N_{tor}\\ & + C_{sol}\sum_{i_C,j}S_i V_j e^{\left(-r_{ij}^2/2\sigma^2\right)}\end{aligned}$$

Equation 1: Original binding free energy equation of Autodock 3.0.

$$\begin{aligned}\Delta G_{bind} = {} & C_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)\\ & + C_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}\right)+C_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}}\\ & + C_{tor}N_{tor}+C^{p}_{sol}\sum_{i_{N,O,S,H},j}S_i V_j e^{\left(-r_{ij}^2/2\sigma^2\right)}\\ & + C_{sol}\sum_{i_C,j}S_i V_j e^{\left(-r_{ij}^2/2\sigma^2\right)}\end{aligned}$$

Equation 2: Modified binding free energy equation of Autodock 3.0.

$$\Delta G_{bind} = C_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + C_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + C_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + C_{tor} N_{tor} + C_{sol} \sum_{i,j} \left(S_i V_j + S_j V_i \right) e^{\left(-r_{ij}^2 / 2\sigma^2 \right)}$$

Equation 3: Binding free energy equation of Autodock 4.0.

In Eqs. 1–3 C_{vdw}, C_{hbond}, C_{elec}, C_{tor} and C_{sol} are the weights of the respective terms, A_{ij}, B_{ij}, C_{ij}, D_{ij} are atom type pair specific terms, r_{ij} is the interatomic distance, E(t) correspond to a directional weight based on the angle t between the probe and the target atom, that gives to the H-bond term the character of "directional", N_{tor} is the number of rotatable bonds, S_i and V_i are solvation parameters (see below), and σ is a distance weighting function set to 3.5 Å.

In contrast to a previously reported modification of Autodock 3.0 [20], the changes in the scoring function were not restricted to the weighting coefficients but involved the functional form of the terms describing desolvation.

The main differences between the original and the modified binding free energy equation of Autodock 3.0 are the removal of the constant E_{hbond} in the second term and the inclusion of an additional solvation term with the same functional form of the original one but including the polar heteroatoms. The inclusion of the constant term in the standard Autodock 3.0 scoring function increases the tendency of ligand polar atoms to interact with the protein. This enhanced interaction could partially mimicks the effect of protein side chain flexibility providing alternative hydrogen bonding partners for the ligand atoms. However, the constant hydrogen bond term has the undesirable effect of forcing un-natural docking poses. The differentiation between "polar" and "non-polar" heteroatoms was done considering hybridization and connectivity, including hydrogen atoms. "Polar" heteroatoms are those possessing an acceptor or acceptor/donor capability with respect to H-bond formation and are assigned a capital letter symbol. Hydrogen atoms connected to any oxygen, nitrogen or sulfur atoms are considered as "polar" hydrogens. Correspondences between polar and nonpolar heteroatoms and mol2 atom types are shown in Table 1.

New solvation parameters (S_i and V_i) were generated for all the polar atom types (N, O, S, H) starting from the values used by "addsol" program tool included in the Autodock 3.0 package used to assign solvation parameters to the protein atoms. These parameters were optimized iteratively using a set of 13 out of 30 complexes from the original Autodock 3.0 parametrization set that do not involve direct interaction between ligand and metal groups in the binding site (listed as supplementary material). In this first stage, the coefficient for the solvation of polar (C_{sol}^{P}) and apolar (C_{sol}^{A}) were assigned the same original C_{sol} value ($C_{sol} = 0.1711$) and only the C_{hbond} coefficient was modified from its original value of 0.0656 to 0.1312 in order to compensate for the removal of the constant term. The optimized new solvation parameter values are shown on Table 2. The correlation between the predicted binding free energies of 13 complexes using the modified and the standard solvation treatment is shown in Fig. 3 ($R^2 = 0.94$).

In a later stage, the C_{sol}^{P} coefficient was optimized using a set of 189 complexes extracted from the AffinDB database keeping the previously optimized solvation parameters (S_i and V_i) fixed. The set of 189 complexes was generated by selecting only those complexes involving ligands with molecular weights less than 600 Da and no contacts between ligand and metal centers in protein binding sites (see supplementary material). The experimental structures were used. In some cases, local refinement was applied using the local search of standard Autodock 3.0 but the refined structures never had RMSD larger than 0.2 Å with respect to the crystal structure. After selection, all 189 protein–ligand complexes were examined and the protonation state of ionizable groups at pH 7.0 was determined using the cxcalc module of the JChem software which calculates the major microspecies at a given pH [21]. Protonation states for amino acid side chains were assigned using the standard conventions, glutamic, aspartic and terminal carboxylic acids were set to carboxylates and arginine, lysine and terminal amino groups were protonated. In addition, histidine side chains were maintained neutral and the assignment of the tautomeric form was carried out manually depending on the surrounding amino acid chains. On the other hand, crystallographic waters involved in hydrogen bond interactions with the ligands were kept and hydrogens added were using the MAB algorithm as implemented in Moloc [22]. Optimization of C_{sol}^{P} was carried out by applying an iterative process. In this process, 85% of the complexes in the database were selected as training set and the remaining 15% as validation set. The optimized coefficient value of C_{sol}^{P} as well as the value for the rest of coefficients are given in Table 3. We refer to the original Autodock 3.0 scoring function as Original Solvation Treatment (OST) and the modified version as Modified Solvation Treatment (MST).

The MST Autodock 3.0 was developed independently of Autodock 4.0. The recently released new version of Autodock has in fact implemented, among other improvements, a similar treatment of solvation by replacing the constant desolvation term by a specific estimation of the solvation of polar atoms based on the Stouten method. A larger number

Table 1 Heteroatom classification

Mol2 atom type	Description	Bound atoms	Bound H	HB-character	Character assigned
N.1	Nitrogen sp	1	No	Acceptor	N
N.2	Nitrogen sp2	2	Yes	Acceptor/donor	N
N.2	Nitrogen sp2	2	No	Acceptor	N
N.2	Nitrogen sp2	3	Yes	Donor	n
N.2	Nitrogen sp2	3	No	–	n
N.4	Nitrogen sp3 positively charged	3	Yes	Acceptor/donor	N
N.4	Nitrogen sp3 positively charged	3	No	Acceptor	N
N.4	Nitrogen sp3 positively charged	4	Yes	Donor	n
N.4	Nitrogen sp3 positively charged	4	No	–	n
N.am	Nitrogen amide	3	Yes	Donor	n
N.am	Nitrogen amide	3	No	–	n
N.ar	Nitrogen aromatic	2	No	Acceptor	N
N.ar	Nitrogen aromatic	3	Yes	Donor	n
N.ar	Nitrogen aromatic	3	No	–	n
N.3	Nitrogen sp3	3	Yes	Donor	n
N.3	Nitrogen sp3	3	No	–	n
N.pl3	Nitrogen trigonal planar	3	Yes	Donor	n
N.pl3	Nitrogen trigonal planar	3	No	–	n
O.3	Oxygen sp3	2	Yes	Acceptor/donor	O
O.3	Oxygen sp3	2	No	Acceptor	O
O.2	Oxygen sp2	1	No	Acceptor	O
O.co2	Oxygen in carboxylate and phosphate groups	1	No	Acceptor	O
O.co2	Oxygen in carboxylate and phosphate groups	2	Yes	Donor	o
S.3	Sulfur sp3	2	Yes	Acceptor/donor	S
S.3	Sulfur sp3	2	No	Acceptor	S
S.3	Sulfur sp3	3	No	–	s
S.3	Sulfur sp3	4	No	–	s
S.2	Sulfur sp2	1	No	Acceptor	S
S.O	Sulfoxide sulfur	3	No	–	s
S.O2	Sulfone fulfur	4	No	–	s

Table 2 Volumes and solvation parameters for the original and the modified (boldface) model

Atom type	OST		MST	
	V	S	V	S
C	12.77	4.0	12.77	4.0
A	10.80	0.6	10.80	0.6
N	–	–	**7.00**	**−8.0**
O	*	*	**9.00**	**−12.0**
S	–	–	**19.90**	**−5.4**
H	*	*	**1.00**	**−9.5**

V is the solvation volume and S is the solvation parameter. Atom types: C is aliphatic carbon, A is the aromatic carbon, N, O, S, H correspond to polar nitrogen, oxygen, sulphur and hydrogen, respectively. (*) Hydrogen and oxygen atoms did not have either solvation volume or solvation parameter in the original model

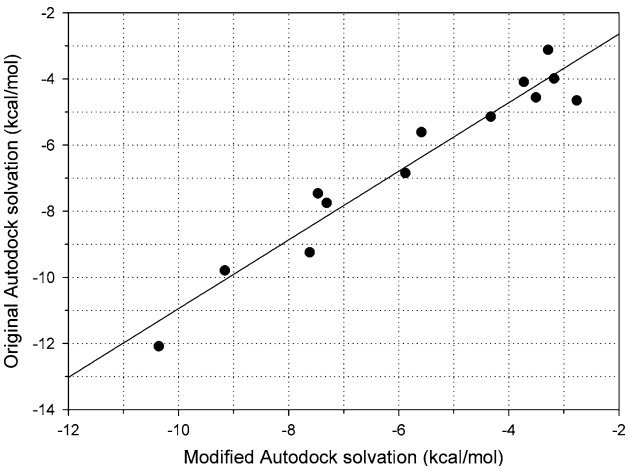

Fig. 3 MST versus OST docking energy comparison of the 13 complexes in the original Autodock 3.0 calibration set [see supplementary material]

Table 3 Coefficient of the original and modified (boldface) Autodock 3.0 scoring function

Coeff.	OST value	MST value
C_{vdw}	0.1485	0.1485
$\mathbf{C_{hbond}}$	**0.0656**	**0.1312**
C_{elec}	0.1146	0.1146
C_{tors}	0.3113	0.3113
C_{sol}^{A}	0.1711	0.1711
$\mathbf{C_{sol}^{P}}$	–	**0.2497**

of new atoms types have been defined in Autodock 4.0 and, in contrast to our modification of Autodock 3.0, the partial charge of each atom is used explicitly in the solvation treatment. The modified version can be obtained from the authors by authorized Autodock 3.0 users. The limited modifications introduced in the MST version of Autodock 3.0 preserves its full integration with the present implementation of the MPA algorithm.

In the comparison of the original and modified methods the same standard docking parameters were used: grids were built using Autogrid 3.0 with a grid space of 0.3 Å and a smoothing paramether of 0.5. Ligand rotatable bonds were assigned using Autotors and a previously described home written script [6].

Comparison between MST and OST scoring functions of Autodock 3.0 was carried out using the previously described optimized version of Autodock 3.0 that uses a parallelized search algorithm [13]. Each docking experiment consisted in a variable number of repetitions of a basic docking run. The basic docking run used a population of 400 divided in 16 niches of 25 molecules each. The maximum number of energy evaluations in each basic run was of 160000 and the maximum number of generations was 1000. At the end of each run, solutions were clustered using a similarity threshold of 0.8 Å and the lowest energy cluster was taken as the solution for the individual run. The basic run was repeated from 5 to 120 times depending on the number of rotatable bonds in the ligand (the total maximum number of energy evaluation for every complex varied between 800000 and 19200000).

Autodock 4.0 experiments were carried out using ten basic runs with the standard searching parameters set by AutoDockTools with a "medium" maximum number of energy evaluation. The settings use for every run a population of 150, a maximum number of energy evaluation of 2500000 and a maximum number of generation of 27000 (the total maximum number of energy evaluation for every complex was 25000000).

Finally, for both MST, OST and Autodock 4.0 experiments the best solution was taken as the result with the lowest binding energy.

Results

A plot of the experimental and calculated energies for the training and the validation sets are shown in Fig. 4. The correlation coefficients are 0.32 and 0.31 for the MST and OST, respectively. The upper and lower tendency lines correspond to the 95% confidence interval, calculated using sigmaplot10.0.

Figure 5 compares the orientation of the lower energy poses predicted by the MST of the complexes of phosphoric acid phenyl ester and benzamidine with *lmwPTP*. The modified scoring function indeed provides the expected different orientation for the two ligands and the polar group of benzamidine is found solvated outside the cavity while the phosphate group is forming the expected hydrogen bond with the complementary protein polar groups.

The change in orientation of the benzamidine ligand, but not of the phosphoric acid phenyl ester when the modified solvation term is used is a consequence of the changes in the relative energies: While the predicted binding free energy of the phosphoric acid ester using the MST (−7.03 kcal/mol) is only slightly higher than with the OST (−7.58 kcal/mol) with the same geometry, the energy of the benzamidine pose that buried the polar group and was the lowest energy minimum for this complex in the OST (−5.39 kcal/mol) increased by almost 3 kcal/mol with the MST (−2.44 kcal/mol). As a result the expected orientation, with the non-interacting polar atoms solvated, became the predicted conformation.

Figure 6 shows the best poses obtained with the MST for the complexes of benzoic acid and benzamidine with *factor Xa*. Consistently, the benzamidine polar group is now found forming a hydrogen bond inside the cavity of the protein while the carboxilate group of benzoate remains solvated outside the binding pocket, which seems to be a more realistic pose, although no experimental evidence has been described. The difference in binding energy of *factor Xa* to the complementary and the non-complementary ligands is increased using the MST, leading to a better discrimination.

MST–OST model comparison

In order to evaluate the ability of the modified scoring function to reproduce the experimental geometry of real complexes we performed docking experiments of the 189 protein–ligand pairs with known complex geometries using the standard and the modified scoring function under otherwise identical docking conditions and we compared the RMSD differences between the lowest energy solution and the crystallographic structure. The average RMSD of the complex structures calculated with the modified solvation

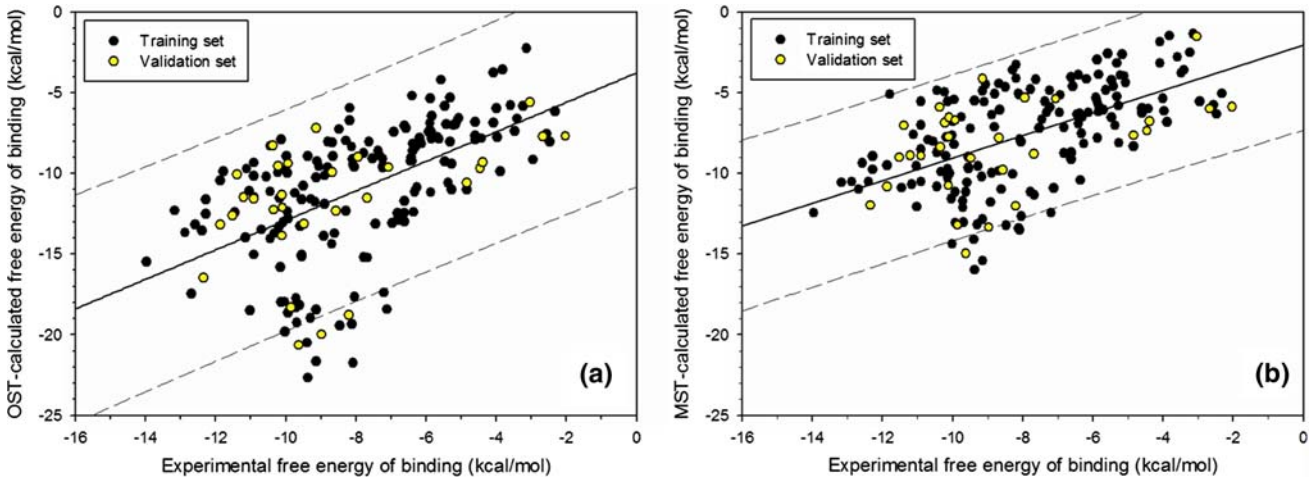

Fig. 4 Experimental and calculated binding free energies using the original (**a**) and modified (**b**) solvation treatments. The energies are calculated with the experimental geometry of the complexes. Open symbols correspond to the validation set not used to develop the model

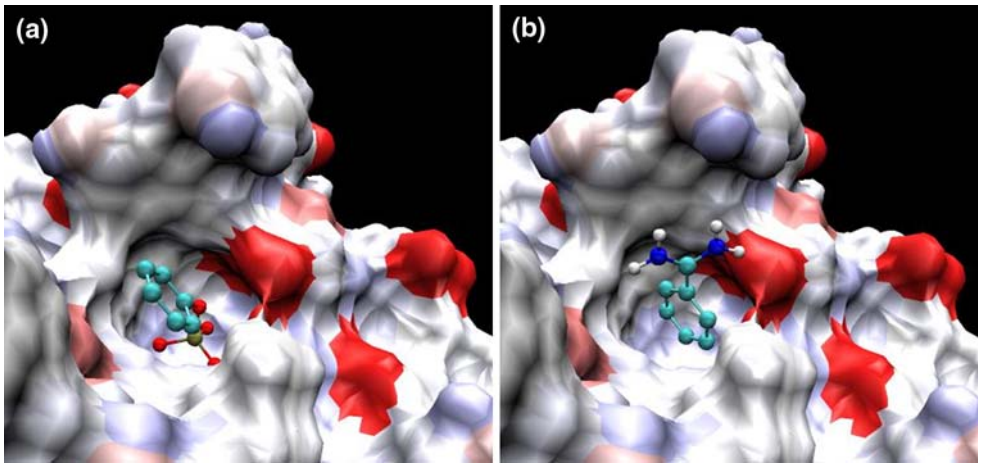

Fig. 5 Orientation of phosphoric acid phenyl ester (**a**) ($^{MST}\Delta G_{bind} = -7.03$ kcal/mol) and benzamidine (**b**) ($^{MST}\Delta G_{bind} = -3.90$ kcal/mol) complexed with *low molecular weight protein tyrosine phosphatase* calculated with the modified Autodock 3.0 scoring function

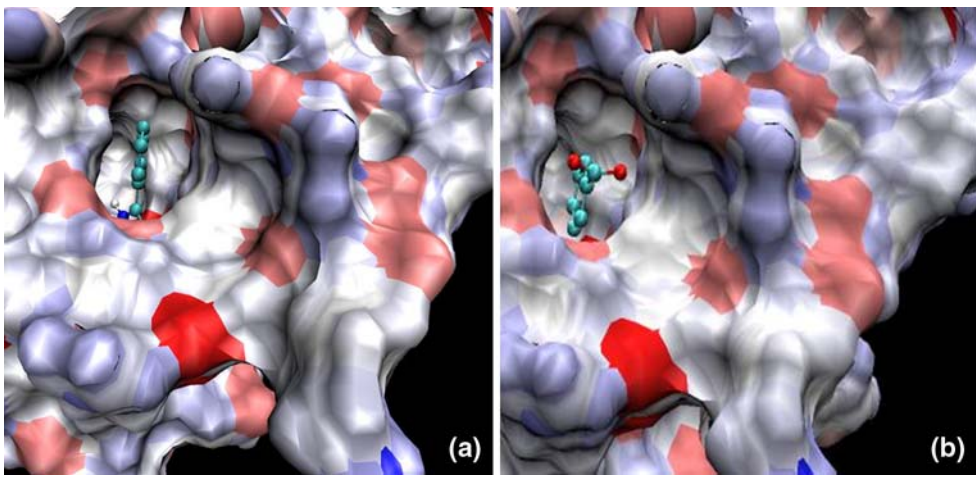

Fig. 6 Orientation of benzamidine (**a**) ($^{MST}\Delta G_{bind} = -6.19$ kcal/mol) and benzoate (**b**) ($^{MST}\Delta G_{bind} = -3.90$ kcal/mol) complexed with *Factor Xa* calculated with the modified Autodock 3.0 scoring function

treatment is 1.20 Å ($\sigma = 1.25$). For the complexes computed with the original Autodock 3.0 scoring function the average RMSD is 1.22 Å ($\sigma = 1.08$). Individual RMSD differences are listed in the supplementary material.

As the main difference between the two versions of Autodock 3.0 is the treatment of the solvation of polar heteroatoms, we have compared the binding energies computed with the original and modified treatments using

the X-ray structure of the complex and plotted the difference as a function of the number of "polar" heteroatoms. The plot is shown in Fig. 7. A linear correlation is observed between the *maximum* difference and the number of polar atoms present in the ligand. However, as the solvation contribution in the MST depends on the particular environment of these atoms, different ligands with the same number of polar heteroatoms show different degrees of correction with respect to the predictions of the OST and the points in Fig. 7 are located within a triangular area.

The differences between experimental and calculated binding energies using both models and the experimental geometry are plotted as a function of the number of polar heteroatoms in Fig. 8. The modified solvation treatment clearly corrects the tendency of the original Autodock 3.0 scoring function to overestimate the binding affinities, especially for ligands with a large number of polar heteroatoms leading to a much better agreement between theory and experiment.

One of the motivations of this work is to improve the performance of Autodock for high energy complexes. Weak complexes are extensively found during virtual screening. These complexes are not the final solution but are used by the MPA genetic algorithm to direct the search toward better complexes. Figure 9 compares the performance of the original and MST versions of Autodock 3.0 in reproducing the experimental binding energies of a collection of experimentally studied complexes that includes a wide range of stabilities. In both versions of Autodock 3.0, the binding affinities are underestimated for the stronger complexes and overestimated for the weaker complexes. However, the modified version of Autodock gives more realistic estimates in a wider range of affinities.

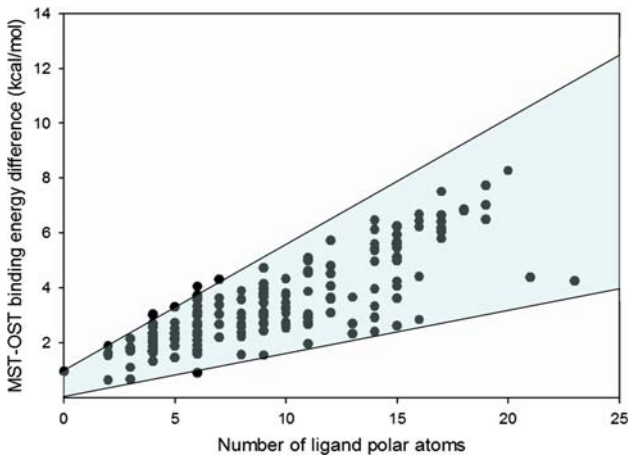

Fig. 7 Binding free energies differences between MST- and OST-calculated complexes versus the number of "polar" heteroatoms (i.e. O, N, S, H) in the ligand

MST–Autodock 4.0 comparison

The modified Autodock 3.0 version was developed independently of Autodock 4.0 but it shares a similar approach. Both Autodock 4.0 and the modified Autodock 3.0 model correctly predict that the polar atoms of benzamidine should remain solvated in the high energy complex with *lmwPTP* in contrast with the prediction of the original Autodock 3.0 model. However, both standard versions of Autodock (3.0 and 4.0) predict that benzoic acid interacting with *factor Xa* would have its polar group buried in the binding site, in spite of the non-complementary nature of the binding site meaning that the desolvation penalty would not be compensated. In the case of Autodock 4.0, however, the correct orientation is only 0.1 kcal/mol higher in energy. The MST minimum energy pose has the correct orientation.

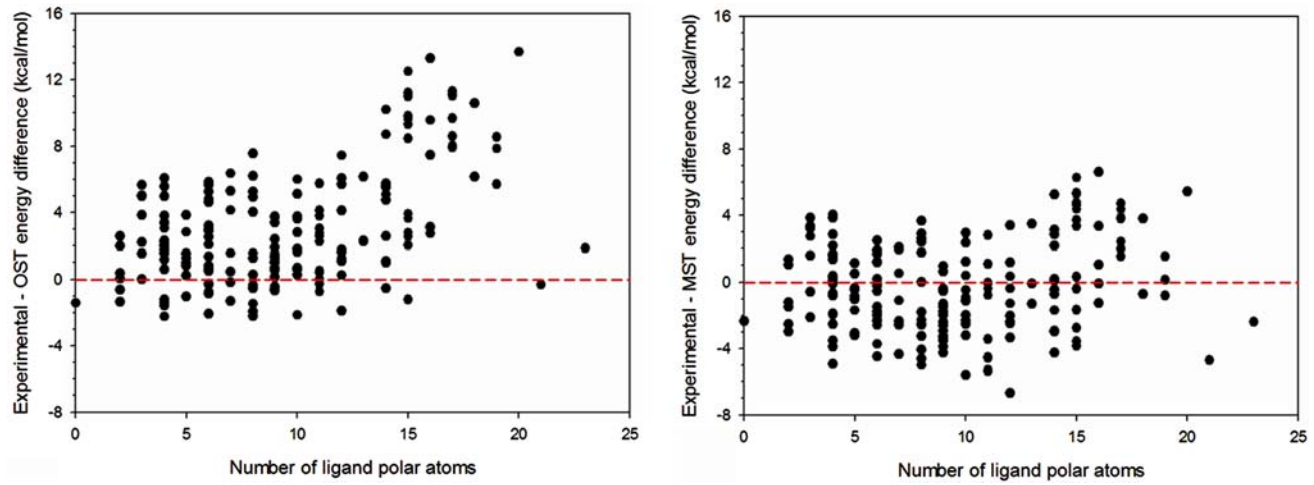

Fig. 8 Comparison of experimental and calculated binding free energies as a function of the number of polar heteroatoms in the ligand. The energies are calculated with the experimental geometry of the complexes. (**a**) OST, (**b**) MST

Autodock 4.0 binding energies were computed for 44 complexes spanning the same range of energies as in Fig. 9 selected by taking the four complexes which had deviations from the experimental values which were closer to the mean of each bin for both MST and OST models. The result is presented in Fig. 10. Autodock 4.0 clearly corrects the tendency of Autodock 3.0 and the MST to underestimate the energies of the best complexes. However, this comes at the expense of an even stronger overestimation of the affinities of weak complexes. The MST model clearly predicts more realistic affinities for weak complexes, which is advantageous for learning-based screening applications.

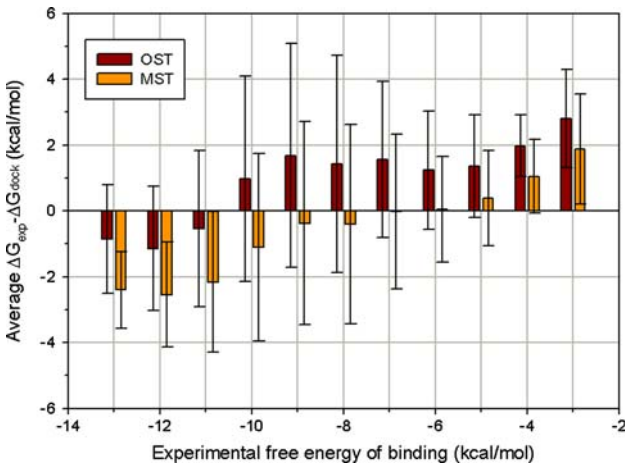

Fig. 9 Histogram showing the mean and standard deviations of the difference between experimental and calculated binding free energies (MST or OST) of 189 complexes showing experimental binding energies in ±1 kcal/mol bins centered at different values. Positive values indicate overestimated affinities with respect to experiment

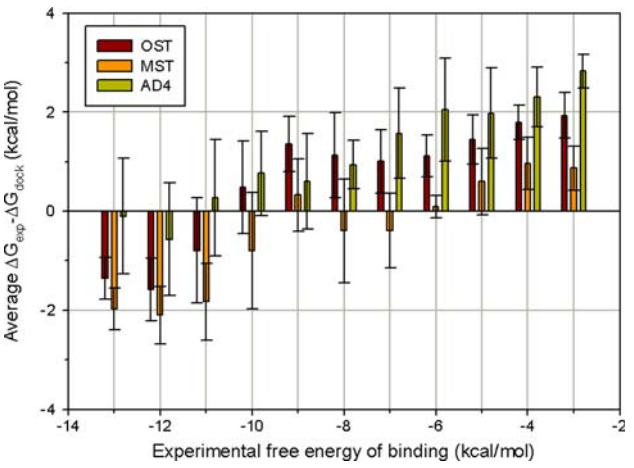

Fig. 10 Histogram showing a comparison of Autodock 4.0 and the two versions of Autodock 3.0 as a function of the experimental binding energy. Forty-four complexes out of the 189 shown in Fig. 9 were selected by taking the four complexes in each bin that are closer to the bin average for both the MST and OST models. The complexes are listed in the supplementary material

Discussion

The modified solvation treatment was introduced as a small modification of Autodock 3.0 preserving the basic features of its scoring function but correcting the original treatment of solvation that lead to the prediction of unrealistic poses for weak complexes of ligands containing polar atoms lacking a matching partner in the target (Figs. 1, 2). Our aim was to preserve the good performance of Autodock for low energy complexes and to extend it to weaker complexes that, although usually not interesting as a final result of a virtual search, may convey enough information to guide the selection of better ligands (i.e. by similarity) in search strategies designed for the exploration of very large databases.

Figure 4 shows that both scoring functions for Autodock 3.0 give similar correlations between calculated and experimental energies, using the experimental geometry of the complexes. However, the MST docking energies are consistently larger: the average differences between experimental and calculated binding energies are +1.12 kcal/mol for the OST and −0.45 kcal/mol for the MST.

The energies calculated by MST and OST for 13 complexes used in the original Autodock 3.0 calibration show a good correlation. In the ensemble of 189 complexes studied, the difference between the binding energy computed with the original and modified solvation treatment depends on the number of polar heteroatoms in the ligand and their environment (accessibility and complementary atoms in the target). While the maximum correction introduced by the modified solvation treatment is, as expected, proportional to the number of polar heteroatoms, the result of the correction is not a simple scaling but depends on the structural details of the complexes. For some complexes containing large numbers of heteroatoms the two approaches predict very similar energies. The structural dependency of the solvation correction leads to the triangular distribution of points shown in Fig. 7.

Flexible docking experiments, where the structure of the complex was optimized, show that the ability of Autodock 3.0 to predict the correct experimental structure (as measured by the average RMSD) is very similar for both scoring functions.

The main difference between the original and modified scoring functions is the treatment of polar heteroatoms and, specially, when there are no complementary sites in the target. In the absence of complementarity one would expect that polar atoms not contributing to the binding energy would be preferentially located to allow solvation by surrounding water molecules. A comparison of Figs. 5 and 6 with Figs. 1 and 2 show that the non complementary complexes between *lmwPTP* and benzamidine and

between *Factor Xa* and benzoate change their orientation using the modified scoring function, and the polar atoms not contributing with direct contacts to the complex stability become solvent exposed only when the modified scoring function is used. In contrast, the complementary complexes *lmwPTP*-phosphoric acid phenyl ester and *Factor Xa*-benzamidine preserve the correct orientation and contacts obtained with the original scoring function.

The performance of the different Autodock versions changes depending on the binding energy of the complex. We have explored a range of experimental ligand–protein complexes affinities, which varies from −2 kcal/mol to a −14 kcal/mol, corresponds to a variation in experimental inhibition constants from several hundred millimolar to subnanomolar. The original Autodock 3.0 version overestimates the binding energies of weak complexes. This trend is even more severe in Autodock 4.0, increasing the probability of false positives in virtual screening. The original Autodock 3.0 scoring function and the MST model show a tendency to underestimate the ligand affinities of strong complexes and this has been clearly improved in Autodock 4.0.

While the final goal of virtual screening is to find high affinity ligands, some successful search strategies, like the one implemented in MPA, make use of suboptimal complexes to guide the non-exhaustive search of chemical space present in very large databases. Weak complexes can be considered as proxys for the suboptimal complexes found during a virtual screening. In this approach, the performance of the docking scoring functions for weak complexes becomes essential.

The MST provides much better estimates of the docking energy for complexes with experimental energy higher than −10 kcal/mol, which include the majority of the complexes in the database. The binding energy of the stronger complexes tends to be underestimated by both models but especially by the modified model. This characteristic formally increases the risk of overlooking good ligands. However, even with the observed underestimation, the computed binding energies are low enough to deserve further investigation so the risk should be minimal.

The better accuracy of the MST model observed for most of the intermediate affinity complexes is clearly advantageous in the context of a virtual HTS strategy that use the affinities of sub-optimal complexes to guide the search of better ligands. Overestimation of the affinities of weak ligands results in the lack of discrimination between true weak complexes, that may direct the search towards "productive" regions of chemical space and those that do not contain any relevant information. In the context of an evolutionary algorithm, the efficiency of the search depends strongly on the right choice of the complexes to be retained in the evolving population. A wrong choice will result in an unproductive exploration of regions of chemical space with a low probability of containing good ligands for the target of interest.

Acknowledgments This work has been supported by funds from the Spanish Ministerio de Educación y Ciencia-FEDER BIO2004-5436 and BIO2007-63458 and the Generalitat de Catalunya. DV and GC gratefully acknowledge predoctoral grants from the Spanish Ministerio de Educación y Ciencia.

References

1. Walters WP, Sthal MT, Murcko MA (1998) Drug Discov Today 3:160. doi:10.1016/S1359-6446(97)01163-X
2. Seifert MHJ, Wolf K, Vitt D (2003) BIOSILICO 1:143. doi:10.1016/S1478-5382(03)02359-X
3. Lengauer T, Lemmen C, Rarey M, Zimmermann M (2004) Drug Discov Today 9:27. doi:10.1016/S1359-6446(04)02939-3
4. Mestres J (2002) Biochem Soc Trans 30:797. doi:10.1042/BST0300797
5. Jones G (1998) Encyclopedia of computational chemistry: genetic and evolutionary algorithms. Wiley Publisher, Chichester
6. Vidal D, Thormann M, Pons M (2006) J Chem Inf Model 46:836. doi:10.1021/ci050458q
7. Barbosa F, Horvath D (2004) Curr Top Med Chem 4:589. doi:10.2174/1568026043451186
8. Shanmugasundaram V, Maggiora GM, Lajiness MS (2005) J Med Chem 48:240. doi:10.1021/jm0493515
9. Warmuth MK, Liao J, Rätsch G, Mathieson M, Putta S, Lemmen C (2003) J Chem Inf Comput Sci 43:667. doi:10.1021/ci025620t
10. Morris GM, Goodsell DS, Halliday SR, Huey R, Hart WE, Belew RK et al (1998) J Comput Chem 19:1639. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B
11. Sousa SF, Fernandes PA, Ramos MJ (2006) Proteins 65:15. doi:10.1002/prot.21082
12. Vidal D, Thormann M, Pons M (2005) J Chem Inf Model 45:386. doi:10.1021/ci0496797
13. Thormann M, Pons M (2001) J Comput Chem 22:1971. doi:10.1002/jcc.1146
14. Hetényi C, Paragi G, Maran U, Timár Z, Karelson M, Penke B (2006) J Am Chem Soc 128:1233. doi:10.1021/ja055804z
15. Huey R, Morris GM, Olson AJ, Goodsell DS (2007) J Comput Chem 28:1145. doi:10.1002/jcc.20634
16. Stouten PFW, Frömmel C, Nakamura H, Sander C (1993) Mol Simul 10:97. doi:10.1080/08927029308022161
17. Block P, Sotriffer CA, Dramburg I, Klebe G (2006) Nucleic Acids Res 34:D522. doi:10.1093/nar/gkj039
18. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) Nucleic Acids Res 35:D198. doi:10.1093/nar/gkl999
19. AffinDB is freely accessible at http://www.agklebe.de/affinity. Accessed Jan 2005
20. Barbault F, Zhang L, Fan BT (2006) Chemom Intell Lab Syst 82:269. doi:10.1016/j.chemolab.2005.05.014
21. ChemAxon Ltd Budapest, Hungary. http://www.chemaxon.com. Accessed Jan 2005
22. Gerber PR MOLOC—a molecular design software suite. http://www.moloc.ch. Accessed Jan 2005