Estimation of the Binding Free Energy by Linear Interaction Energy Models

O. Nicolotti*, M. Convertino, F. Leonetti, M. Catto, S. Cellamare and A. Carotti

Dipartimento Farmaco-Chimico, University of Bari, via Orabona 4, I-70125 Bari, Italy

Abstract: Since Hansch's extra thermodynamic multi-parameter approach, originally coined as Linear Free Energy Relationship, great efforts in medicinal chemistry have been made to properly estimate the binding free energy. Despite the often small amount, its value is however very critical in determining a successful binding. As a result, its correct estimation may provide a guide for a prospective rational drug design. The calculation of the absolute binding free energies is however a very challenging task as it requires a rigorous treatment of a number of physical terms that are both very time demanding and to some extent not immediately interpretable. In view of this, the introduction of some numerical approximations has permitted to develop the so called Linear Interaction Energy method that, at present, constitutes the best compromise among accuracy, speed of computation and easy interpretation. The initially developed Linear Interaction Energy method was subsequently revisited and several important improvements have been made. Significant examples are the Extended Linear Response, the surface generalized Born LIE, the molecular mechanics generalized Born surface area, the linear interaction energy in continuum electrostatics as well as its quantum mechanics variant. Principles and selected applications of these methods will be herein reviewed.

Keywords: Binding free energy, linear interaction energy, extended linear response, molecular mechanics generalized born surface area, linear interaction energy in continuum electrostatics.

INTRODUCTION

Free energy is a central concept of physical and medicinal chemistry, describing the tendency of molecular systems to spontaneously associate and react. The accurate prediction of this valuable amount of energy, via computational approaches, is thus a really important goal to achieve for physicists, chemists, and biochemists. In recent years, much progress has been made applying molecular dynamics (MD) or Monte Carlo (MC) methods to carry out the free energy calculation with a particular emphasis in the area of ligand binding to macromolecules such as proteins and nucleic acids. In this scenario, in fact, molecular docking scoring functions have failed the prediction of Binding Free Energy (BFE) of even structurally close compounds with the results that no statistically significant correlation between experimentally measured biological activities and scoring values has so far emerged [1,2].

As a matter of fact, Free Energy Perturbation (FEP) and thermodynamic integration (TI) are the most rigorous, MD or MC-based, computational approaches currently used to calculate the BFE [3,4]. Despite their wide use, both of them are however very time demanding and, thus, of often limited feasibility [5-8]. Moreover, limitations in conformational sampling [9,10] and force field parameterization might compromise FEP and TI accuracy. For instance, it is well known that molecular force fields are sometimes too simplified to properly describe intermolecular interactions with the serious risk of obtaining unconfident or even unreliable predictions of BFE values. On the other hand, especially over a course of a long MD or MC calculation, it can be difficult to exhaustively explore the search space for an efficient sampling of all the ligand-protein conformations and to assess the loss of entropy due to the occurrence of intermolecular interactions as well [11,12].

In view of this, a reasonable treatment of the errors and uncertainties in the determination of free energy has been made possible by the development of approximated BFE calculation approaches. A notable example is the Linear Interaction Energy (LIE) method that gives an estimation of the BFE from only two simulations of the solvated protein-ligand complex and the ligand in solution. A side benefit is that LIE is about one order of magnitude faster than the FEP and TI methods [13]. However, LIE needs to be calibrated with a training set of ligands provided with known and definite experimental binding data. [14].

Further advancements of the LIE model were achieved in the last ten years leading to the development of improved models in which implicit solvent and different descriptors are used. Among others, this survey will be focused on the Extended Linear Response (ELR) [15-17]; the surface generalized Born LIE (SGB-LIE) [18]; the molecular mechanics generalized Born surface area (MM-GB/SA) and its Poisson-Boltzman variant (MM-PB/SA)[19-23], the linear interaction energy in continuum electrostatics (LIECE) and its quantum mechanics variant (QMLIECE) [24].

LINEAR INTERACTION ENERGY: LIE

Initially developed by Åqvist on a set of endothiapepsin inhibitors [13], the Linear Interaction Energy (LIE) method

*Address correspondence to this author at the via E. Orabona 4, I-70126 BARI, Italy; Tel: ++39.080.5442551; Fax: ++39.080.5442230;
E-mail: nicolotti@farmchim.uniba.it

1875-5607/12 $58.00+.00

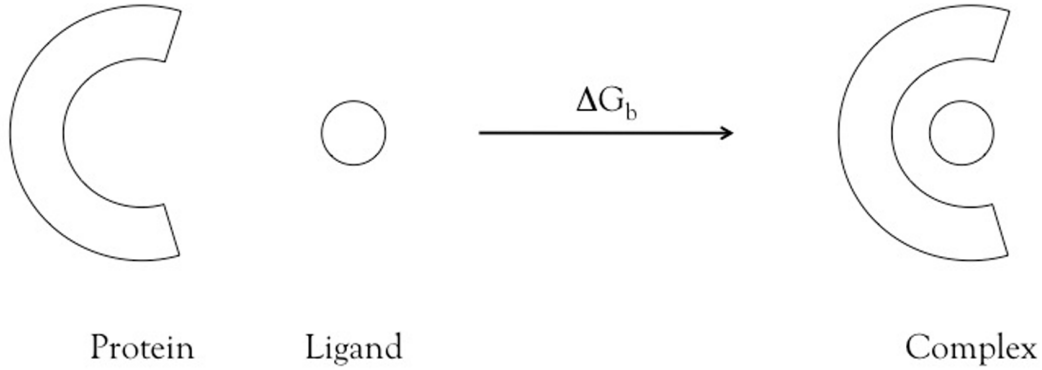


Fig. (1). Schematic representation of a protein-ligand binding event.

is a semiempirical approach, which requires only two simulations of the solvated protein-ligand complex and the ligand in solution for estimating the BFE. A general version of the equation used to evaluate the BFE is:

$$\Delta G_b = \alpha\langle \Delta E_{vdw} \rangle + \beta\langle \Delta E_{Coul} \rangle + \gamma \quad (1)$$

where $\langle ... \rangle$ stands for MD or MC ensemble energy average of non-bonded van der Waals (E_{vdW}) and electrostatic (E_{Coul}) interactions between the ligand and its surrounding environment, i.e., either the solvated receptor binding site (bound state) or just solvent (free state). The Δ in equation (1) indicates the difference between such averages in the bound and free state (Fig. **1**). α and β are the weight coefficients for the non-polar and electrostatic BFE contributions, respectively. The parameter α = 0.161 was derived to give the best fit to experimental binding data (note that $\Delta G_{exp} \approx$ -RTln(K_i) or $\approx$ - RTln(IC_{50})) and the electrostatic scaling factor β = 0.50 follows from the quadratic dependence of free energy on solute charge, as embodied in the Born model for ion solvation [25]. Further works demonstrated that β = 0.50 is valid only for ligands bearing charged groups, while neutral dipolar compounds have a systematic dependence of their content of dipolar groups [6,26]. The parameter γ is an additional constant number which was initially set to zero, while in following investigations conducted over a series of human thrombin inhibitors [27] was set to ~ -3 kcal/mol to accurately predict the experimental data showing that systematic error. In some cases γ has been expressed in terms of solvent accessible surface area (SASA) of the solute [28,29] and, thus, equation (1) has been changed as follows:

$$\Delta G_b = \alpha\langle \Delta E_{vdW} \rangle + \beta\langle \Delta E_{Coul} \rangle + \gamma\langle \Delta SASA \rangle \quad (2)$$

In the modified approach expressed in (2), both α and β coefficients can vary in order to achieve the best fit to experimental binding data, while the γ-scaled SASA term is included as a means of accounting for possible positive free energies of hydration caused by solute cavity formation in the solvent [30,31].

EXTENDED LINEAR RESPONSE: ELR

Given the success of the previously described MD [6,13,25,32-34] or MC-based [28,29,35] LIE binding studies, Jorgensen started to treat larger data set (thrombin inhibitors [15], HIV-1 reverse transcriptase non-nucleoside inhibitors [16], COX-2 inhibitors [17]) with the purpose of investigating the existence of possible correlations with experimental data at the increase of the ratio of data points to parameters. In this study a large number of physicochemical descriptors was applied, including among others, counts of hydrogen bonds, hydrophilic, hydrophobic, aromatic surfaces area. A multivariate fitting approach was used leading to the following general expression:

$$\Delta G_b = \sum_n c_n \xi_n + C \quad (3)$$

where c_n is an optimizable coefficient for its associated descriptor ξ_n and C is a constant. In principle, any physical meaningful quantity could be used as descriptor in the Extended Linear Response (ELR) method, although descriptors with an understandable physical interpretation are likely to afford models having greater interpretability and prediction utility.

However, it should be noted that ELR are derived using experimental data in conjunction with descriptors obtained from computer simulations to derive a regression expression. Once a solid cross-validated equation is derived, no additional experimental data are thus necessary for activity predictions of novel compounds.

From these applications, the importance of few key descriptors emerged and, as a result, was widely recognized. For instance descriptors accounting for van der Waals interactions and for the change in number of hydrogen bonds for the ligand upon binding are considered of great importance, as pointed out by Jorgensen in a regression-based study over a series of non-nucleoside reverse transcriptase inhibitors (Fig. **3**), potentially useful in anti-HIV1 therapies [16]. In the cited example, the most common significant descriptors for individual regression equations were those terms accounting for the ligand-protein Lennard-Jones interactions (EXX_{LJ}) and the change in total number of hydrogen bonds for the inhibitor upon binding ($\Delta HBtot$), as shown in Table **1**.

LINEAR INTERACTION ENERGY AND CONTINUUM SOLVENT MODELS: SGB-LIE

From a computational standpoint, LIE and ELR models are attractive compared to FEP because they are not so much time demanding and may thus permit the study of larger number of ligands. The cheaper computational cost is

Table 1. Significant Descriptors Used in Regression Equations that Incorporate Multiple Non-Nucleoside Reverse Transcriptase Inhibitors as Reported in Fig. (3)

Core	Name	r^2	EXX_{LJ}	ΔHB_{tot}	Water-Bridges	qp_#Rotor	ΔFOSA	ΔWPSA	ΔPISA	ΔDipole	qp_ΔG_{hyd}	DtoProPi
3	Nevirapine	0.54	+	-	-	+						-
4	ASBN	0.74	+									
5a	Sustiva	0.67	+					+				
8	9-Cl-TIBO	0.79	+	-								

[a]Descriptors in common share a + or – in the same column, which also indicates the sign of the fitted coefficient. Desciptors codes; EXX_{LJ}: ligand-protein Lennard-Jones interactions; ΔHB_{tot}: change in total number of hydrogen bonds for the inhibitor upon binding; *water-bridges*: number of bridging water molecole that mediate hydrogen bonding between ligand and protein; *qp_#rotor*: number of rotatable bonds in the ligand;*ΔFOSA*: change in hydrophobic solvent accessibile surface area (SASA); *ΔWPSA*: change in the weakly polar (halogens, P, and S) SASA; *ΔPISA*: change in aromatic SASA; *Δdipole*: change in dipole moment of inhibitor; $qp\_\Delta G_{hyd}$: estimate of the free energy of hydration for the inhibitor obtained using the QikProp program (v.1.67 Schrödinger Inc., New York, 2001); *DtoProPri*: number of hydrogen bonds donated by the ligand to a protein π system. Data were taken from Rizzo *et al*. [16].

determined by the fact that only interactions between the ligand and either the protein or the aqueous environment are explicitly considered as descriptors in the resulting model (protein-protein and protein-water interactions are only used to generate conformations via MD or MC). As a result, such approaches enabled to prevent a considerable amount of noise and systematic uncertainties in the calculations as, for example, those arising from different conformations of the protein obtained from co-crystallized structures of different ligands.

Further developments were introduced by Zhou; in his work the use of explicit solvent in the simulation was replaced by a continuum surface generalized Born (SGB) solvent model [18], which demonstrated to be more than one order of magnitude faster than the previously described LIE models. Interestingly, the SGB-LIE approach performs a deeper exploration of the conformational space and it is even much more rapid to reach the convergence threshold of the simulations because of the absence of the explicit water friction (see Table **2**). The SGB-LIE model replaces the solvent accessible surface area term in Jorgensen's LIE [28] formulation (2) by the cavity term (ΔU_{cav}) in continuum solvent model:

$$\Delta G_b = \alpha\left\langle \Delta E_{vdW} \right\rangle + \beta\left\langle \Delta E_{Coul} \right\rangle + \gamma\left\langle \Delta U_{cav} \right\rangle \quad (4)$$

In the generalized Born models the solvation free energy is given by the sum of two terms called reaction field energy (U_{rnx}) and cavity energy (U_{cav}) [19,20]:

$$U_{SGB} = U_{rnx} + U_{cav} \quad (5)$$

On the other hand, in the SGB-LIE there is not explicit electrostatic and van der Waals energy between solute and solvent anymore; the van der Waals energy is thus implicitly included in the cavity term, which is based on the total accessible surface area (SASA):

$$U_{cav} = c_1 SASA + c_2 \quad (6)$$

where c_1 and c_2 are empirical coefficients with c_1 = 0.00486 kcal/(mol Å^2) and c_2 = 1.092 kcal/mol [18,20,21]. The total electrostatic energy term in (4) is given by the sum of the possible Coulomb interactions between ligand and protein (U_{Coul}) and two times the reaction field energy (U_{rxn}):

$$E_{Coul} = U_{Coul} + 2U_{rxn} \quad (7)$$

This simplification is made possible because the reaction field energy (U_{rxn}) is half of the Coulomb energy between solute and solvent (U_{Coul}) in the linear interaction approximation in the case of small molecules' solvation, as proved by Jorgensen [30]. The calculation of the reaction field energy (U_{rxn}) is straightforward for the ligand in its free state [20] while it is more difficult for the ligand in the bound state because SGB accounts for the U_{rxn} for the whole solute, although the U_{rxn} from the protein alone is not needed in SGB-LIE. Zhou, therefore, screened the pairwise coulombic component of the U_{rxn} as a whole for atoms belonging to the ligand, as half for atom belonging to the ligand and protein and zero for atoms belonging to the protein only. The reader interested to know the exact mathematical details is referred to elsewhere [18]. The SGB-LIE has been applied to three different ligand sets: HIV-1 reverse transcriptase inhibitors (1-[(2-hydroxyethoxy) methyl]-6-(phenylthio)thymine (HEPT) analogues, 20 ligands), human thrombin inhibitors (sulfonamide derivatives, 7 ligands) and various ligands binding the coagulation factor Xa (8 ligands). The results obtained with SGB-LIE seem to be quite good in terms of quality of the prediction (fitting and cross-validation results show that about 1.0 kcal/mol accuracy is achievable) and time calculation performance, as shown in Table **2** [18].

LINEAR INTERACTION ENERGY AND CONTINUUM SOLVENT MODELS: MM-GB/SA AND MM-PB/SA

The encouraging results obtained with the continuum solvent model inspired different authors in continuing the work to adopt and improve, the so-called molecular mechanics generalized Born surface area (MM-GB/SA) and its Poisson-Boltzmann variant (MM-PB/SA) [19,22,23]. The major efforts were directed to the rescoring of docking results and in the combined use of molecular mechanics and continuum solvation to compute average BFE of ligands, taking into account their bound and unbound states and using MD or MC simulations. An interesting application was presented by Guimarães [36], who investigated the performance of MM-GB/SA rescoring in structure-based lead optimization for quite diverse sets of pharmaceutically

Table 2. CPU Comparison of the SGB-LIE Model with the Explicit Solvent LIE Model[a]

Model	State	Data Collected	CPU/ps	Total CPU
explicit	free	125 ps	0.370 h	3.74 d
	bound	110 ps	0.336 h	3.08 d
continuum	free	30 ps	0.11 min	4.95 min
	bound	30 ps	0.291 h	0.546 d

[a]Data were taken from Zhou *et al*. [18], who collected them for HEPT (H01) binding to HIV-1RT, and the explicit solvent model is based on the solvation of a 20 Å water sphere from the center of the ligand, as used by Åqvist *et al*. [6]. The CPU timing is obtained from IBM Power3-375MHZ SP2 cluster and the MD was run with a time step of 2 fs in both simulations using multiple time step algorithm RESPA [M. Tuckerman *et al*., *J. Chem. Phys.,* **1992**, *112*, 1990] ("min" for minutes, "h" for hours and "d" for days).

relevant targets such as CDK2, factor Xa, thrombin and HIV-1 reverse transcriptase. In his implementation each ligand was docked into the protein binding site followed by a MC search to perform the conformational analysis of the ligand in the bound state, and by a conformational search for the inhibitor in the unbound state. An energy minimization was thus performed over the ligand-protein complex with the aim of increasing the computational efficiency compared to MD simulation. However, the lack of sampling in this method might represent a serious drawback since the protein would not be able to fully relax when accommodating different scaffolds after docking. Nevertheless this issue should be minimized when scoring congeneric series of ligands (because of a systematic error which will globally not affect the ligand ranking). The binding energy was then computed as follows:

$$\Delta G_b = \Delta E_{intra} + \Delta G_{solv} - T\Delta S_{conf} + E_{vdW} + E_{Coul} + E_{PTN} \quad (8)$$

where ΔE_{intra} and ΔG_{solv} are the intramolecular and desolvation penalties for each ligand upon binding while E_{vdW} and E_{Coul} are the intermolecular protein-ligand van der Waals and electrostatic interaction energies, respectively, and E_{PTN} is the protein energy as extracted from the complexes. Relevant is the presence of the entropic term $T\Delta S_{conf}$, which accounts for the ligand conformational entropy penalty (multiplied by the temperature), computed from the Boltzmann distribution in water. In fact, assuming a Boltzmann distribution, the probabilities for each conformer (P_i) were calculated as follows:

$$P_i = \frac{exp(-E_i / kT)}{\sum_i exp(-E_i / kT)} \quad (9)$$

where E_i is the sum of internal energy and hydration energy of the ligand i and k is the Boltzmann constant. The Boltzmann-averaged intermolecular energy and solvation free energy in the unbound state for every compound were also obtained and the conformational entropies (S_{conf}) were calculated from the probabilities according to the following equation:

$$S_{conf} = -k\sum_{i=1}^{n} P_i \ln P_i \quad (10)$$

MM-GB/SA, as developed by Guimarães, proved to be far superior to docking scoring functions and at the same time to be as accurate as FEP or TI methods, with the main advantage to handle more structurally dissimilar compounds with less computational time. This study demonstrated that the conformational entropy penalty term, derived by a Boltzmann distribution, was very small and similar for all the examined sets of compounds even for molecules with different degree of flexibility. This suggests that the entropic contribution should not be so important for ranking-ordered, especially in congeneric series. However, it should be considered that the poor estimation of the conformational entropy penalty term can derive from inaccuracies in the currently available force fields. Further approximation may also derive from the algorithms used in GB or PB continuum solvent models as remarked by Cavalli [37].

A convincing variant to the described MM-GB/SA technique is the MM-PB/SA. Initially developed by Kollman [22], MM-PB/SA uses a combination of molecular mechanics and continuum solvation (computed solving the Poisson-Boltzmann equation) to calculate the average binding energies for a set of ligands.

An interesting study was presented by Thompson, [38] who applied the MM-PB/SA to discriminate docking poses, using a subset of the CCDC/Astex test set [39] and a set of actives/inactives from the DUD data set [40]. Each of the terms used for the prediction of the binding free energy (ΔG_b) is approximated by computing values for the ligand-protein complex and the two molecular structures are individually taken from the minimized complexes.

The following equation was used for the prediction of the binding free energy:

$$\Delta G_b = \Delta H_{Elec}^{Sol} + \Delta H_{vdW} - T\Delta S_{Hphob} - T\Delta S_{RotB} \quad (11)$$

where ΔH_{Elec}^{Sol} and ΔH_{vdW} represent the Poisson-Boltzmann electrostatic and the van der Waals energy, respectively. $T\Delta S_{Hphob}$ is the hydrophobic term computed as follows:

$T\Delta S_{Hphob}$ = (surface area buried upon complex formation) × 0.006 kcal/(mol Å^2) (12)

the numerical coefficient (0.006 kcal/(mol Å^2)) accounts for the partitioning of solute molecules between aqueous and organic phases [41,42]; $T\Delta S_{RotB}$ is a rotatable bond penalty included to quantify the loss of binding energy due to the freezing of the internal degrees of freedom of the ligand [43].

$T\Delta S_{RotB}$ = (number of rotatable bonds) × 0.7 kcal/mol (13)

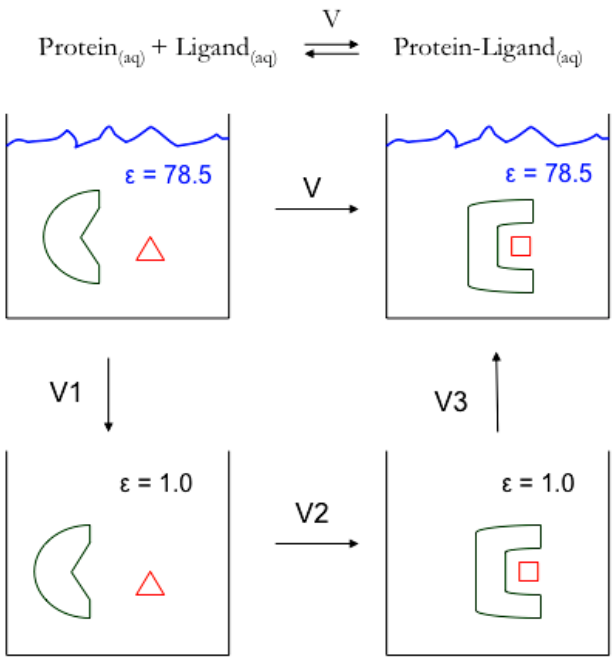


Fig. (2). Thermodynamic cycle used in LIECE calculations.

Interestingly, also in this study it was revealed that the entropic terms tend to be less important in the determination of ΔG_b; in fact, the examination of $T\Delta S_{Hphob}$ and $T\Delta S_{RotB}$ revealed that for any given complex these two terms are similar in magnitude but of opposite sign, and their weight is accounting only for the 10% of the binding free energy final value while the contribution of the electrostatic and van der Waals terms is approximately equal to the 90% of the total ΔG_b in the above described implementation.

LINEAR INTERACTION ENERGY IN CONTINUUM ELECTROSTATICS: LIECE AND QMLIECE

Another important variant of LIE was initially developed by Huang [44], in which the MD sampling was replaced by a simple energy minimization and combined the LIE method with a rigorous treatment of continuum electrostatics based on the numerical solution of the Poisson equation by finite difference techniques [45].

The new technique was named LIECE (Linear Interaction Energy in Continuum Electrostatics) and revealed a higher computational time efficiency being more than two orders of magnitude faster than LIE, as well as, a satisfying predictive accuracy, reaching about 1.0 kcal/mol for thirteen and twenty-nine inhibitors of β-secretase (BACE) and HIV-1 protease (HIV-PR), respectively. The approach was also successfully applied to a series of virtual screening experiments which finally lead to the discovery of cell-permeable β-secretase inhibitors with inhibitory activity in the low-micromolar range, as well as a potent and selective nanomolar inhibitor of the tyrosine kinase erythropoietin producing human heptocellular carcinoma receptor B4 (EphB4) [46-50]. The two cited targets were used for deriving a two-parameter model [13]:

$$\Delta G_b = \alpha \Delta E_{vdW} + \beta \Delta G_{elec} \quad (12)$$

and a three-parameter model [25]:

$$\Delta G_b = \alpha \Delta E_{vdW} + \beta \Delta G_{elec} + \Delta G_{tr,rot} \quad (13)$$

where ΔE_{vdW} is the ligand-protein van der Waals interaction energy, ΔG_{elec} is the sum of the ligand-protein Coulomb energy *in vacuo* and the change in solvation energy of ligand and protein upon binding (see Fig. **2**), and $\Delta G_{tr,rot}$ is an entropic term representing the loss of translational and rotational degree of freedom upon binding. As it was already discussed in the previous paragraphs, also in this case the presence of an entropic term does not improve the prediction accuracy [42].

Interestingly, LIECE was also successful over different sets of protein kinase inhibitors (CDK2, Lck, p38, Wee1) [51,52]; in this study five different models were created and validated on three different protein kinases. In addition to equation (12) a one-parameter model:

$$\Delta G_b = \alpha \Delta E_{vdW} \quad (14)$$

and a three-parameter model:

$$\Delta G_b = \alpha \Delta E_{vdW} + \beta_1 \Delta E_{Coul} + \beta_2 \Delta G_{solv} \quad (15)$$

were used to fit the calculated energy terms. In the latter equation the electrostatic term of (12) (ΔG_{elec}) is split in its two energetic components: ligand-protein Coulomb energy *in vacuo* (ΔE_{Coul}) and the change in solvation energy of ligand and protein upon binding (ΔG_{solv}). Further interesting models were also proposed with the introduction of an entropic term, as the third parameter in (12) and (13), but again, this contribution did not furnish any substantial

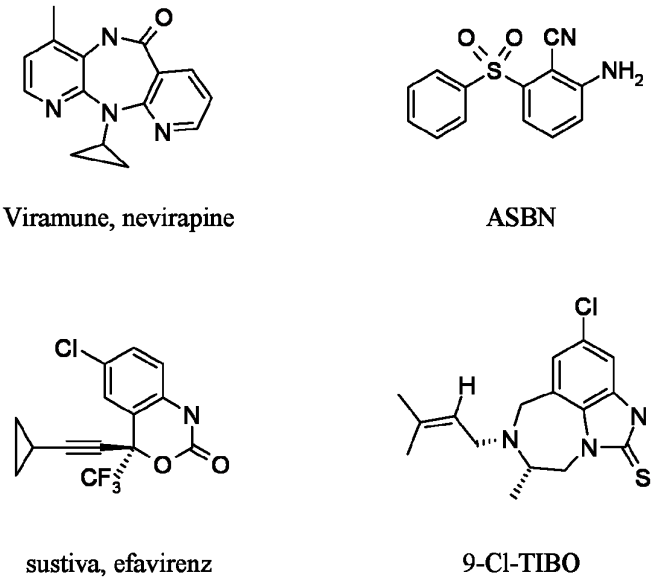


Fig. (3). Some of non-nucleoside reverse transcriptase inhibitors used in LRE model by Rizzo R. C. *et al.* [16]: nevirapine, efavirenz, 9-Cl-4,5,6,7-tetrahydroimidazo-[4,5,1-jk][1,4]benzodiazepine-2(1H)-thione (9-Cl-TIBO), 2-amino-6-phenylsulfonylbenzonitrile (ASBN).

improvement to the model predictive accuracy (see Table **3** for a summary of LIECE performance over different targets).

Huang and Kolb performed detailed studies about the parameter transferability among their respective data sets. In contrast with Huang's observation on BACE and HIV-1 PR, Kolb showed that the parameter transferability among protein kinases was due to the predominance of the van der Waals interaction (whose multiplicative parameter is between five and ten times larger than the electrostatic term one). This is likely due to the typical shape and structure of ATP-binding site that is highly conserved with the results that different kinases share a high degree of steric similarity.

On the other hand, LIECE parameters are not transferable between human BACE and HIV-1 PR, despite they are both aspartic proteases. The lack of generality in this case is probably due to the significant diversity existing between the substrate-binding site of mammalian and viral aspartic proteases.

In this regard, the large number of computational experiments performed with LIECE remarked the extreme sensibility of this technique even to small conformational changes [50], as well as its obvious tendency to give better performance in interpolating rather than in extrapolating new external data [53]. It is worthy saying that LIECE model tends also to be affected by the presence of false positives (due to misleading interaction energy values) which, however, can be easily filtered out by a preliminary selection of docking conformations by means of filtering criteria based on energy efficiency or, alternatively, monitoring the occurrence of specific ligand-protein interactions in the obtained ligand binding modes [51].

A recent application of a LIECE model was also developed and validated on a series of benzamidine-based inhibitors of human thrombin [50]. In this example, a large number of observations (n_{obs} = 27), primarily extracted from X-ray complexes and further complemented with other benzamidine thrombin inhibitors, was successfully used to derive a model with satisfactory ability in prediction (over an almost three times larger external test set, n_{test} = 88). Interestingly such a model demonstrated to outperform scoring functions [54] and shape-based similarity ranking [55] in a virtual screening experiment conducted over a large combinatorial library of mimic compounds experimentally tested against thrombin (Fig. **4**) [56]. The analysis of the true positives signaled the relevance of the privileged molecular substructures in the case of thrombin enzyme, supposedly engaging localized interactions with the D subsite. These results strengthened the confidence in LIECE as a promising tool for reliable estimation of anti-thrombin activity with the aim to offer to the medicinal chemists involved in research on thrombin and thrombin-related disease (e.g., coagulation cascade, blood clot formation) a successful strategy for lead finding and optimization.

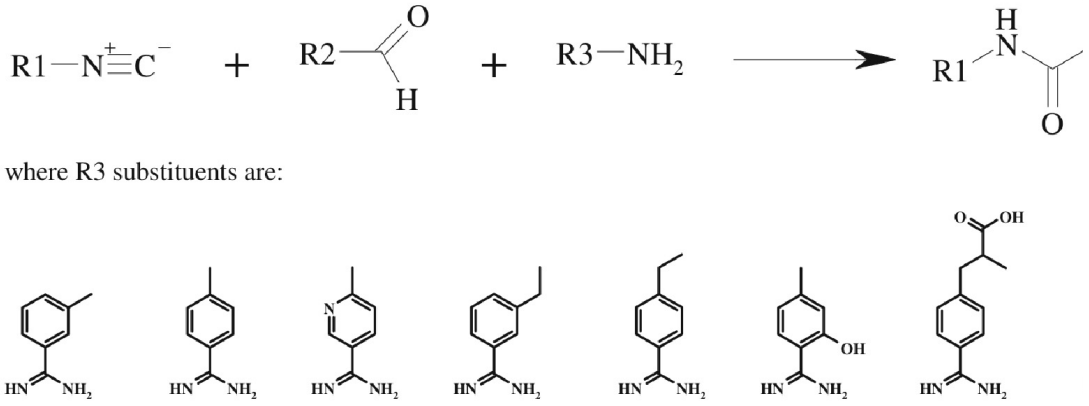
However, as already pointed out by Kollman [4], one of the most important inaccuracy in the current used force fields is the evaluation of the electrostatic contribution to the ligand binding. For this reason, further efforts were spent to capture polarization effects for a better description of the compound atomic charges. *Ab initio* and semiempirical approaches were used by Åqvist [57] who suggested that semiempirically derived CM1A charges [58] emerged as a fast and reliable alternative for fully automated LIE based virtual screening with the OPLS-AA force field [59,60]. On the other hand, Caflisch further improved LIECE by using a linearly scaling semiempirical QM method [61] associated with the CHARMM force field [62,63] to calculate the electrostatic interaction energy between the ligand and the protein, adopting the so-called divide and conquer approach [61] and using MOPAC [64] and the recently developed semi-empirical Hamiltonian RM1 [65].

The new approach was termed QMLIECE [24] and its predictive ability (Table **3**) is increasing with respect to LIECE for compounds having significantly diverse charge-

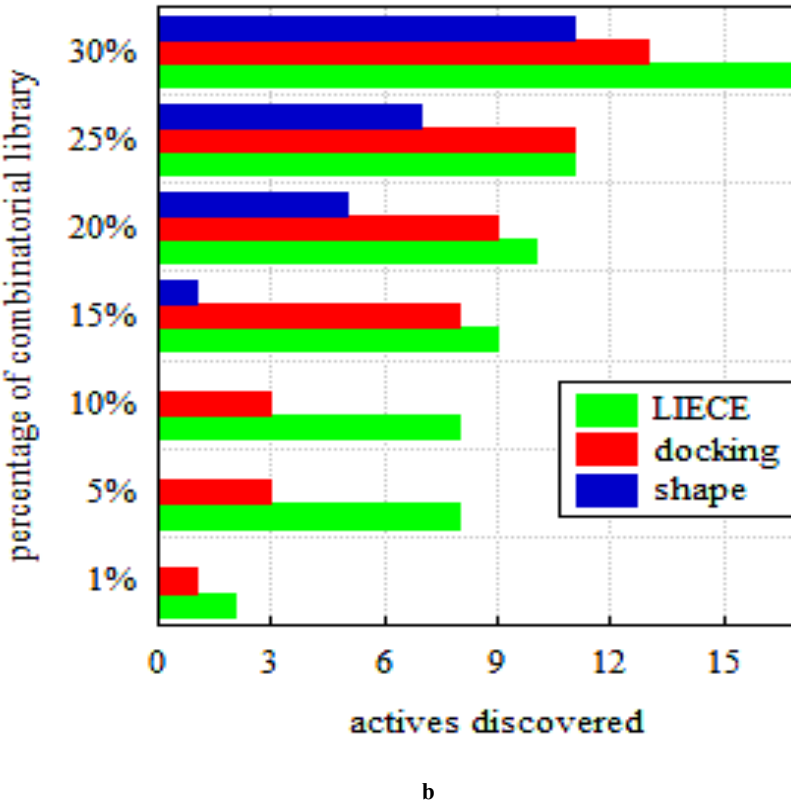
Table 3. LIECE Models[a]

model	No. objects	No. parameters	parameters	Energy RMS error (kcal/mol)	LOO q^2
BACE	13	2	ΔE_{vdW}, ΔG_{elec}	1.16	0.71
BACE	13	3	ΔE_{vdW}, ΔG_{elec}, $\Delta G_{tr,rot}$	0.95	0.65
HIV-1 PR	24	2	ΔE_{vdW}, ΔG_{elec}	0.89	0.64
HIV-1 PR	24	3	ΔE_{vdW}, ΔG_{elec}, $\Delta G_{tr,rot}$	0.73	0.77
CDK2	73	1	ΔE_{vdW}	0.98	0.80
CDK2	73	2	ΔE_{vdW}, ΔG_{elec}	0.93	0.82
CDK2	73	3	ΔE_{vdW}, ΔE_{coul}, ΔG_{solv}	0.89	0.83
Lck	51	1	ΔE_{vdW}	0.93	0.47
Lck	51	2	ΔE_{vdW}, ΔG_{elec}	0.93	0.44
Lck	51	3	ΔE_{vdW}, ΔE_{coul}, ΔG_{solv}	0.84	0.53
p38	41	1	ΔE_{vdW}	1.01	0.40
p38	41	2	ΔE_{vdW}, ΔG_{elec}	0.98	0.43
p38	41	3	ΔE_{vdW}, ΔE_{coul}, ΔG_{solv}	0.80	0.59
CDK2+Lck	124	1	ΔE_{vdW}	1.13	0.66
CDK2+Lck	124	2	ΔE_{vdW}, ΔG_{elec}	1.03	0.72
CDK2+Lck	124	3	ΔE_{vdW}, ΔE_{coul}, ΔG_{solv}	1.02	0.72
CDK2+p38	114	1	ΔE_{vdW}	0.99	0.81
CDK2+p38	114	2	ΔE_{vdW}, ΔG_{elec}	0.97	0.82
CDK2+p38	114	3	ΔE_{vdW}, ΔE_{coul}, ΔG_{solv}	0.93	0.83
Lck+p38	92	1	ΔE_{vdW}	1.19	0.39
Lck+p38	92	2	ΔE_{vdW}, ΔG_{elec}	1.00	0.56
Lck+p38	92	3	ΔE_{vdW}, ΔE_{coul}, ΔG_{solv}	0.98	0.57
CDK2+Lck+p38	165	1	ΔE_{vdW}	1.13	0.69
CDK2+Lck+p38	165	2	ΔE_{vdW}, ΔG_{elec}	1.03	0.74
CDK2+Lck+p38	165	3	ΔE_{vdW}, ΔE_{coul}, ΔG_{solv}	1.03	0.74
WNV PR (QMLIECE)	44 ($0 \leq Q \leq 3$)	2	ΔG_{QMelec}, $\Delta G_{tr,rot}$	0.67	0.65
WNV PR	44 ($0 \leq Q \leq 3$)	2	ΔG_{MMelec}, $\Delta G_{tr,rot}$	0.91	0.35
WNV PR (QMLIECE)	37 ($2 \leq Q \leq 3$)	2	ΔG_{QMelec}, $\Delta G_{tr,rot}$	0.64	0.70
WNV PR	37 ($2 \leq Q \leq 3$)	2	ΔG_{MMelec}, $\Delta G_{tr,rot}$	0.84	0.49
HIV-1 PR (QMLIECE)	24	3	$\Delta E_{vdW}\Delta G_{QMelec}$, $\Delta G_{tr,rot}$	0.64	0.80
HIV-1 PR	24	3	$\Delta E_{vdW}\Delta G_{MMelec}$, $\Delta G_{tr,rot}$	0.67	0.78
CDK2 (QMLIECE)	73	2	$\Delta E_{vdW}\Delta G_{QMelec}$	0.99	0.79
CDK2	73	2	$\Delta E_{vdW}\Delta G_{MMelec}$	0.98	0.79
fIIa	27	2	ΔE_{vdW}, ΔG_{elec}	1.03	0.662

[a]Q indicates the number of formal charges of compounds in the used training sets. Data were taken from Huang *et al.* [45], Kolb *et al.* [57], Zhou *et al.* [24], Nicolotti *et al.* [50].



a



b

Fig. (4). LIECE model on fXa performance tested over a large combinatorial mimetic library. **a**) Ugi-type three-component reaction involving 16 isonitriles, 80 aldehydes and 8 amines to generate a combinatorial library of 10240 mimetic compounds. **b**) Histogram indicating the number of actives discovered at different early percentages of the screened combinatorial library. This figure is taken from Nicolotti *et al.* [50].

charge interactions, that is a large variability of polarized charges of protein atoms upon binding different inhibitors.

CONCLUSIONS

Progress in medicinal chemistry and biochemistry is upon the understanding of the type and nature of the molecular interactions that, in turn, are dependent on the amount of free energy variation arising from the binding process involving a bioactive compound and its biological counterparts. As properly stated in 1993 by Kolmann "*The goals of any numerical theoretical approach applied to chemical phenomena are to calculate numerical values that agree with experiment, provide mechanistic insight into the phenomena, and be predictive.*" [4] On this purpose, a series of MD or MC-based computational approaches have been here above proposed to accurately predict the free energy (BFE) involved in the binding. Starting from the efficient Linear Interaction Energy (LIE) method, which gives an estimation of the BFE from only two simulations of the solvated protein-ligand complex and the ligand in solution, a

series of LIE-derived approaches has been reviewed. From the initial LIE evaluation of non-bonded van der Waals and electrostatic intermolecular interaction ensembles, the list of descriptors was expanded to the Extended Linear Response (ELR), in order to include, via multivariate fitting, a number of descriptors with desirable and interpretable physical meaning.

In this regard, the herein described methodologies benefit of a cheap computational cost because only interactions between the ligand and either the protein or the aqueous environment are explicitly considered as descriptors whereas further developments have also led to the replacement of explicit solvent by a continuum surface generalized Born (SGB) solvent model. Known as SGB-LIE, the new method was proved to be more than one order of magnitude faster than the previously described models because of a much more rapid and deeper exploration of the conformational space as the water frictions were not explicitly taken into account. The encouraging results obtained with the continuum solvent model inspired the development of the so-called molecular mechanics generalized Born surface area (MM-GB/SA) and its Poisson-Boltzmann variant (MM-PB/SA) to rescore docking results with the combined use of molecular mechanics and continuum implicit solvation to compute average BFE of ligands after MD or MC simulations.

The need of speeding up calculations as well as of reducing computational costs for rescoring docking solutions resulting from large molecular databases screening has led to another important variant of LIE named LIECE (Linear Interaction Energy in Continuum Electrostatics) and its quantum mechanics variant QMLIECE. In this approach, the MD sampling was replaced by a simple energy minimization that was profitably combined with the LIE method to ensure a rigorous treatment of continuum electrostatics based on the numerical solution of the Poisson equation by finite difference techniques.

In LIECE, as in other cited techniques, the lack of sampling (due to the replacement of MD simulation with the energy minimization) is however a negligible and secondary issue when the aim is focused on the scoring congeneric series of ligands. In such a case, that is very much frequent in QSAR practice, it was proven that a systematic error did not affect the accuracy of the ranking.

The successful application to a series of relevant pharmaceutical targets (among the others factor Xa, thrombin, HIV-1 reverse transcriptase, β-secretase, different protein kinases), as well as the simplicity of the cited computational techniques allow their use for the rapid assessment of million-compound molecular libraries as nowadays routinely required by the pharmaceutical market's time pressure.

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGEMENT

The authors thank the University of Bari "Aldo Moro" (Progetto IDEA Giovani Ricercatori 2008) for financial support.

REFERENCES

[1] Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M.H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, **2006**, *49*, 5912-5931.

[2] Kontoyianni, M.; McClennan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.*, **2004**, *47*, 558-565.

[3] Beveridge, D. L.; Di Capua F. M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Chem.*, **1989**, *18*, 431-492.

[4] Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.*, **1993**, *93*, 2395-2417.

[5] Tembe, B. L.; McCammon, J. A. In: *Ligand-Receptor interactions*; Comput. Chem., Ed.; Elsevier: Oxford, **1995**.

[6] Hansson, T.; Åqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.*, **1995**, *8*, 1137-1144.

[7] Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interaction. *J. Comput. Chem.*, **1996**, *17*, 1587-1597.

[8] Lamb, M. L.; Jorgensen, W. L. Computational approaches to molecular recognition. *Curr. Opin. Chem. Biol.*, **1997**, *1*, 1587-1597.

[9] Nicolotti, O.; Miscioscia, T. F.; Carotti, A.; Leonetti, F.; Carotti, A. An integrated approach to ligand- and structure-based drug design. Development and application to a series of serine protease inhibitors. *J. Chem. Inf. Model.*, **2008**, *48*, 1211-1226.

[10] Nicolotti, O.; Giangreco, I.; Miscioscia, T. F.; Carotti, A. Improving Quantitative Structure-Activity Relationships through Multi-Objective Optimization. *J. Chem. Inf. Model.*, **2009**, *49*, 2290-2302.

[11] Mitchell, M. J.; McCammon, J. A. Free-energy differences calculations by thermodinamic integration: difficulties in obtaining a precise value. *J. Comput. Chem.*, **1991**, *12*, 271-275.

[12] Hodel, A.; Simonson, T; Fox, R. O.; Brunger, A. T. Conformational substates and uncertainty in macromolecular free-energy calculations. *J. Phys. Chem.*, **1993**, *97*, 3409-3417.

[13] Åqvist, J.; Medina, C.; Samuelsson, J. E. New Method for predicting binding affinity in computer-aided drug design. *Protein Eng.*, **1994**, *7*, 385-391.

[14] Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding costants of neuraminidase inhibitors: an investigation of the linear interaction energy method. *J. Med. Chem.*, **1999**, *42*, 5142-5152.

[15] Pierce, A. C.; Jorgensen, W. L. Estimation of binding affinities for selective thrombin inhibitors via Monte Carlo simulation. *J. Med. Chem.,* **2001**, *44*, 1043-1050.

[16] Rizzo, R. C.; Udier-Blagović, M.; Wang, D.; Watkins, E. K.; Smith, M. B. K.; Smith, R. H. J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of activity for non-nucleoside inhibitors with HIV-1 reverse transcriptase based on Monte Carlo Simulation. *J. Med. Chem.*, **2002**, *45*, 2970-2987.

[17] Wesolowski, S. S.; Jorgensen, W. L. Estimation of binding affinities for celecoxib analogues with COX-2 via Monte Carlo extended linear response. *Bioorg. Med. Chem. Lett.*, **2002**, *12*, 267-70.

[18] Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. New linear interaction method for binding affinity calculations using a continuum solvent model. *J. Phys. Chem. B*, **2001**, *105*, 10338-10397.

[19] Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, **1990**, *112*, 6127-6129.

[20] Gosh, A.; Rapp, C. S.; Friesner, R. A.; Generalized Born model based on surface integral formulation. *J. Phys. Chem. B*, **1998**, *102*, 10983-10990.

[21] Friedrich, M.; Zhou, R.; Edinger, S.; Friesner, R. A. Poisson-Boltzmann Analytical Gradients for Molecular Modeling Calculations. *J. Phys. Chem. B*, **1999**, *103*, 3057-3061.

[22] Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative predicition of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.*, **2000**, *43*, 3786-3791.

[23] Barril, X.; Gelphì, J. L.; Lopez, J. M.; Orzoco, M.; Luque, F. J. How accurate can molecular dynamics/linear response and Poisson-Boltzman/solvent accessible surface calculations be for predicting relative binding affinities? Acetylcholinesterase as huprine inhibitors as a test case. *Theor. Chem. Acc.*, **2001**, *106*, 2-9.

[24] Zhou, T.; Huang, D.; Caflisch A. Is quantum mechanics necessary for predicting binding free energy? *J. Med. Chem.*, **2008**, *51*, 4280-4288.

[25] Roux, H. A.; Yu, B.; Karplus, M. Molecular basis for the Born model of ion solvation. *J. Phys. Chem.*, **1990**, *94*, 4683-4688.

[26] Hansson, T.; Marelius J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided. Mol. Des.*, **1998**, *12*, 27-35.

[27] Ljungberg, K. B.; Marelius, J.; Musil, D.; Svensson, P.; Norden, B.; Åqvist, J. Computational modeling of inhibitor binding to human thrombin. *Eur. J. Pharm. Sci.*, **2001**, *12*, 441-446.

[28] Jones-Hertzog, D. K.; Jorgensen, W. L. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *J. Med. Chem.*, **1997**, *40*, 1539-1550.

[29] Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the binding affinities of FKBP12 inhbitors usning a linear response method. *Bioorg. Med. Chem.*, **1999**, *7*, 851-860.

[30] Carlson, H. A.; Jorgensen, W. L. An extended linear response method for determining free energies of hydration. *J Phys. Chem.*, **1995**, *99*, 10667-10673.

[31] McDonald, N. A.; Carlson, H. A.; Jorgensen W. L. Free energies of solvation in chloroform and water from a linear response approach. *J. Phys. Org. Chem.*, **1997**, *10*, 563-576.

[32] Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.*, **1996**, *9*, 567-571.

[33] Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Åqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 protease inhibitors derived from mannitol: synthesis, inhibitory potencies, and computational predictions of binding affinities. *J. Med. Chem.*, **1997**, *40*, 885-897.

[34] Wang, W.; Wang, J.; Kollman, P. A. What determines the van der Waals coefficient beta in the LIE (Linear Interaction Energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins*, **1999**, *34*, 395-402.

[35] Smith, R. H.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Smith, M. B. K. Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in linear response methodology. *J. Med. Chem.*, **1998**, *41*, 5272-5286.

[36] Guimarães, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.*, **2008**, *48*, 958-970.

[37] Grazioso, G.; Cavalli, A.; De Amici, M.; Recanatini, M.; De Micheli, C. Alpha 7 nicotinic acetycholine receptor agonist: prediction of their binding affinity through a molecular mechanics Poisson-Boltzmann surface areao approach. *J. Comput. Chem.*, **2008**, *29*, 2593-2603.

[38] Thompson, D. C.; Humblet, C.; Joseph-McCarhty, D. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.*, **2007**, *48*, 1081-1091.

[39] Nissink, J. W. M.; Murray, C. W.; Hartshorn, M. J.; Verdonk, M. K.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein-ligand interaction. *Proteins*, **2002**, *49*, 457-471.

[40] Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.*, **2006**, *49*, 6789-6801.

[41] Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science*, **1991**, *252*, 106-109.

[42] Sitkoff, D.; Ben-Tal, N.; Honig, B. Calculation of alkane to water solvation free energies using continuum solvent models. *J. Phys. Chem.*, **1996**, *100*, 2744-2752.

[43] Bohm, H. J. The development of a simple empirical scoring function to estimate the binding costant for a protein ligand complex of known 3-dimensional structure. *J. Comput.-Aided Mol. Design*, **1994**, *8*, 243-256.

[44] Huang, D.; Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.*, **2004**, *47*, 5791-5797.

[45] Warwicker, J.; Watson, H. C. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.*, **1982**, *157*, 671-679.

[46] Huang. D.; Lüthi, U.; Kolb, P.; Edler, K.; Cecchini, M.; Audetat, S.; Barberis, A.; Caflisch, A. Discovery of cell-permeable non-peptidic inhibitors of β-secretase by high-throughput docking and continuum electrostatics calculation. *J. Med. Chem.*, **2005**, *48*, 5108-5111.

[47] Huang. D.; Lüthi, U.; Kolb, P.; Cecchini, M.; Barberis, A.; Caflisch, A. In silico discovery of β-secretase inhibitors. *J. Am. Chem. Soc.*, **2006**, *128*, 5436-5443.

[48] Kolb, P.; Kipouros, K. B.; Huang, D.; Caflisch A. Structure-based tailoring of compound libraries for highthroughput screening: discovery of novel EphB4 kinase inhibitor. *Proteins*, **2008**, *73*, 11-18.

[49] Lafleur, K.; Huang, D.; Zhou, T.; Caflisch A.; Nevado C. Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *J. Med. Chem.*, **2009**, *52*, 6433-6446.

[50] Nicolotti, O.; Giangreco, I.; Miscioscia, F.; Convertino, M.; Leonetti, F.; Pisani, L.; Carotti, A. Screening of benzamidine-based thrombin inhibitors via a linear interaction energy in continuum electrostatics model. *J. Comput.-Aided Mol. Des.*, **2010**, *24*, 117-129.

[51] Kolb, P.; Huang, D.; Dey, F.; Caflisch, A. Discovery of kinase inhibitors by high-throughput docking and scoring based on a trasferable linear interaction energy model. *J. Med. Chem.*, **2008**, *51*, 1179-1188

[52] Wichapong, K.; Lawson, M.; Pianwanit, S.; Kokpol, S.; Sippl, W. Postprocessing of protein-ligand docking poses using Linear Response MM-PB/SA: application to Wee1 kinase inhibitors. *J. Chem. Inf. Model.*, **2010**, *50*, 1574-1588.

[53] Alfarano, P. Structure-based methods for virtual screening, protein design and protein function studies. PhD Thesis, Universität Zürich: Zürich, July **2010**.

[54] Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R. L.; Taylor R. Development and validation of a generic algorithm for flexible docking. *J. Mol. Biol.*, **1997**, *267*, 727-748.

[55] ROCS, version 2.4.2, OpenEye scientific software Inc. Santa Fe, NM, USA, **2005**, http://www.eyesopen.com.

[56] Illgen, K.; Enderle, T.; Broger, C.; Weber, L.; Simulated molecular evolution in a full combinatorial library. *Chem. Biol.*, **2000**, *7*, 433-441.

[57] Wallin, G.; Nervall, M.; Carlsson, J.; Åqvist, J. Charges for large scale binding free energy calculations with the linear interaction energy method. *J. Chem. Theory Comput.*, **2009**, *5*, 380-395.

[58] Storer, J. W.; Giesen, D. J.; Cramer. C. J.; Trhular, D. G.; Class IV charge models a new empirical approach in quantum chemistry. *J. Comput.-Aided. Mol. Des.*, **1995**, *9*, 87-110.

[59] Jorgensen ,W. L.; Tirado-Rives J. The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.*, **1988**, *110*, 1657-1666.

[60] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.,* **1996**, *118*, 11225-11236.

[61] Dixon, S. L.; Merz, K. M. Semiempirical molecular orbital calculation with linear system size scaling. *J. Chem. Phys.*, **1996**, *104*, 6643-6649.

[62] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **1983**, *4*, 187-217.

[63] Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **2009**, *30*, 1545-1614.

[64] Stewart, J. J. P. Optimization of parameters for semiempirical methods. 1. *J. Comput. Chem.*, **1989**, *10*, 209-220.

[65] Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1:a reparametrization of AM1 for H, C, N, O, P, S, F, Cl, Br and I. *J. Comput. Chem.*, **2006**, *27*, 1101-1111.

Received: February 21, 2011 Revised: June 20, 2011 Accepted: June 21, 2011