

# Screening of benzamidine-based thrombin inhibitors via a linear interaction energy in continuum electrostatics model

Orazio Nicolotti · Ilenia Giangreco · Teresa Fabiola Miscioscia ·  
Marino Convertino · Francesco Leonetti ·  
Leonardo Pisani · Angelo Carotti

Received: 4 December 2009 / Accepted: 28 January 2010 / Published online: 11 February 2010  
© Springer Science+Business Media B.V. 2010

**Abstract** A series of 27 benzamidine inhibitors covering a wide range of biological activity and chemical diversity was analysed to derive a Linear Interaction Energy in Continuum Electrostatics (LIECE) model for analysing the thrombin inhibitory activity. The main interactions occurring at the thrombin binding site and the preferred binding conformations of inhibitors were explicitly biased by including into the LIECE model 10 compounds extracted from X-ray solved thrombin-inhibitor complexes available from the Protein Data Bank (PDB). Supported by a robust statistics ( $r^2 = 0.698$ ;  $q^2 = 0.662$ ), the LIECE model was successful in predicting the inhibitory activity for about 76% of compounds ( $r_{\text{ext}}^2 \geq 0.600$ ) from a larger external test set encompassing 88 known thrombin inhibitors and, more importantly, in retrieving, at high sensitivity and with better performance than docking and shape-based methods, active compounds from a thrombin combinatorial library of 10240 mimetic chemical products. The herein proposed LIECE model has the potential for successfully driving the design of novel thrombin inhibitors with benzamidine and/or benzamidine-like chemical structure.

**Keywords** LIECE · Virtual screening ·  
Thrombin inhibitors · Benzamidine

## Introduction

The successful identification of new hits and their potential evolution to leads is often the result of compromises involving blind serendipity and rational approaches. While the former is a self-consistent event, the latter have been continuously pursued through the development of new advanced computational methodologies inspired to the basic principles of structure-based molecular design. In recent years, the availability of an always increasing number of solved protein three-dimensional structures enabled the birth of newer technologies for better understanding whether and how small molecules can interact with proteins [1–9]. In this scenario, molecular docking is playing a prominent role and is nowadays the up-front strategy for drug discovery research. As widely recognized, molecular docking was over the years challenged for three main goals: (1) the accurate reproduction of binding mode of known active ligands; (2) the identification of hits from a decoy pool; (3) the prediction of binding affinities of structurally close compounds. Of these three purposes, the last was indeed the most difficult as no statistically significant correlation between measured affinity and scoring function has so far emerged [10, 11]. Recently, some of us approached this issue by devising a two term fitness function: the first accounted simultaneously for best regression among experimental affinity and docking score and the second for minimal atom displacements from a given crystallographic binding hypothesis [12–14]. However, despite the enormous progress in improving and calibrating scoring functions, the accurate prediction of binding affinity still remains an unmet goal. As brilliantly reported by Leach and colleagues the pitfall is that scoring “is a problem of subtraction of large numbers, inaccurately calculated, to arrive at a small number” [15]. The Binding

**Electronic supplementary material** The online version of this article (doi:10.1007/s10822-010-9320-1) contains supplementary material, which is available to authorized users.

O. Nicolotti (✉) · I. Giangreco · T. F. Miscioscia ·  
M. Convertino · F. Leonetti · L. Pisani · A. Carotti  
Dipartimento Farmaco-Chimico, University of Bari,  
Via Orabona 4, 70125 Bari, Italy  
e-mail: nicolotti@farmchim.uniba.it

Free Energy (BFE) is the small number resulting by the subtraction of two large numbers: (a) on one side, the interaction energy enabling contacts between ligand and protein; (b) on the other side, the energetically expensive desolvation to squeeze, both ligand and protein, out of the water to form the ligand–protein complex.

Huge efforts have been dedicated to cope with this problem. Normally, the free energy difference is computed by using solid standard formula such as Free Energy Perturbation (FEP), Thermodynamic Integration (TI) and Umbrella Sampling (US) [16, 17]. Based on Molecular Dynamics (MD) or Monte Carlo (MC) simulations, these techniques are very time demanding and, thus, of often limited feasibility [18–21]. Alternatively, Åqvist and colleagues proposed a faster methodology known as Linear Interaction Energy (LIE) that improved convergence and decrease computational cost by empirical approximation of FEP and TI formulas [22]. Later, Jorgensen and colleagues proposed a modified LIE approach by implementing a cavitation term, proportional to the solvent accessible surface area of the solute, to estimate hydration free energies [23, 24]. Other LIE related approaches based on Generalized Born (GB) approximation [25–28] of electrostatic solvation or on a continuum electrostatic solvent Poisson-Boltzmann [29–31] model have been recently reported in literature. Lately, Caffisch and colleagues enforced LIE methods with a rigorous treatment of continuum electrostatics based on the numerical solution of the Poisson equation by the finite difference technique. Popular as LIECE [29] (standing for Linear Interaction Energy in Continuum Electrostatics), the new approach improved the efficiency of the LIE methods by replacing explicit water MD (or MC) simulations with energy minimization in a continuum model of the solvent. Another smart development of the LIECE method led to QMLIECE, an approach implementing the linear-scaling semiempirical quantum mechanical calculation of the intermolecular electrostatic energy. LIECE and QMLIECE were successfully applied to several interesting cases of study [32–34].

The present investigation is aimed at modeling the thrombin inhibitory activity through the derivation of a LIECE model obtained from a series of 27 benzamidine-based compounds [35–43]. The LIECE model was profitably challenged for (a) predicting binding affinity of an even larger test set of compounds [44, 45] and, more importantly, for (b) screening a sizeable combinatorial library [46], both biased toward thrombin. The LIECE model was also challenged against docking and shaped-based methods, in retrieving active hits from a large pool of chemically similar decoys. In this respect, the current work is aimed to offer to the medicinal chemists involved in research on thrombin and thrombin-related diseases (e.g., coagulation cascade, blood clot formation) new computational routes for a more solid

and reliable estimation of anti-thrombin activity and, ultimately, for the discovery of new therapeutic agents potentially useful in several blood-related disfunctions.

## Materials and methods

### Target structure

The Cartesian coordinates for thrombin were obtained from the 1ETS crystal structure available from the Protein Data Bank (PDB) [7]. As well known, the catalytic triad of the serin-proteases is constituted by the residues Asp102, His57 and Ser195 which are located in the subsite S2. Three main additional subsites in the thrombin binding pocket have been also highlighted. The first, termed S1, is formed by a deep, narrow hydrophobic cavity at the bottom of which Asp189 forms a salt bridge with positively charged moieties of substrates and inhibitors. The second subsite, named D (distal), is basically lined by aromatic residues. The third subsite, known as P (proximal), is made up by the loop Tyr60A- Pro60B- Pro60C- Trp60D.

Protein was prepared by adding hydrogen atoms, inserting and optimizing missing residues, and removing water and the co-crystallized inhibitor  $N^{\alpha}$ -(2-naphthyl-sulphonyl-glycyl)-DL-*p*-amidinophenyl-alanyl-piperidine (4-NAPAP). On the basis of documented literature [44], controls were carefully carried out on the protonation state at physiological pH, assigning a formal positive charge to lysine and arginine residues and a formal negative charge to aspartate and glutamate residues. Disulphide bridges were built between cysteine thiols closer than 2.0 Å. Using the Protein Preparation module in Maestro 7.5 [47] implementing the OPLS2001 all-atom force field [48], light relaxation was performed to optimize hydroxyl and thiol torsions followed by all-atom constrained minimization to relieve steric clashes until the rmsd reached 0.18 Å.

### Training set

The LIECE model was derived on the basis of a training set comprising a number of 27 benzamidine derivatives (Tables 1, 2) with known thrombin inhibition data in the  $pK_i$  range 3.64–8.40. The model reliability was insured by the inclusion of a number of 12 X-ray solved ligands (1–12, Table 1) extracted from thrombin crystal complexes available from the PDB [35–42]. Such inhibitors shared a benzamidine chemical core variably substituted at the position 4 with known  $pK_i$  values in the range 4.35–8.40. An additional number of 17 ligands, manually docked biasing the X-ray binding conformations, was taken into consideration to enrich the structural diversity domain as

**Table 1** LIECE model training set: thrombin inhibitors extracted from X-ray complexes available from the PDB

No	PDB code	Structure	$q$	$pK_i$	No	PDB code	Structure	$q$	$pK_i$
1	1AHT <sup>a</sup>		0	6.21	7	1YPE		+2	8.10
2	1GHV		+1	4.35	8	1YPG		+2	8.00
3	1GHW <sup>a</sup>		+1	5.40	9	2CF8		+2	8.10
4	1K22		0	8.40	10	2CF9		+2	7.82
5	1KTS		+1	8.35	11	1YPJ		+2	7.02
6	1VZQ		+2	7.44	12	1ETS		+1	8.22

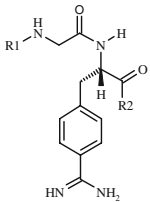
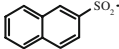

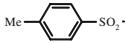

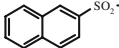

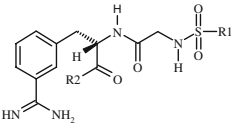


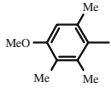


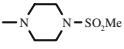
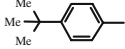

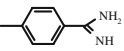



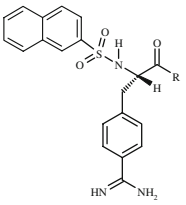


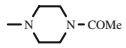
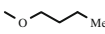
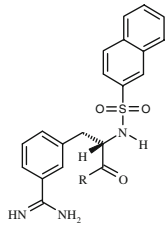
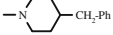

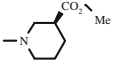

<sup>a</sup> Molecules discarded from the LIECE model (see text)

well as to better harmonize the  $pK_i$  interval [43]. This pool of ligands was arranged in four groups formally deriving from two basic chemical structures related to NAPAP and TAPAP that were both 3- and 4-amidinophenylalanine derivatives whose crystal structures were available.

The first group (**13–15**, Table 2a) of compounds included 4-NAPAP derivatives. The second group (**16–21**, Table 2b) encompassed analogues of 3-NAPAP (standing

for  $N^{\alpha}$ -(2-naphthyl-sulphonyl-glycyl)-*DL*-*m*-amidinophenyl-alanyl-piperidine). The third group (**22–25**, Table 2c) comprised the congeners of 4-TAPAP (standing for  $N^{\alpha}$ -[4-toluenesulphonyl]-*D,L*-*p*-amidino-phenylalanyl-piperidine). Molecules affiliated to 3-TAPAP (standing for  $N^{\alpha}$ -[4-toluenesulphonyl]-*D,L*-*m*-amidino-phenylalanyl-piperidine) belonged to the last group (**26–29**, Table 2d). According to previous works [44, 45], the protonation state was

**Table 2** LIECE model training set: a) 4-NAPAP; b) 3-NAPAP; c) 4-TAPAP and d) 3-TAPAP derivatives, respectively. R, R1 and R2 substituents, formal charges (i.e.,  $q$ ) and  $pK_i$  values are reported

a	No	R1	R2	q	p <i>K</i> <sub>i</sub>	No	R1	R2	q	p <i>K</i> <sub>i</sub>
	13			+2	6.26	15			+1	6.13
	14			0	6.29					
b	No	R1	R2	q	p <i>K</i> <sub>i</sub>	No	R1	R2	q	p <i>K</i> <sub>i</sub>
	16			+1	5.44	19			+1	6.38
	17			+1	5.05	20			+1	5.85
	18			+1	4.43	21			+1	5.92
c	No	R	q	p <i>K</i> <sub>i</sub>	No	R	q	p <i>K</i> <sub>i</sub>		
	22		+1	6.38	24		0	3.64		
	23		+1	4.89	25		+1	6.43		
d	No	R	q	p <i>K</i> <sub>i</sub>	No	R	q	p <i>K</i> <sub>i</sub>		
	26		+1	5.11	28		+1	6.92		
	27		+1	6.44	29		+1	4.55		

assumed to be the following: the benzamidine and basic functional groups were protonated while aromatic amino groups were left uncharged, carboxylic moieties were considered to be deprotonated. All of these structures were used for the derivation of the LIECE model and, thus, constituted its source of physicochemical attributes.

#### External test set

A molecular collection of 88 3-amidinophenylalanine derivatives was used to sample the LIECE model applicability and predictive ability [44, 45]. As well known, this data set represented a standard for QSAR practitioners; it was intensively used for training QSAR models of serine

protease inhibitors as well as for the development of new computational methods. All the chemical structures encompassed into this molecular series were taken from the Tripos database [49] and subjected to preliminary minimization using the standard *Prem* script released in MacroModel [50]. Structural variations were made at the substituents R1 and R2 of the benzamidine core leading to  $pK_i$  values spanning over 4 logarithmic units (see Table 4 of the Electronic Supplementary Material). Each structure was subjected to runs of molecular docking simulations and a number of 20 docking poses *per* compound was obtained.

### Virtual screening benchmark

Retrieved from a combinatorial library biased toward thrombin, a large external benchmarking set of 10240 compounds was taken into account [46]. To minimize random retrieving and to perform an even stringent test, all of the screened structures shared a similar benzamidine-based core and were, thus, both physically and topologically close to compounds comprised into the LIECE model. Effectively, the chosen 10240 set of compounds were obtained by reacting a number of 80 aldehydes, 8 amines and 16 isonitriles according to the Ugi-type three component reaction (Fig. 1).

As reported in the original paper [46], the library components spanned a broad range of chemical diversity mainly derived from the large structural variations of the isonitriles and aldehydes reactants including small aliphatic, aromatic, heteroaromatic, benzylic groups. Incorporating structural pattern suitable for the thrombin binding S1 pocket, the amines were instead known to act as arginine mimetic with affinity toward thrombin in the high micromolar range.

As normally done in standard docking screens, experiments of ligand enrichment among top-ranking hits and Receiver Operator Characteristic (ROC) [51] curves were

assumed as a key metrics for assessing LIECE model performances.

### LIECE model fundamentals

The estimation of BFE was obtained on the basis of the following general two-parameter LIECE model [29]:

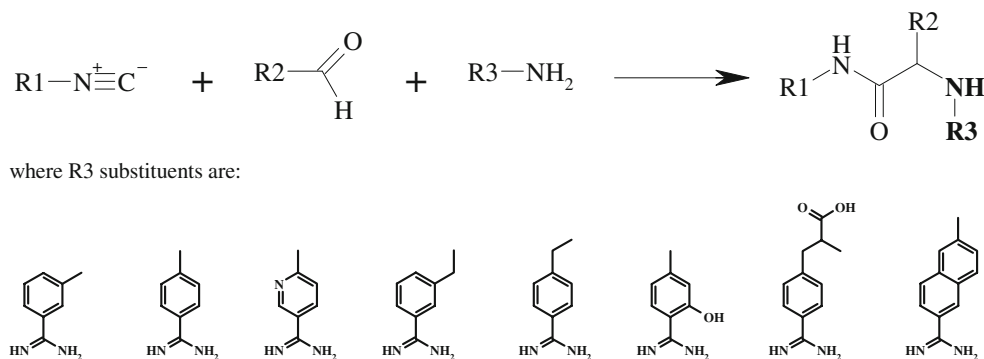
$$\Delta G_{bind} = \alpha \Delta E_{vdW} + \beta \Delta G_{elec}$$

where  $\Delta G_{bind}$  quantifies the amount of BFE,  $\Delta E_{vdW}$  represents the ligand–protein van der Waals interaction energy,  $\Delta G_{elec}$  accounts for the sum of ligand–protein Coulomb energy computed in vacuo and the change in solvation energy of protein and ligand upon binding, while  $\alpha$  and  $\beta$  are numerical coefficients derived through linear regression.

The energetic terms occurring in the BFE equation were derived by subtraction of the corresponding values of isolated ligands and protein from those of their complexes. In particular, after ligand minimization into the fixed thrombin binding site, the van der Waals energy was calculated with CHARMM [52] using the CHARMM22 force field (Accelrys Inc.) with the nonbonding cutoff set to 14 Å while the Coulomb energy was computed by using infinite cutoff without considering interactions between atoms separated by one or two covalent bonds. The solvation energy was instead calculated by the finite-difference Poisson method [53] (using the PBEQ module [54] in CHARMM [52]) for ligands, protein and ligand–protein complexes.

Based on a focusing procedure described elsewhere [29], a grid covering a layer of 20 Å around the solute with a final grid space of 0.3 Å was applied. Using the PBEQ module [54], each calculation was performed twice by setting the exterior dielectric equal to 78.5 and to 1.0 for first and second computation, respectively. The dielectric discontinuity surface was delimited by the molecular surface identified by means of a rolling probe of 1.4 Å; the ionic strength was dropped to zero; the temperature was set to 300 K; and the solute dielectric equalized to 1.0. The

**Fig. 1** Ugi-type three-component reaction involving 16 isonitriles, 80 aldehydes and 8 amines to generate a combinatorial library of 10240 compounds



solvation energy is assumed as difference between the above two described calculations.

The LIECE model approximated the estimation of BFE. In particular, the entropic term was not explicitly considered and, thus, tacitly the effect of the induced fit in the ligand–protein interaction was not taken into account. Steepest descent and conjugated gradient algorithms were adopted for ligand minimization to a rmsd of the gradient of  $0.01 \text{ kcal}\cdot\text{mol}^{-1} \text{ \AA}^{-1}$ . Partial charges were assigned with *Wit!P* molecular modeling package [55] using the Modified Partial Equalization of Orbital Electronegativity (MPEOE) method [56, 57].

#### Molecular docking simulations

GOLD [58] was used for docking simulations of the 88 test set compounds and of the 10240 structures available from a chemical combinatorial library. Among the variety of existing docking software, GOLD was used as it was extensively trained on a large number of complexes selected from the CCDC/ASTEX [59] validation set where serine proteases were very well represented. As reported elsewhere, GOLD settings were tuned and calibrated by docking the co-crystallized ligand (i.e., 4-NAPAP) within its own crystal structure (i.e., 1ETS [42]). Expected binding modes were accurately reproduced and the lowest rmsd value, over heavy atoms between predicted and experimentally solved positions, was equal to  $0.656 \text{ \AA}$ . Using the refined crystal structures of thrombin, GOLD was set to generate 20 and 5 docking poses for each of the test set and the combinatorial library compounds, respectively. Centered on the alpha carbon of Gly216, residues within an active radius of  $11 \text{ \AA}$  were considered in docking simulations. According to descriptions elsewhere [45], ligand–protein complexes were generated by setting physical constraints to favour the occurrence of well-established interactions such as the salt bridge formed by Asp189 and the HBs engaged by Gly216 and Gly219 with the inhibitors.

#### Shape-based similarity analysis

The chemical similarity between each obtained docking pose and the co-crystallized 4-NAPAP inhibitor was calculated on the basis of their molecular shapes. The calculation was operated through the program ROCS [60] (which stands for Rapid Overlay of Chemical Structures, from OpenEye Scientific Software), enabling shape-based superimposition through solid-body optimization process that maximizes the overlap volume between molecules. Each docking pose was therefore attributed its own shape Tanimoto similarity [61] value, ranging from 0 to 1 to indicate lower or higher shape similarity.

## Results and discussion

#### LIECE model derivation

The X-ray solved ligands extracted from complexes deposited into the PDB represented a solid starting point for the construction of a reliable LIECE model.

Preliminarily, X-ray solved complexes were superimposed by using the *Structural Autopairs Fitting Method* [62] implemented in *Wit!P* [55]. The Cartesian coordinates of heavy atoms forming both backbone and side-chains of 1ETS [42] were assumed as a driving ruler. The enzyme residues coordinates were deleted and, thus, only the co-crystallized ligand coordinates were kept. The presumed active conformations of inhibitors **13–29** (Table 2a–d) were instead manually modeled by biasing the prevalent binding modes of similar compounds occurring in X-ray complexes. This task required efforts and expertise acquired in modeling serine protease inhibitors over recent years [12–14]. To avoid inadvertent actions when deriving the molecular alignment, a careful delineation of the residues encompassing the binding site as well as of the molecular room effectively available was rigorously taken into account. To this end, even subtle changes of the molecular conformation of each inhibitor were immediately subjected to soft molecular minimization. In this regard, it has to be stressed that the goodness of the LIECE model is depending upon two main reasons. The first is concerned with its physical validity and, technically, with the algebraic sign of the coefficients (i.e.,  $\alpha$  and  $\beta$ ) resulting after linear regression analysis. It is a known fact that both the coefficients were required to be positive to avoid BFE models with non-sense physical meaning. The second motive is instead related with statistics that is a measure of model ability in representing data and explaining their variation. In this respect, Leave One Out (LOO) analysis [63] was, thus, designated as a criterion for discarding or including compounds. In the present investigation, both these two aspects constituted the guidelines for the obtaining of a reliable LIECE model.

The herein proposed LIECE model comprised all the molecules listed in Tables 1, 2a–d with the exception of inhibitors **1** and **3** as their inclusion dropped down statistics ( $r^2 = 0.519$ ,  $q^2 = 0.365$ ). Both these inhibitors were supposed to act as outliers for the lacking of R1 and R2 substituents normally ensuring stable antiparallel hydrogen bond interactions with Gly216.

Noteworthy, the herein proposed 27 objects LIECE model was however based on a number of data points by great far larger than those accounted for other published models of thrombin inhibitors [64–66]. More specifically, the LIECE model was as follows:



$$\Delta G_{\text{bind}} = 0.224 (\pm 0.007) \cdot \Delta E_{\text{vdW}} + 0.085 (\pm 0.013) \cdot \Delta G_{\text{elec}}$$

$$n = 27, r^2 = 0.698, q^2 = 0.662, \text{PRESS} = 28.938$$

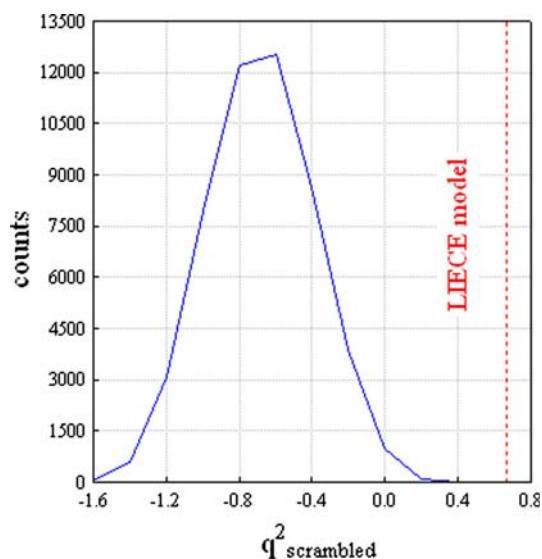
$$\text{RMS}_{\text{error}} = 1.025 \text{ kcal mol}^{-1}, \text{MUE} = 0.800 \text{ kcal mol}^{-1}$$

where  $n$ ,  $r^2$ ,  $q^2$ , PRESS,  $\text{RMS}_{\text{error}}$  and MUE represent the number of objects in the model, the squared correlation coefficient, the leave-one-out cross-validated correlation coefficient, the predictive sum of squares, the root mean squares of errors and the mean unsigned error, respectively. Model statistics was validated by intensive randomization tests [67]. A number of 50000 y-scrambling runs was conducted with a consistent drop of statistical parameters. The analysis of y-scrambling disclosed that only few  $q^2$  values resulted positive values the average being equal to  $-0.580$  (Fig. 2).

The molecular superimposition of the analyzed inhibitors embedded into the X-ray crystal structure of thrombin as well as the data spread of the LIECE model is shown in Fig. 3.

#### LIECE model validation via a larger external test set

Unlike other 3D-QSAR grid-based methods, [68, 69] LIECE takes explicitly into consideration the energetic values (i.e.,  $\Delta E_{\text{vdW}}$  and  $\Delta G_{\text{elec}}$ ) resulting from the interaction between each single aligned ligand and its biological counterpart. However, the reliability of the obtained models can be ensured only from severe validation tests. Typically tailored to this purpose, predictions in 3D-QSAR are performed on a test set consisting of a limited number of compounds extracted from the entire data set according to widely accepted criteria based on structural diversity and



**Fig. 2** Validation of LIECE model by randomization tests: distribution of  $q^2$  values after 50000 y-scrambling runs. The dashed red line indicates  $q^2$  value of the LIECE model

regular distribution of activity values. In the present work, the test set used for validation was an external molecular collection about three times larger than the training set. In addition, docking was used to sample the conformational space to unequivocally avoid manual modelling.

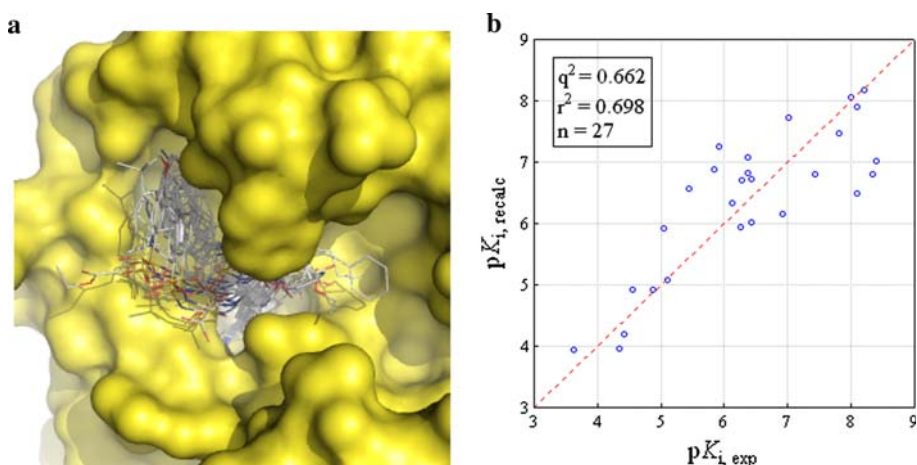
In this regard, each docking pose generated through GOLD, for a general figure of  $20 \times 88$  structures, was subject to a soft molecular minimization within the fixed thrombin binding site by using the CHARMM22 force field in order to calculate  $\Delta E_{\text{vdW}}$  and  $\Delta G_{\text{elec}}$  and, then, to rescore the corresponding BFE values by extrapolation on the basis of  $\alpha$  and  $\beta$  coefficients of the LIECE model.

The projections of experimental versus predicted BFE values (Fig. 4a) disclosed for each inhibitor a consistent spread. Nor the selection of the top-scored docking results neither the choice of the lowest BFE solutions or the pose with the highest shape similarity provided any convincing correlations and very disappointing values of  $r_{\text{ext}}^2$  equal to 0.104, to 0.045 and to 0.050, respectively, were observed. This was a typical trap of the *Kubinyi's paradox* stating the impossibility of proofing the relationship among internal and external predictability [70]. However, the retrospective analysis of poses with smallest BFE residuals disclosed a potential promising correlation ( $r_{\text{ext}}^2 = 0.606$ ). In this respect, it resulted clear that LIECE model failed to calculate BFE of a number of molecules (i.e., **30**, **39**, **58**, **69**, **89** and **102** see Table 4 in Electronic Supplementary Material) having large energetic residuals (i.e.,  $>2.133 \text{ kcal mol}^{-1}$ ) even retrospectively considering the better performing docking pose. As expected, the linear fit was significantly raised up ( $r_{\text{ext}}^2 = 0.914$ ) after their removal. Despite not always straight and easy, a few hypotheses were herein proposed to explain why the LIECE model did not success in calculating the  $\Delta G_{\text{bind}}$  values of these test set compounds. Allegedly, it is worthy saying that the excluded molecules were generally characterised by R1 and/or R2 substituents not sufficiently sampled, or not even covered, in the LIECE model. A few examples are the dansyl R1 substituent in **30** and the 3-carboxyl-1,2,3,4-tetrahydroquinol-1-yl, the 3-carboxy-1,2-dihydroquinol-1-yl, 2-carboxypiperidin-1-yl, 2-carboxypyrrolidin-1-yl R2 substituents in **39**, **58**, **69** and **89**, respectively. On the other hand, the incorporation of a different molecular spacer connecting the benzamidine core with the R1 substituent was the likely reason for the adopting of unusual conformations by **102** whose  $\Delta G_{\text{bind}}$  value was, thus, erroneously estimated.

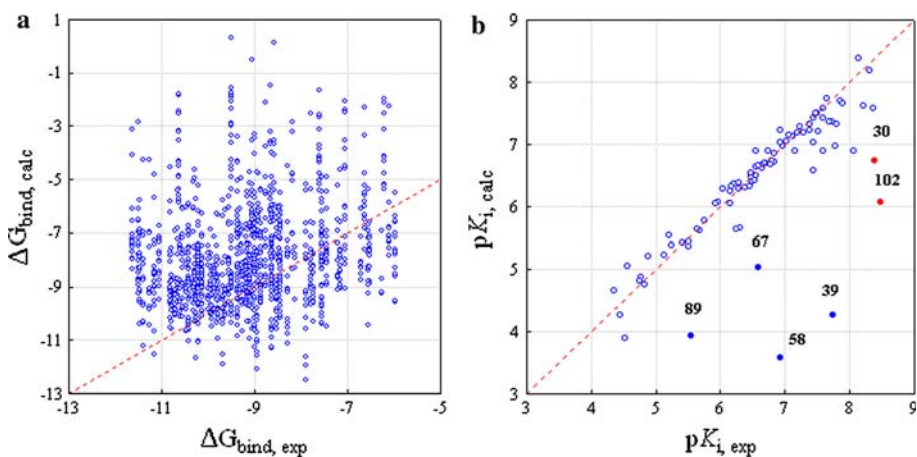
Of course, the purpose of this retrospective analysis was only diagnostic and, thus, aimed at ensuring the existence of really predictive poses among those generated during docking simulations. In this respect, consensus-based methods [71] were investigated with the intention of designating predictive poses among those available after docking campaigns. Accordingly, a number of molecular

**Fig. 3** The LIECE model:

**a** zoomed-in view of molecular superimposition and **b** training set spread of experimental and recalculated  $pK_i$  values. The red dashed line represents the ideal correlation



**Fig. 4** Correlation between experimental and calculated binding free energies: **a** test set spread encompassing projections for all the docking poses; **b** test set spread focused on the pose having the lowest  $pK_i$  residual *per* compound (test set compound numbers are related to structures in Table 4 in the Electronic Supplementary Material). Red and blue solid circles indicate badly calculated compounds (see text). The red dashed lines represent the ideal correlation



features was taken into account to prioritize, on the basis of different ranking rules, the hypothetical predictive poses. The most successful trial involved the assignment of three diverse attributes *per* pose. The first was the Tanimoto shape-based similarity, the second and the third were the van der Waals ( $\Delta E_{vdW}$ ) and the electrostatic ( $\Delta G_{elec}$ ) terms, respectively. Each pose was, thus, represented through a numerical triad encoding its position after incremental sorting of each individual attribute. A global rank was unequivocally calculated through a consensus maximum and the supposed predictive pose accordingly designated.

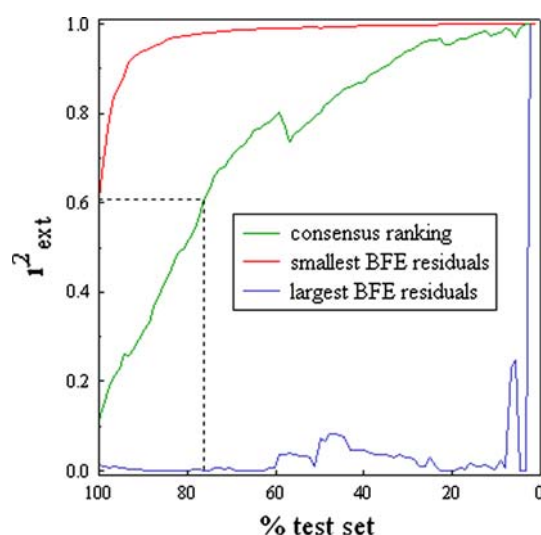
As shown in Fig. 5, a comfortable predictive trend with  $r_{ext}^2 \geq 0.600$  was observed for about the 76% of the test set. Despite derived from a smaller number of observations ( $n = 27$ ), the LIECE model was however successful in properly predicting BFE values for at least 67 out of 88 test set compounds. In the pool of badly foreseen inhibitors were comprised, as expected, the molecules (i.e., **30**, **39**, **58**, **69**, **89** and **102**) already discussed in our retrospective analysis because of their large energetic residuals. Disappointingly, the consensus strategy failed to designate predictive poses for an additional number of 15 compounds (i.e., **35**, **55**, **71**, **78**, **93**, **86**, **88**, **90**, **91**, **97**, **98**, **101**, **112**, **116**

and **117**). Presumably, it was likely due to the different chemical types in R1 and/or R2, as well as to the molecular spacer of the above compounds, that were not sufficiently covered by the LIECE model.

#### Virtual screening experiments

The LIECE model was ultimately challenged for its performance to fish active hits (i.e., actives) from a large mimetic combinatorial library of 10240 products whose affinity toward thrombin was experimentally determined [46]. Considering a threshold equal to 100 nM (i.e.,  $pK_i = 7.00$ ), the number of actives was equal to 34 with the inactives, that are decoys, thus being over the 99%. The overwhelming amount of decoys and, in addition, their closely related chemical structures ensured that the random selection did not compete with the effective real recognition and, thus, made the screening experiments even harder. In this view, the LIECE model was severely tested with the intention to compare its performance to other screening procedures such as those related to docking and shape-based methods. In this respect, a knowledge-based criterion was adopted to designate the docking pose, among





**Fig. 5** Profiles of the predictive squared correlation coefficient  $r^2_{\text{ext}}$  obtained by the LIECE model at different percentages of the test set. The red and blue profiles indicate baseline  $r^2_{\text{ext}}$  as they were derived by intercepting docking poses with the smallest and largest residuals, respectively. The green profile represent the  $r^2_{\text{ext}}$  changes after selecting the docking poses through a consensus scheme

those available *per* compound, for the effective ranking of the entire combinatorial library. More specifically, the ranking was based: (a) on the pose assigned with the minimum  $\Delta G_{\text{bind}}$  extrapolated value (i.e., the maximum  $pK_i$ ) for LIECE; (b) on the top scored pose for docking; (c) on the pose with the highest Tanimoto similarity value for the shape-based analysis. Screening performances were, thus, assessed by comparing several metrics. Initially, the fishing of actives was quantified by measuring the enrichment factor (EF), calculated as follows:

$$EF(X\%) = \frac{100}{X} \cdot \frac{\text{actives}_{\text{fished}}}{\text{actives}_{\text{total}}}$$

where  $\text{actives}_{\text{fished}}/\text{actives}_{\text{total}}$  is the fraction of actives found at a given X percentage of the screened database [72]. The most pragmatic use of the EF metric is in the analysis of the early recognition enrichment by monitoring the top few percent of the ranked library. Interestingly, in the very early screening, at the top 1% of the ranked combinatorial library, LIECE model outperformed with an  $EF_{1\%}$  equal to 5.88 while docking screening achieved a value of 2.94. A similar trend hold true also when considering the top 10% threshold. In this regard, the LIECE model resulted an  $EF_{10\%}$  equal to 2.35 while docking resulted a value of 0.88. Shape-based similarity analyses were instead very disappointing as no actives were retrieved within the 1% as well as the 10% of the ranked combinatorial library (Fig. 6).

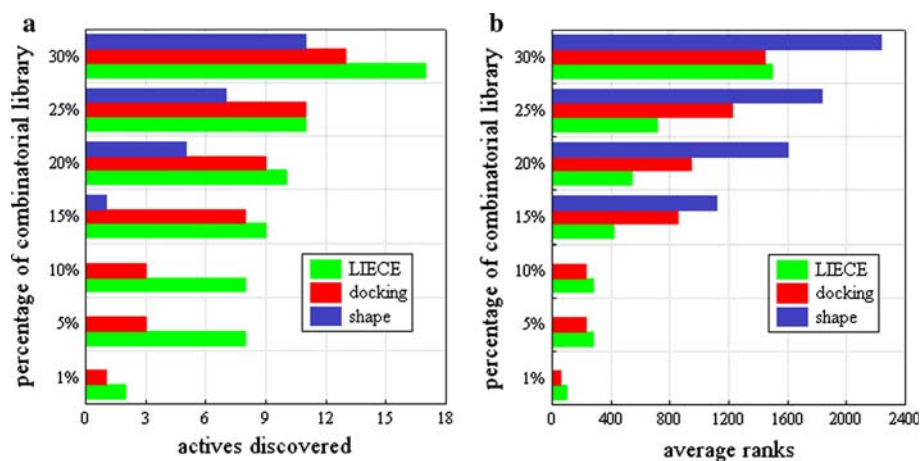
A closer look at the enrichments revealed that LIECE achieved more promising results in almost all experimental

screening conditions. And in fact, this tendency was further reinforced by inspecting the number of actives as well as the values of average ranking observed irrespective of the percentage of the combinatorial library screened. In Fig. 7 the chemical structures of the actives retrieved at percentage equal to the 20% are shown. As can be seen, the phenoxyphenyl substructure, incorporated by reacting the 4-phenoxy-benzaldehyde, was the most frequently represented in the pool of the actives selected by the LIECE model. Interestingly, this group emerged as a privileged substructure and was one of the of most represented molecular frameworks [73, 74] in the Comprehensive Medicinal Chemistry (CMC) [75] database, a collection of two-dimensional and predicted three-dimensional structures as well as important biochemical properties for known drugs. As shown in Fig. 7, the phenoxyphenyl fragment recurred 7 out of 10 for the actives retrieved from LIECE screening and, implicitly, this observation ensured that the chance of random fishing was actually minimal. In fact, the entire combinatorial library encompassed a number of 34 actives (i.e.,  $K_i < 100$  nM) with 11 of them including the phenoxyphenyl substructure. In addition, a closer inspection revealed that this group was joined to the *para* or *meta* position in case of the 4- and 3-substitued benzamidine core, respectively.

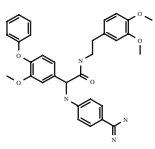
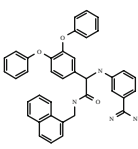
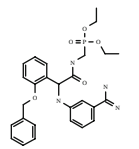
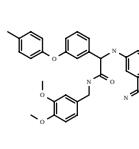
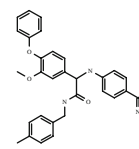
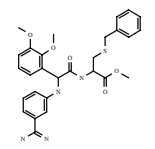
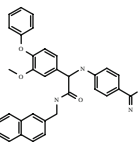
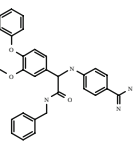
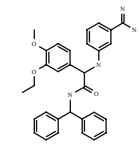
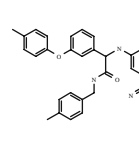
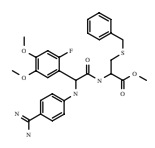
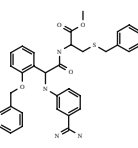
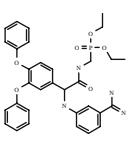
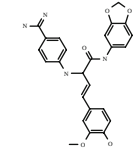
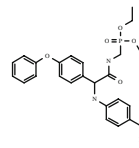
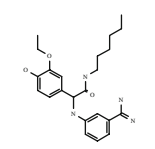
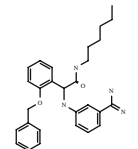
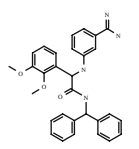
However, the phenoxyphenyl group was also intercepted by the docking screening (i.e., 6 out of 9) as well as the shape-based retrieving (i.e., 1 out of 5). The visual analysis of this fragment within the thrombin binding pocket revealed that it was likely to be engaged in hydrophobic interactions within the subsite D. Incidentally, the benzyl-oxyphenyl group, which to some extent can be considered a close chemical variant of the phenoxyphenyl, occurred as well. These results prompted us to investigate the occurrence of phenoxyphenyl fragment in the WOMBAT [76–78] (standing for World of Molecular BioAcTivity) collection which covered a large number of annotated entries from medicinal chemistry journals between 1975 and 2008. As shown in the pie plot in Fig. 8, the phenoxyphenyl fragment enabled the recovery of 2996 compounds having measured biological activity toward different target types (i.e., from receptors, proteins, to nucleic acids, to ion channels and to enzymes according to the definition for drug targets stated by Imming and coll. [79]). Interestingly, the inspection of WOMBAT outcomes resulted a number of thrombin inhibitors and among them a subset of 17 benzamidine structures incorporating the phenoxyphenyl group.

For the ease of completeness, the ROC curves were thus profiled to relate the discovered inactive fraction (i.e., the term [1–specificity]) versus the discovered active fraction (i.e., the term sensitivity) [51]. As can be seen in Fig. 9, the LIECE model resulted the dominant ROC curve with a profile even more dominant in the early stages. The

**Fig. 6** Histograms indicating **a** the number of actives discovered and **b** the average ranking at different early percentages of the screened combinatorial library



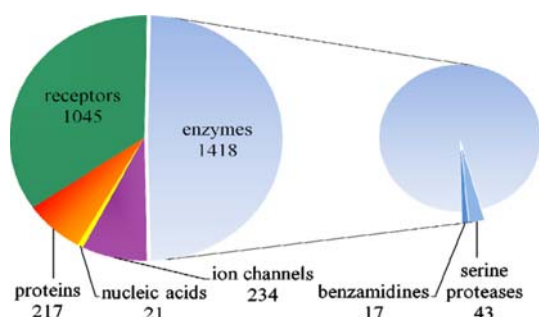
**Fig. 7** Chemical structures of the actives fished at the top 20% of the ranked combinatorial library

 L93, G1396	 L97	 L216, C1607	 L231	 L329, G1254
 L343	 G355, L441	 L509, G1057, C1694	 L1497	 L1680
 G61	 G281	 G1047	 G1385	 G1683
 C1117	 C1544	 C2043	The letters L, G and C stand for LIECE, docking and shape-based methods, respectively. The numbers indicate the actual rank.	

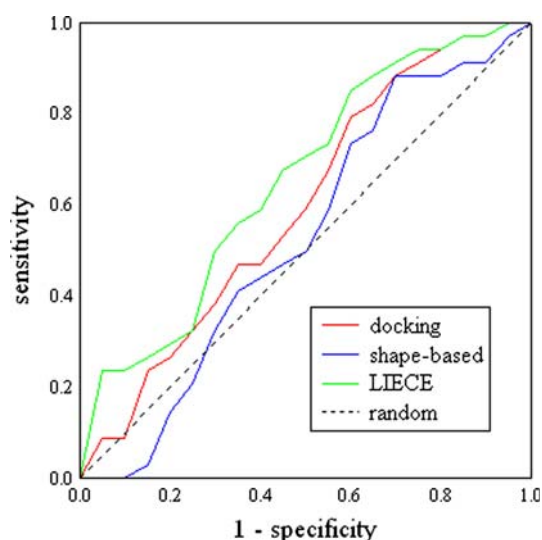
comparison of the area under curve (AUC) disclosed a value equal to 0.655 for the LIECE model with the docking and shape-based methods achieving worse performance with AUC values of 0.596 and 0.528, respectively.

Finally a series of test was executed to estimate the variation of screening performances at the change of the

$pK_i$  threshold designated to discriminate actives from decoys. The border to designate actives was varied into the  $pK_i$  range 6.00–7.50 with a step equal to 0.5. In that interval, the number of actives decremented from 348, to 102, to 34 and to 10, respectively. Satisfactorily, the LIECE model demonstrated always better performances



**Fig. 8** Pie plots showing the occurrence of the phenoxyphenyl fragment over different drug target types after inspecting the WOMBAT molecular collection



**Fig. 9** Red, green and blue ROC curves illustrating the performances of the LIECE model, the docking and shape-based methods, respectively. The black dashed line represents the random retrieving

irrespective of the threshold set for designating the number of the actives. Further details can be found in the Electronic Supplementary Material.

Finally, it should be remarked that the ultimate aim of the current analysis is not the election of the best virtual screening method above all the others. As well known, the failure or the success of screening experiments depends upon many factors. In our case, the use of a mimetic combinatorial library enhanced the likelihood of success for the LIECE model. And in fact, the LIECE model is explicitly local and, thus, is intrinsically structured to capture differences, even rather small, as those occurring among mimetics. On this basis, it goes without saying that docking and shape-based methods are expected to give more convincing results when the database under investigation comprises a more remarked molecular diversity spanning numerous chemotypes as those typically present in vendors and corporate collections. In our opinion, the choice of the screening method is strictly related to the scope of a given

investigation. In this view, the LIECE model is indeed a good tool when the basic interest is focused on lead optimization rather than on lead identification.

## Conclusion

The LIECE model described in the present paper was derived from a large number of observations ( $n = 27$ ) and demonstrated a solid statistics reliability ( $r^2 = 0.698$ ;  $q^2 = 0.662$ ) as well as a sound and unequivocal physical meaning. Based on benzamidine structures, primarily extracted from X-ray complexes and further complemented by other benzamidine thrombin inhibitors, the LIECE model disclosed a satisfactory and, to some extent, unexpected ability in prediction and, more importantly, screening analyses. The former was challenged by using an even larger external test set comprising a series of 88 thrombin inhibitors. On the other hand, the LIECE model exhibited a remarkable capability in fishing actives from a large combinatorial library of 10240 mimetic compounds. Interestingly, the LIECE model outperformed other standard methods (i.e., docking and shape-based methods) irrespective of the metrics and/or of the experimental conditions designated for the screening. Incidentally, the analysis of the true positives signalled the relevance of the phenoxyphenyl fragment, a privileged substructure recurrent for different drug target types and, in the case of thrombin enzyme, supposedly engaging localised hydrophobic interactions with the subsite D. Taken together, these results strengthened the confidence in the proposed LIECE model as a promising tool for the successful optimization of benzamidine-like leads and ultimately for the design of novel thrombin inhibitors. In this view, the nearer immediate and pragmatic use will be the virtual screening analysis of an *in silico* designed library biased toward thrombin.

**Acknowledgments** Credits are due to Prof. Amedeo Caflisch for fruitful and critical discussions. The authors thank the talented and dedicated graduate students Dr. Nicola Labianca, Dr. Giuseppe Mangiatordi, Dr. Lydia Siragusa and Dr. Paola Tedeschi. A grateful acknowledgement goes to the University of Bari and MIUR (Rome, Italy; PNR project RBNE01F5WT).

## References

- Jain AN, Koile K, Chapman D (1994) Compass: predicting biological activities from molecular surface properties: performance comparisons on a steroid benchmarks. *J Med Chem* 37:2315–2327
- Mestres J, Rohrer DGC, Maggiora G (1997) MIMIC: a molecular field matching program. Exploiting applicability of molecular surface approaches. *J Comput Chem* 18:934–954

3. Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB (2002) Consensus scoring for ligand/protein interactions. *J Mol Graphics Model* 20:281–295
4. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* 16:11–26
5. Wei H, Tsai K, Lin T (2005) Modeling ligand-receptor interaction for some MHC class II HLA-DR4 peptide mimetic inhibitors using several molecular docking and 3D QSAR techniques. *J Chem Inf Model* 45:1343–1351
6. Prathipati P, Saxena AK (2006) Evolutionary of binary QSAR models derived from LUDI and MOE scoring functions for structure based virtual screening. *J Chem Inf Model* 46:39–51
7. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhatt N, Weissig M, Weissig M, Bourne PE (2000) The protein data bank. *Nucleic Acid Res* 28:235–242
8. Nicolotti O, Miscioscia TF, Leonetti F, Muncipinto G, Carotti A (2007) Screening of matrix metalloproteinases available from the PDB: insights into biological functions, domain organization and zinc binding groups. *J Chem Inf Model* 47:2439–2448
9. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CLIII (2004) Assessing scoring functions for protein-ligand interactions. *J Med Chem* 47:3032–3047
10. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* 49:5912–5931
11. Kontoyianni M, McClellan LM, Sokol GS (2004) Evaluation of docking performance: comparative data on docking algorithms. *J Med Chem* 47:558–565
12. Nicolotti O, Miscioscia TF, Carotti A, Leonetti F, Carotti A (2008) An integrated approach to ligand- and structure-based drug design: development and application to a series of serine protease inhibitors. *J Chem Inf Model* 48:1211–1226
13. Nicolotti O, Giangreco I, Miscioscia TF, Carotti A (2009) Investigating enzyme selectivity and hit enrichment by automatically interfacing ligand- and structure-based molecular design. *QSAR & Comb Sci* 28:861–864
14. Nicolotti O, Giangreco I, Miscioscia TF, Carotti A (2009) Improving quantitative structure–activity relationships through multiobjective optimization. *J Chem Inf Model* 49:2290–2302
15. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J Med Chem* 49:5851–5855
16. Beveridge DL, DiCapua FM (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu Rev Biophys Chem* 18:431–492
17. Kollman P (1993) Free-energy calculations: applications to chemical and biochemical phenomena. *Chem Rev* 93:2395–2417
18. Tembe BL, McCammon JA (1984) In: *Comput Chem* (ed) *Ligand-Receptor Interactions*. Elsevier, Oxford
19. Hansson T, Åqvist J (1995) Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng* 8:1137–1144
20. Åqvist J (1996) Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J Comput Chem* 17:1587–1597
21. Lamb ML, Jorgensen WL (1997) Computational approaches to molecular recognition. *Curr Opin Chem Biol* 1:449–457
22. Åqvist J, Medina C, Samuelsson JE (1994) New method for predicting binding affinity in computer-aided drug design. *Protein Eng* 7:385–391
23. Carlson HA, Jorgensen WL (1995) An extended linear response method for determining free energies of hydration. *J Phys Chem* 99:10667–10673
24. McDonald NA, Carlson HA, Jorgensen WL (1997) Free energies of solvation in chloroform and water from a linear response approach. *J Phys Org Chem* 10:563–576
25. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 112:6127–6129
26. Lee MC, Yang R, Duan Y (2005) Comparison between generalized-born and poisson-boltzmann methods in physics based scoring functions for protein structure prediction. *J Mol Model* 12:101–110
27. Lee MR, Sun Y (2007) Improving docking accuracy through molecular mechanics generalized born optimization and scoring. *J Chem Theory Comput* 3:1106–1119
28. Guimarães CRW, Cardozo M (2008) MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J Chem Inf Model* 48:958–970
29. Huang D, Caflisch A (2004) Efficient evaluation of binding free energy using continuum electrostatics solvation. *J Med Chem* 47:5791–5797
30. Thompson DC, Humblet C, Joseph-McCarty D (2008) Investigation of MM-PBSA rescoring of docking poses. *J Chem Inf Model* 48:1081–1091
31. Grazioso G, Cavalli A, De Amici M, Recanatini M, De Micheli C (2008) Alpha7 nicotinic acetylcholine receptor agonists: prediction of their binding affinity through a molecular mechanics poisson-boltzmann surface area approach. *J Comput Chem* 29:2593–2602
32. Huang D, Caflisch A (2009) Library screening by fragment-based docking. *J Mol Recognit*. doi: [10.1002/jmr.981](https://doi.org/10.1002/jmr.981)
33. Zhou T, Huang D, Caflisch A (2008) Is quantum mechanics necessary for predicting binding free energy? *J Med Chem* 51:4280–4288
34. Kolb P, Huang D, Caflisch A (2008) Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J Med Chem* 51:1179–1188
35. Chen Z, Li Y, Mulichak AM, Lewis SD, Shafer JA (1995) Crystal structure of human  $\alpha$ -thrombin complexed with hirugen and p-amidinophenylpyruvate at 1.6 Å resolution. *Arch Biochem Biophys* 322:198–203
36. Katz BA, Elrod K, Luong C, Rice MJ, Mackman RL, Sprengeler PA, Spencer J, Hataye J, Janc J, Link J, Litvak J, Rai R, Rice K, Sideris S, Verner E, Young W (2001) A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. *J Mol Biol* 307:1451–1486
37. Dullweber F, Stubbs MT, Musil D, Sturzebecher J, Klebe G (2001) Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J Mol Biol* 313:593–614
38. Huel NH, Nar H, Pripke H, Ries U, Stassen JM, Wienen W (2002) Structure-based design of novel potent non peptide thrombin inhibitors. *J Med Chem* 45:1757–1766
39. Schaerer K, Morgenthaler M, Seiler P, Diederich F, Banner DW, Tschopp T, Obst-Sander U (2004) Enantiomerically pure thrombin inhibitors for exploring the molecular-recognition features of the oxyanion hole. *Helv Chim Acta* 87:2517–2539
40. Fokkens J, Klebe G (2006) A simple protocol to estimate differences in protein binding affinity for enantiomers without prior resolution of racemates. *Angew Chem Int Ed Engl* 45:985–989
41. Schweizer E, Hoffmann-Roeder A, Olsen JA, Seiler P, Obst-Sander U, Wagner B, Kansy M, Banner DW, Diederich F (2006) Multipolar interactions in the D pocket of thrombin: large differences between tricyclic imide and lactam inhibitors. *Org Biomol Chem* 4:2364–2369
42. Brandstetter H, Turk D, Hoeffken HW, Grosse D, Stürzebecher J, Martin PD, Edwards BF, Bode W (1992) Refined 2.3 Å X-ray



- crystal structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA. A starting point for improving antithrombotics. *J Mol Biol* 226:1085–1099
43. Breu B, Silber K, Gohlke H (2007) Consensus adaptation of field for molecular comparison (AFMoC) models incorporate ligand and receptor conformational variability into tailor-made scoring function. *J Chem Inf Model* 47:2283–2400
  44. Böhm M, Stürzebecher J, Klebe G (1999) Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J Med Chem* 42:458–477
  45. Murcia M, Morreale A, Ortiz AR (2006) Comparative binding energy analysis considering multiple receptors: a step toward 3D-QSAR models for multiple targets. *J Med Chem* 49:6241–6253
  46. Illgen K, Enderle T, Broger C, Weber L (2000) Simulated molecular evolution in a full combinatorial library. *Chem Biol* 7:433–441
  47. Maestro, version 7.5.112 (2006) Schrödinger LLC, New York
  48. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS All-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236
  49. SYBYL, version 7.1 (2007) Tripos Inc, St. Louis, MO
  50. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrikson T, Still WC (1990) Macro-model: an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J Comput Chem* 11:440–467
  51. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. *J Med Chem* 49:6789–6801
  52. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S et al (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4: 187–217
  53. Warwicker J, Watson HC (1982) Calculation of the electric potential in the active site cleft due to R-helix dipoles. *J Mol Biol* 157:671–679
  54. Im W, Beglov D, Roux B (1998) Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput Phys Commun* 111: 59–75
  55. Widmer A, Novartis Pharma, Basel, Switzerland
  56. No K, Grant J, Scheraga H (1990) Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides. *J Phys Chem* 94:4732–4739
  57. No K, Grant J, Jhon M, Scheraga H (1990) Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 2. Application to ionic and aromatic molecules as models for polypeptides. *J Phys Chem* 94:4740–4746
  58. Jones G, Willett P, Glen RC, Leach ARL, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267:727–748
  59. Nissink JW, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R (2002) A new test set for validating predictions of protein-ligand interaction. *Proteins* 49:457–471
  60. ROCS, version 2.4.2 (2005) OpenEye Scientific Software Inc, Santa Fe, NM, USA, <http://www.eyesopen.com>
  61. Willett P (2006) Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today* 11:1046–1053
  62. Gerstein M, Levitt M (1998) Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins. *Protein Sci* 7:445–456
  63. Wold S, Trygg J, Berglund A, Antti H (2001) Some recent developments in PLS modeling. *Chemom Intell Lab Syst* 58: 131–150
  64. Jones-Hertzog DK, Jorgensen WL (1997) Binding affinities for sulfonamide inhibitors with human thrombin using monte carlo simulations with a linear response method. *J Med Chem* 40:1539–1549
  65. Ljungberg KB, Marelus J, Musil D, Svensson P, Norden B, Åqvist J (2002) Computational modeling of inhibitor binding to human thrombin. *Eur J Pharm Sci* 12:441–446
  66. Zhou R, Friesner RA, Ghosh A, Rizzo RC, Jorgensen WL, Levy RM (2001) New linear interaction method for binding affinity calculations using a continuum solvent model. *J Phys Chem* 105:10388–10397
  67. Nicolotti O, Carotti A (2006) QSAR and QSPR studies of a highly structured physico-chemical domain. *J Chem Inf Mod* 46:264–276
  68. Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 28:849–857
  69. Cramer RD III, Patterson DE, Bounce JD (1988) Comparative Molecular Field Analysis (CoMFA) 1. Effect of shape on binding of Steroids to carrier proteins. *J Am Chem Soc* 110:595–5967
  70. Van Drie JH (2003) Pharmacophore discovery—lessons learned. *Curr Pharm Des* 9:1649–1664
  71. Akifumi O, Keiichi T, Tadakazu T, Noriyuki Y, Shuichi H (2006) Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *J Chem Inf Model* 46:380–391
  72. Nicholls A (2008) What do we know and when do we know it? *J Comput Aided Mol Des* 22:239–255
  73. Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. *J Med Chem* 39:2887–2893
  74. Bemis GW, Murcko MA (1999) Properties of known drugs. 2. Side chains guy. *J Med Chem* 42:5095–5099
  75. Comprehensive Medicinal Chemistry (CMC-3D) Release 94.1. MDL Information Systems Inc, San Leandro, CA
  76. Olah M, Mracec M, Ostopovici L, Rad R, Bora A, Hadaruga N, Olah I, Banda M, Simon Z, Mracec M, Oprea TI (2004) In: Chemoinformatics in Drug Discovery. Oprea TI (ed), WOMBAT: World of Molecular Bioactivity. Wiley-VCH, New York
  77. Olah M, Oprea TI (2006) Comprehensive Medicinal Chemistry II vol 3. Taylor JB, Triggle DJ (eds). Bioactivity Databases. Elsevier, Oxford
  78. Olah M, Rad R, Ostopovici L, Bora A, Hadaruga N, Hadaruga D, Moldovan R, Fulias A, Mracec M, Oprea TI (2007) Chemical biology: from small molecules to systems biology and drug design. In: Schreiber SL, Kapoor TM, Wess G (eds), WOMBAT and WOMBAT-PK: Bioactivity Databases for Lead and Drug Discovery. Wiley, New York
  79. Imming P, Sinning C, Meyer A (2006) Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* 5:821–834