Discovery of Kinase Inhibitors by High-Throughput Docking and Scoring Based on a Transferable Linear Interaction Energy Model

Peter Kolb,[†] Danzhi Huang,[†] Fabian Dey,[†] and Amedeo Caflisch*

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received June 7, 2007

The linear interaction energy method with continuum electrostatics (LIECE) is evaluated in depth on five kinases. The two multiplicative coefficients for the van der Waals energy and electrostatic free energy are shown to be transferable among different kinases. Moreover, good enrichment factors are obtained for a library of 40375 diverse compounds seeded with 73 known inhibitors of CDK2. Therefore, a general two-parameter LIECE model for kinases is derived by combining large data sets of inhibitors of CDK2, Lck, and p38. This two-parameter model is cross-validated on two kinases not used for fitting; it shows an average error of about 1.5 kcal/mol for the prediction of absolute binding affinity of 37 and 128 known inhibitors of EphB4 and EGFR, respectively. High-throughput docking and ranking by two-parameter LIECE models are shown to be able to identify novel low-micromolar EphB4 and CDK2 inhibitors of low-molecular weight ($\leq$355 g/mol).

1. Introduction

Accurate and efficient approaches for the evaluation of binding affinities are required for in silico screening of large libraries of compounds by high-throughput docking.[1–10] Rigorous methods based on free energy perturbation molecular dynamics simulations have recently been developed to improve efficiency by enhancing convergence. However, these methods still require about 10–20 days of computer time per compound.[11]

The LIE (linear interaction energy) method was proposed to calculate free energies of binding by averaging interaction energies from molecular dynamics simulations of the ligand and the ligand/protein complex.[12,13] In LIE, the free energy of binding is approximated by

$$\Delta G = \alpha(\langle E^{\mathrm{vdW}}\rangle_{\mathrm{bound}} - \langle E^{\mathrm{vdW}}\rangle_{\mathrm{free}}) + \beta(\langle E^{\mathrm{elec}}\rangle_{\mathrm{bound}} - \langle E^{\mathrm{elec}}\rangle_{\mathrm{free}})$$

where E^{vdw} and E^{elec} are the van der Waals and electrostatic interaction energies between the ligand and its environment. The environment is either the solvent (free) or the solvated ligand/protein complex (bound). The $\langle\rangle$ denotes an ensemble average sampled over a molecular dynamics[12] or Monte Carlo[14] trajectory. The coefficient α is determined empirically.[12] Originally, α was fixed to a value of $^1/_2$, as predicted by the linear response approximation.[12] Later studies have shown, however, that improved models for a large variety of systems could be obtained by considering β as a free parameter.[15] Consequently, both coefficients are obtained by a fit of experimentally determined values of ΔG to the calculated values of E^{elec} and E^{vdw} for a training set of known ligands.

The LIE method and modifications thereof have been applied to a large number of existing inhibitor/protein data sets.[12,16–22] Moreover, LIE-based scorings of ligands were shown to perform better than established scoring functions.[23] Interestingly, recent applications to pharmaceutically relevant enzyme targets have documented the predictive ability and usefulness in lead-discovery projects. As an example, the LIE method with explicit water molecular dynamics sampling was successfully used in the design of a series of inhibitors of the malarial aspartic proteases plasmepsin I and II.[24] Unfortunately, LIE cannot be used for high-throughput docking because of its computational requirements (the currently fastest implementation needs about 6 h for each compound[23]). Therefore, we have replaced the explicit water molecular dynamics (or Monte Carlo) sampling with a simple energy minimization and combined the LIE method with a rigorous treatment of solvation within the continuum electrostatics approximation,[22] i.e., the numerical solution of the Poisson equation by the finite-difference technique.[25] The LIECE[a] approach, where the last two letters stand for continuum electrostatics, is about 2 orders of magnitude faster than previous LIE methods and shows a similar precision on the targets tested. In fact, an error of about 1 kcal/mol was observed for 13 and 29 peptidic inhibitors of β-secretase and HIV-1 protease, respectively.[22] Most importantly, the LIECE approach has played a key role in the recent discoveries of three novel series of β-secretase inhibitors (phenylurea derivatives,[9] triazine derivatives,[10] and a set of five cell-permeable, nonpeptidic, low-micromolar inhibitors with a different scaffold (D. Huang and A. Caflisch, unpublished results)).

Here, we further validate the predictive ability of LIECE by a critical assessment of five different protein kinases and 330 inhibitors. Protein kinases are important pharmacological targets,[26,27] and several three-dimensional structures with inhibitors and binding affinity data are available.[28,29] Recently, a generalization of LIE to additional energy terms and parameters (i.e., the extended linear response approach) has been applied to three protein kinases;[21] the results and predictive ability are used as a basis of comparison here. The present work was motivated by three main questions: Are the parameters of the LIECE approach transferable between enzymes of the same class? Which physicochemical properties of the binding site are responsible for the parameter transferability? Is it possible to combine experimental data from different kinases to derive a predictive LIECE model with a minimal number of parameters?

* To whom correspondence should be addressed. Phone: (+41 44) 635 55 21. Fax: (+41 44) 635 68 62. E-mail: caflisch@bioc.uzh.ch.
† These authors contributed equally to this work.

[a] Abbreviations: LIECE, linear interaction energy with continuum electrostatics; CDK2, cyclin-dependent kinase 2; Lck, lymphocyte-specific kinase; p38, p38 mitogen-activated protein kinase; EGFR, epidermal growth factor receptor; EphB4, erythropoietin producing human hepatocellular carcinoma receptor B4.

10.1021/jm070654j CCC: $40.75
Published on Web 02/14/2008

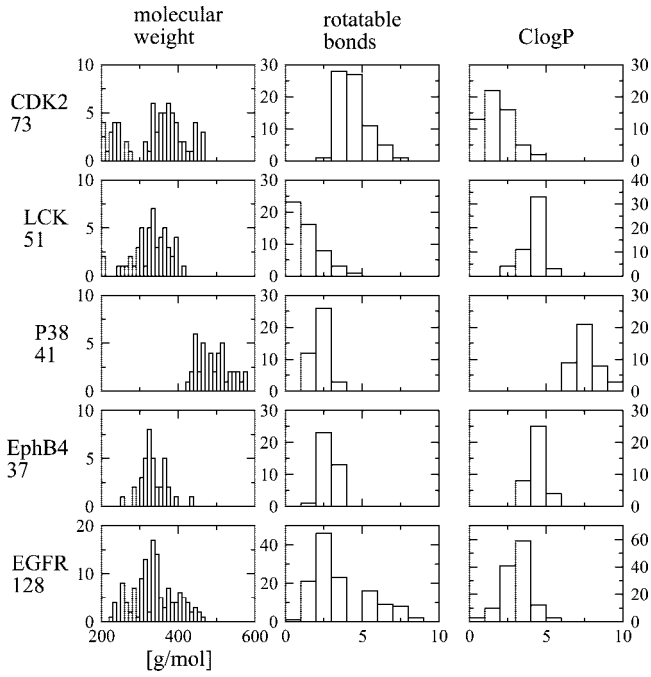


Figure 1. Distributions of some key properties of the five sets of inhibitors. All values have been calculated with DAIM.[52]

Answers to these questions help to determine the usefulness and range of applicability of the LIECE method in structure-based ligand design.

2. Methods

Selection of the Kinases. Three tyrosine kinases, lymphocyte-specific kinase (Lck), erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4), and epidermal growth factor receptor (EGFR), as well as two Ser-Thr protein kinases, cyclin-dependent kinase 2 (CDK2) and p38 mitogen-activated protein kinase (p38), were selected for the present study. CDK2, Lck, and p38 (with their inhibitors, see below) were chosen as training set to directly compare with the extended linear response approach published recently.[21] Cross-validation was performed on 37 EphB4 inhibitors (in-house data, unpublished) and 128 EGFR inhibitors;[30] these two tyrosine kinases and their inhibitors had not been employed for fitting.

Selection of the Inhibitors for the Training Set. For CDK2, 50 of the 94 inhibitors published by Bramson et al.[31] were used. Tominaga and Jorgensen[21] had used a subset of 57 compounds, which was further reduced in our study by excluding the seven inhibitors for which only lower bounds of the IC_{50} were available (compounds **C14**, **C34**, **C38–40**, **C48**, and **C49** in ref 21). To extend the range of IC_{50} values, this set was supplemented with the 23 inhibitors and IC_{50} values described in Gibson et al.[32] Hence, the training set for CDK2 comprised a total of 73 known inhibitors with two different scaffolds. In the case of Lck, the 51 inhibitors used in ref 21, which had previously been reported by Chen et al.,[33,34] were selected. Finally, 41 inhibitors of p38 identified by Stelmach et al.[35] were used (compounds **9a**, **9b**, **9j**, and **9s** in ref 35 were discarded because only a lower bound for the IC_{50} had been published). In contrast to the inhibitors used in ref 21, only inhibitors binding to the ATP binding site and forming hydrogen bonds with the hinge region were used. On the other hand, the three ligand sets differ significantly in molecular weight, number of rotatable bonds, and ClogP (Figure 1), and despite the similar binding sites and modes, it is not obvious a priori that all LIECE models can be combined into one.

Selection of the Inhibitors for the Test Set. For EphB4, 37 inhibitors originating from an imidazo[1,2-*a*]pyrazine lead identified in a yeast-based high-throughput screen[36] were selected. The inhibitory activity of all compounds had been verified in two enzymatic assays: a fluorescence-based assay (Panvera Z'Lyte Tyr2 kinase assay PV3191, Invitrogen) and a radioactivity-based assay (KinaseProfiler Assay Protocols, Upstate Ltd., Dundee, U.K.). The EGFR test set comprised 128 known inhibitors published by Aparna et al.[30]

Preparation of the Inhibitor and Protein Structures. The three sets of inhibitors used to derive the LIECE models were generated by manual modification of the scaffold present in the respective X-ray structures (PDB codes 1KE5 for CDK2, 1QPE for Lck, and 1M7Q for p38). Newly added bonds were minimized in the absence of the protein structure. This and all subsequent minimizations were carried out with CHARMM,[37] using the CHARMm22 force field (Accelrys Inc.[38]). Hydrogen atoms were added to all protein structures according to the protonation states of side chains and termini at pH 7. Partial charges were then assigned to both the inhibitor and protein structures using the MPEOE method.[39,40] Protein hydrogen atoms were minimized in the absence of the inhibitor.

The 3D structures of the inhibitors of the two test sets, EphB4 and EGFR, were obtained from Dr. S. Audétat (Oncalis AG) and Dr. V. Aparna, respectively, and were visually checked for accuracy. For the calculations with EphB4 a modeled protein structure was used; the building of this model is described in the next section. The calculations for EGFR were carried out on the X-ray structure of the complex with erlotinib (PDB code 1M17). The 37 inhibitors of EphB4 and the 128 inhibitors of EGFR were manually docked into the binding site such that at least one hydrogen bond with the hinge region was formed. Each pose was minimized within the rigid protein using CHARMM. For the minimizations of the 128 compounds in EGFR, the water molecule WAT10 was retained. This water molecule is supposed to mediate a hydrogen bond between N(3) of the pyrimidine ring of the inhibitor scaffold and Thr766. It is described as essential for the successful reproduction of the crystallographic binding mode,[30] which is supported by the fact that its inclusion in the calculations decreased the average error for the prediction of this set of inhibitors substantially.

Preparation of EphB4. Since the structure of the kinase domain of EphB4 is not publicly available, a homology model was built using the structure of EphB2 (mouse, PDB entry 1JPA) as template. The overall amino acid sequence identity between the human EphB4 sequence obtained from SWISS-PROT (accession code P54760) and the sequence derived from the mouse EphB2 structure is 88.4%, and there are no gaps or insertions in the aligned region. An additionally generated binary sequence alignment with the sequence derived from the human Eph kinase structure EphA2 (PDB entry 1MQB) revealed a lower sequence identity of 63.1% and also three short regions containing gaps or insertions. Therefore, only the sequence alignment between EphB4 (human) and EphB2 (mouse) was used in the initial phase of the homology modeling procedure. All sequence alignments were performed using the program ClustalW.[41] The 1JPA crystal structure comprises two chemically identical subunits in the crystallographic asymmetric unit. A structural superpostion of the two subunits results in an average root-mean-square deviation of 0.32 Å for 269 C_α atoms. Since subunit A of the 1JPA crystal structure has better main-chain dihedral angles (φ and ψ) and lower *B* values, this subunit was initially chosen as the template structure. However, the two subunits of the 1JPA structure show some clear structural differences around the active site region. The differing amino acid residues were analyzed with respect to atomic *B* values, possible contacts, and stereochemical criteria. Subsequently, some side chain rotamer conformations of the initial template structure (subunit A) were replaced by the corresponding side chain conformations of subunit B of the 1JPA crystal structure. The resulting modified structure was then used as template for homology modeling, and a total of 100 different models were generated using the program Modeller.[42,43]

The obtained initial models were ranked and analyzed on the basis of energetical and stereochemical criteria. The best model was manually improved by comparison with the active sites of the two known Eph kinase structures (PDB entry codes 1JPA and

1MQB) and other related kinase structures (PDB entry codes 1BYG, 1FMK, 1FPU, 1IEP, 1M14, 1M17, 1M52, 1MP8, 1OPK, 1OPL, and 2SRC) found by PSI-BLAST[44] and DALI[45] searches of protein structure databases. The conformational information contained in these homologous structures was used for manually adjusting side chain conformations of conserved or similar amino acid residues. For some of the mutated residues, statistically preferred χ-angles were chosen and favorable hydrophobic or polar contacts were also considered. Additionally, an ATP molecule was modeled into the active site to avoid structural changes during minimization. Hydrogen atoms (considering appropriate ionization states for acidic and basic amino acid residues), CHARMm22[38] atom types, and partial charges were assigned to the protein and ATP using the program WITNOTP and the MPEOE method.[39,40] Manual rebuilding and refinement with CHARMM[37] using the CHARMm22[38] force field led to a model exhibiting excellent stereochemical quality, with 92.9% of the φ/ψ values in the most favored regions and 6.3% in additionally allowed regions of the Ramachandran plot as evaluated with the program PROCHECK.[46] Finally, the ATP molecule was removed.

Energy Minimization of the Complexes. All inhibitor/protein complexes were minimized by the conjugate gradient algorithm to a root mean square of the energy gradient of 0.01 kcal mol^{-1} Å^{-1}. During minimization, the electrostatic energy term was screened by a distance-dependent dielectric of $4r$ to prevent artificial deviations due to vacuum effects, and the default nonbonding cutoff of 14 Å was used. Furthermore, the positions of all protein atoms were fixed. The minimized structures were used for the evaluation of the van der Waals energy and the finite-difference Poisson calculations.

As an illustrative example, the binding mode of the manually placed and minimized compound **14f** (shown in the Supporting Information as Figure 1) is very similar to the X-ray structure of compound **14e** in p38 (PDB code 1M7Q, compound names according to ref 35). As described above, compound **14f** has been obtained by manual deletion/addition of atoms to the scaffold of the compound in the pose present in the X-ray structure (**14e**) and subsequent minimization with CHARMM[37] and the CHARMm22 force field (Accelrys Inc.[38]). Similar small deviations were observed for most compounds.

Binding Energy Evaluation. The van der Waals interaction energy and electrostatic interaction free energy were calculated by subtracting the values of the isolated components from the energy of the complex. The van der Waals energy was calculated with CHARMM and the CHARMm22 force field using the default nonbonding cutoff of 14 Å.

The electrostatic free energy is the sum of the Coulombic energy in vacuo and the solvation energy. The former was calculated with CHARMM using infinite cutoff and neglecting interactions between pairs of atoms separated by one or two covalent bonds. The electrostatic solvation energy was calculated by the finite-difference Poisson approach[25] using the PBEQ module[47] in CHARMM and a focusing procedure with a final grid spacing of 0.4 Å. Test calculations with a grid spacing of 0.3 Å yielded similar solvation energy values (results not shown). The size of the initial grid was determined by considering a layer of at least 20 Å around the solute. The dielectric discontinuity surface was delimited by the molecular surface spanned by the surface of a rolling probe of 1.4 Å. The ionic strength was set to zero and the temperature to 300 K. Two finite-difference Poisson calculations were performed for each of the three systems (inhibitor, protein, and inhibitor/protein complex). The exterior dielectric constant was set to 78.5 and 1.0 for the first and second calculation, respectively, while the solute dielectric constant was set to 1.0, which is consistent with the value used for the parametrization of the charges. The solvation energy is the difference between the two calculations.

Experimental Methods. Panvera. In vitro kinase activity was measured using the Panvera Z'Lyte Tyr2 kinase assay PV3191 (Invitrogen) according to the manufacturer's instructions. Briefly, five dilutions of compound in a 3-fold series were measured, with the highest concentration being 125 μM. The reaction assay (10 μL) contained 7.5 ng of EphB4 kinase (Proqinase, Germany), 10 μM ATP, and 1% DMSO. The reaction was performed at room temperature for 1 h. Since this assay contains Brij, a nonionic detergent, aggregating compounds (i.e., promiscuous binders) should not show inhibition.[48]

Cerep. The assays performed at Cerep (Celle l'Evescault, France) were done in duplicate at eight different concentrations ranging from 10 nM to 20 μM. The concentrations of ATP were 0.75 and 0.8 μM in the assays for EphB4 and CDK2, respectively. The concentrations of EphB4 and CDK2 were 0.2 and 1.25 μg/mL, respectively.

Biosource. To provide more evidence against unspecific binding,[49] compound **1** was tested in the Omnia Tyr Recombinant Kit KNZ4051 (Biosource) twice, i.e., without detergent and with 0.01% Triton X-100. In each of the two experiments, the compound was measured in duplicate at eight different concentrations between 50 nM and 100 μM. EphB4 kinase (Proqinase, Germany) and ATP were used at final concentrations of 25 ng/μL and 125 μM, respectively. The assay was run at 303 K for 1 h.

3. Results and Discussion

LIECE Models. The equations used for fitting the calculated energy terms to the experimental free energies of binding ($\Delta G = RT \ln(IC_{50})$ are a one-parameter model

$$\Delta G = \alpha \Delta E_{vdW} \quad (1)$$

a two-parameter model with continuum electrostatics[22]

$$\Delta G = \alpha \Delta E_{vdW} + \beta \Delta G_{elec} \quad (2)$$

and a three-parameter model with decomposed electrostatics

$$\Delta G = \alpha \Delta E_{vdW} + \beta_1 \Delta E_{coul} + \beta_2 \Delta G_{solv} \quad (3)$$

where, as detailed above, ΔE_{vdw} is the intermolecular van der Waals energy and ΔG_{elec} is the sum of two terms: the intermolecular Coulombic energy in vacuo (ΔE_{coul}) plus the change in solvation energy of inhibitor and protein upon binding (ΔG_{solv}). Additional models were tested by taking into account the loss of translational and rotational degrees of freedom upon binding and the freezing of rotatable bonds of the inhibitor. No improvement was observed for the kinase inhibitors in this study. This result is consistent with the LIECE models of β-secretase but is in contrast to the results for 24 peptidic inhibitors of HIV-1 protease where a third parameter reflecting the penalty for the loss of translational and rotational entropy improved both fitting and predictive ability.[22]

Predictive Accuracy. The parameters obtained by least-squares fitting are given in Tables 1−3 for the LIECE models on individual kinases and combined data sets. The single-protein LIECE models derived from the 73 inhibitors of CDK2 show a higher level of predictivity than those derived from Lck or p38 (Table 1) probably because of the large range of activity values (Figure 2A). Notably, the derivation of predictive models using more than one kinase is possible despite the differing properties of the three ligand sets (Figure 1). As an example, the two-parameter CDK2-Lck-p38-model (eq 2) shows a root mean square of the error (rmse) of 1.29 and 1.46 kcal/mol in predicting the binding affinity of the 37 inhibitors of EphB4 and the 128 inhibitors of EGFR, respectively (Table 3 and Figure 2). Most of the 37 inhibitors of EphB4 are predicted to bind more favorably, which might in part be a consequence of the use of a homology model. The cross-validation on EphB4 and EGFR yields rmse < 1.5 kcal/mol for most LIECE models (Table 4). Such robustness of the LIECE models is especially encouraging in the context of high-throughput screening by docking. In fact, the EphB4 LIECE model generated using only

Table 1. LIECE Parameters for the Single-Protein Models[a]

	α	β or β_1	β_2	rms[b] (kcal/mol)	LOO[c] cv q^2
CDK2 (73 Inhibitors, 1KE5)					
$\alpha\Delta E_{vdW}$	0.2388			0.98	0.80
standard deviation	±0.0030				
$\alpha\Delta E_{vdW} + \beta\Delta G_{elec}$	0.2866	0.0520		0.93	0.82
standard deviation	±0.0171	±0.0183			
$\alpha\Delta E_{vdW} + \beta_1\Delta E_{coul} + \beta_2\Delta G_{solv}$	0.2395	0.0750	*0.0294*	0.89	0.83
standard deviation	±0.0265	±0.0208	*±0.0207*		
Lck (51 Inhibitors, 1QPE)					
$\alpha\Delta E_{vdW}$	0.2700			0.93	0.47
standard deviation	±0.0037				
$\alpha\Delta E_{vdW} + \beta\Delta G_{elec}$	0.2735	*0.0046*		0.93	0.44
standard deviation	±0.0182	*±0.0237*			
$\alpha\Delta E_{vdW} + \beta_1\Delta E_{coul} + \beta_2\Delta G_{solv}$	0.2446	0.1528	*0.0076*	0.84	0.53
standard deviation	±0.0208	±0.0572	*±0.0237*		
p38 (41 Inhibitors, 1M7Q)					
$\alpha\Delta E_{vdW}$	0.2377			1.01	0.40
standard deviation	±0.0032				
$\alpha\Delta E_{vdW} + \beta\Delta G_{elec}$	0.2699	0.0264		0.98	0.43
standard deviation	±0.0210	± 0.0170			
$\alpha\Delta E_{vdW} + \beta_1\Delta E_{coul} + \beta_2\Delta G_{solv}$	0.1827	0.1584	*−0.0013*	0.80	0.59
standard deviation	±0.0316	±0.0397	*±0.0186*		

[a] The LIECE parameters for the single-protein models. Parameters with LOO variation of the same order of magnitude as the parameter itself are statistically not significant and are given in italics. [b] Root mean square of the error when predicting the ΔG values. [c] Leave-one-out cross-validated q^2.

Table 2. LIECE Parameters for the Two-Protein Models[a]

	α	β or β_1	β_2	rms[b] (kcal/mol)	LOO[c] cv q^2	rms[b] (kcal/mol)	R^d
CDK2 + Lck (124 Inhibitors)							
						prediction of 41 inhibitors of p38	
$\alpha\Delta E_{vdW}$	0.2510			1.13	0.66	1.20	0.66
standard deviation	±0.0023						
$\alpha\Delta E_{vdW} + \beta\Delta G_{elec}$	0.3072	0.0657		1.03	0.72	1.17	0.65
standard deviation	±0.0113	±0.0129					
$\alpha\Delta E_{vdW} + \beta_1\Delta E_{coul} + \beta_2\Delta G_{solv}$	0.3118	0.0440	0.0620	1.02	0.72	1.13	0.64
standard deviation	±0.0116	±0.0180	±0.0131				
CDK2 + p38 (114 Inhibitors)							
						prediction of 51 inhibitors of Lck	
$\alpha\Delta E_{vdW}$	0.2383			0.99	0.81	1.52	0.74
standard deviation	±0.0022						
$\alpha\Delta E_{vdW} + \beta\Delta G_{elec}$	0.2632	0.0235		0.97	0.82	1.31	0.72
standard deviation	±0.0103	±0.0095					
$\alpha\Delta E_{vdW} + \beta_1\Delta E_{coul} + \beta_2\Delta G_{solv}$	0.2190	0.0439	*0.0006*	0.93	0.83	1.83	0.76
standard deviation	±0.0193	±0.0121	*±0.0127*				
Lck + p38 (92 Inhibitors)							
						prediction of 73 inhibitors of CDK2	
$\alpha\Delta E_{vdW}$	0.2513			1.19	0.39	1.10	0.90
standard deviation	±0.0024						
$\alpha\Delta E_{vdW} + \beta\Delta G_{elec}$	0.3033	0.0508		1.00	0.56	1.16	0.91
standard deviation	±0.0089	±0.0083					
$\alpha\Delta E_{vdW} + \beta_1\Delta E_{coul} + \beta_2\Delta G_{solv}$	0.2939	0.1186	0.0584	0.98	0.57	1.69	0.92
standard deviation	±0.0098	±0.0311	±0.0090				

[a] The LIECE parameters for the two-protein models as well as the data for the predictions of the third protein not used for the derivation of the parameters. Parameters with LOO variation of the same order of magnitude as the parameter itself are statistically not significant and are given in italics. [b] Root mean square of the error when predicting the ΔG values. [c] Leave-one-out cross-validated q^2. [d] Correlation coefficient.

its 37 known inhibitors is not predictive (data not shown) because of the small range of IC_{50} values (corresponding to binding free energies in the range from −9.4 to −6.3 kcal/mol; Berset et al., Oncalis AG, unpublished results). These results indicate that it is possible to derive predictive LIECE models even in cases where only inhibitors of "cognate" kinases are known.

Interestingly, the three-parameter CDK2-Lck-p38-model does not show an improvement with respect to the two-parameter CDK2-Lck-p38-model in terms of predictivity (i.e., leave-one-out cross-validated q^2, Table 3). The β_1 and β_2 parameters are very close, indicating that the electrostatic term should not be decomposed into Coulombic and solvation terms.

Importance of van der Waals. It is noteworthy that the van der Waals parameter α varies only in a relatively small range of values, i.e., from 0.219 to 0.312, for 20 of the 21 LIECE models in Tables 1−3. The value of α is also much larger than the electrostatic parameter β because of the large correlation between van der Waals energy and experimentally measured binding affinity. As an example, linear regression of these two quantities yields a correlation coefficient of 0.902 for the 73 inhibitors of CDK2. On the same set of inhibitors, there is an

Table 3. LIECE Parameters for the Three-Protein Models[a]

	α	β or β_1	β_2	rms[b] (kcal/mol)	LOO[c] cv q^2	rms[b] (kcal/mol)	R^d	rms[b] (kcal/mol)	R^d
				CDK2 + Lck + p38 (165 Inhibitors)					
						prediction of 128 inhibitors of EGFR		prediction of 37 inhibitors of EphB4	
$\alpha\Delta E_{vdW}$	0.2463			1.13	0.69	1.58	0.53	0.92	0.16
standard deviation	±0.0019								
$\alpha\Delta E_{vdW} + \beta\Delta G_{elec}$	0.2898	0.0442		1.03	0.74	1.46	0.52	1.29	0.15
standard deviation	±0.0075	±0.0074							
$\alpha\Delta E_{vdW} + \beta_1\Delta E_{coul} + \beta_2\Delta G_{solv}$	0.2961	0.0325	0.0454	1.03	0.74	1.46	0.53	1.36	0.16
standard deviation	±0.0089	±0.0115	±0.0075						

[a] The LIECE parameters for the three-protein models as well as the data for the predictions of the two test cases. [b] Root mean square of the error when predicting the ΔG values. [c] Leave-one-out cross-validated q^2. [d] Correlation coefficient.

anticorrelation (correlation coefficient of −0.414) between the electrostatic free energy (ΔG_{elec}) and binding affinity. These results for individual energy terms are consistent with $\alpha = 0.287$ and $\beta = 0.052$ in the LIECE two-parameter model of the 73 inhibitors of CDK2. As a basis of comparison, correlation coefficients of 0.829 (ΔE_{vdw} vs measured binding affinity) and −0.189 (ΔG_{elec} vs measured binding affinity) were obtained for the 13 peptidic inhibitors of β-secretase whose two-parameter LIECE model has $\alpha = 0.274$ and $\beta = 0.180$.[22]

Moreover, the van der Waals energy alone can be used for both fitting and prediction of the test sets (Tables 1−4). In fact, the van der Waals CDK2-Lck-p38-model shows an even better predictive ability for the 37 inhibitors of EphB4 (rmse of 0.92 kcal/mol) than both the two- and three-parameter models on the same set (rmse of 1.29 and 1.36 kcal/mol, respectively). Furthermore, only a marginally worse accuracy is observed for the 128 inhibitors of EGFR (rmse of 1.58 kcal/mol for the van der Waals model vs 1.46 kcal/mol for both the two- and three-parameter models). Such predictive ability indicates that the van der Waals energy is the main factor needed to determine the relative binding strength of the known kinase inhibitors used in this study. In addition, it is likely that a very simple energy function is more predictive because of inaccuracies in the binding site structure, especially when a model of the protein is used as in the case of EphB4. On the basis of docking results obtained with a very simple energy function, it has recently been suggested that steric complementarity is a major factor in specific binding to different kinases.[50]

However, the relative importance of van der Waals and electrostatics might be significantly different for ranking only known inhibitors as opposed to a large database of (mainly) inactive compounds. In fact, the significantly higher enrichment factors (see below and Supporting Information) obtained with the two- and three-parameter models indicate that electrostatics are necessary to reduce false positives in high-throughput docking.

Parameter Transferability. The two-parameter single-protein models are similar (Table 1), and consequently transferable, because of the predominance of the van der Waals term and the steric similarity of the ATP-binding site in the three kinases. The latter observation is supported by the evaluation of the van der Waals interaction energy between the protein and a probe placed at vertices of a grid spanning the entire binding site. The probe consisted of a carbonyl group with a C=O bond length of 1.22 Å. First, the three kinase structures were overlapped (using the Cα atoms), and then the ATP-binding site was filled by a cubic grid of 1 Å spacing. The geometrical center of the probe was put on each grid point in turn, and the carbonyl was oriented along the positive direction of each of the three grid axes. Grid points for which the van der Waals interaction energy between the probe (in any of its three orientations) and the protein had a positive value were discarded. This procedure yielded 81 grid points that define the empty volume in common to the three kinases. On these grid points, the van der Waals interaction energy between the carbonyl probe and the protein is similar for the three kinases (Figure 3). Similar results were obtained for the van der Waals interaction energy of a probe consisting of a single neon atom. These findings explain the parameter transferability if one considers that inhibitors that are energy-minimized in the ATP-binding site do not clash with the protein and therefore occupy a spatial region corresponding to a subset of the 81 grid points. Most importantly, this approach can also be used a priori to judge the similarity of binding sites and hence the transferability of parameters.

False Positives and Enrichment Factors. In the previous sections, the ability of LIECE models for predicting absolute binding free energy values has been tested on known inhibitors. These tests are important but yield only limited information on the usefulness of LIECE for its main application, i.e., ranking of poses of compounds from large libraries. In fact, most of these compounds do not bind at all to the target protein or they bind only weakly and nonspecifically. Therefore, it is essential to estimate the amount of false positives, i.e., compounds whose binding affinity is significantly overestimated by the LIECE approach. A non-negligible amount of false positives is expected to emerge, since the van der Waals interaction energy is a pairwise energy function and thus depends largely on the number of atoms. Moreover, because all docked molecules are subjected to energy minimization, the van der Waals energy is almost always favorable. Consequently, a ranking according to the two-parameter LIECE model will result in large molecules with a high number of atoms in the top ranks, and this ranking will not reflect the binding affinity. An unfeasible pose can only be penalized by the electrostatic term. However, this term has a very small coefficient in the LIECE models presented above. To decrease the number of false positives, poses with unlikely binding modes have to be eliminated before calculation of the LIECE energy (Table 5). One way to discard those poses is by means of a cutoff in the "van der Waals efficiency" (the ratio of the van der Waals interaction energy to the molecular weight). This criterion eliminates compounds with high molecular weight, which have a very favorable van der Waals interaction mainly because of their large number of atoms and not because of ideal steric complementarity with the binding site.[50]

The use of a cutoff in the "Coulombic efficiency" (the ratio of the Coulombic interaction energy [evaluated using a distance-dependent dielectric of $4r$ and a nonbonding cutoff of 14 Å] to

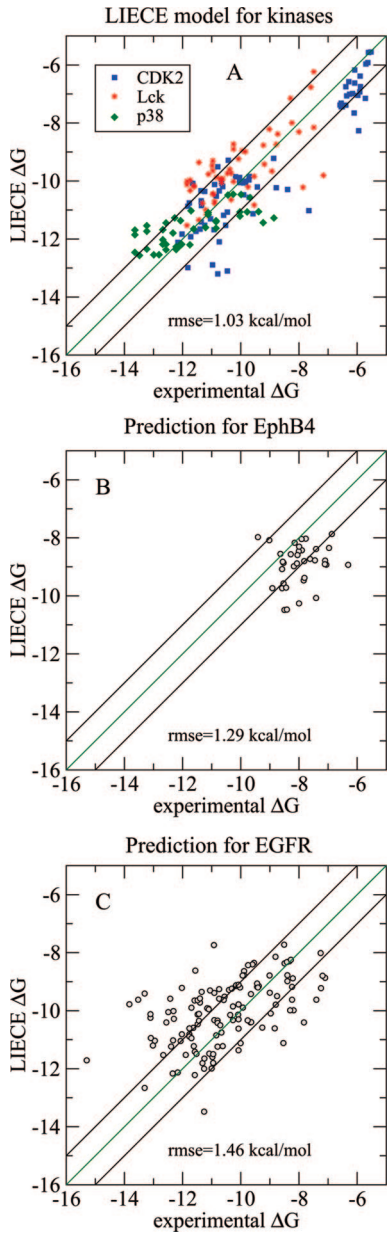


Figure 2. LIECE two-parameter CDK2-Lck-p38-model, with experimental $\Delta G = RT \ln(IC_{50})$. (A) Data used for fitting are 73, 51, and 41 inhibitors of CDK2, Lck, and p38, respectively. Cross-validation was done on 37 inhibitors of EphB4 (B) and 128 inhibitors of EGFR (C). The green diagonal is the ideal line of perfect prediction. The black diagonals delimit the 1 kcal/mol error region. All values are in kcal/mol.

the molecular weight) leads to the elimination of compounds with poor charge complementarity which cannot be discarded by the kinase LIECE models due to the small value of β. Another filter is derived from the fact that inhibitors often show only one binding mode. LIECE models are in general calculated using known inhibitors, and their energy contributions are evaluated using their unique binding mode. In contrast, for the majority of compounds from large libraries, multiple poses are generated by docking programs. To limit the number of possible poses, it is helpful to take advantage of the knowledge from crystal structures. For the ATP-binding site of kinases, the existence of at least one hydrogen bond with the hinge region can be used as a filter, which reduces the average number (±standard deviation) of poses per compound from 11.7 ± 14.0 to 3.3 ± 3.6 (Table 5).

Table 4. Root Mean Square Errors of the Predictions of the Test Sets for the Different LIECE Models

model	rms EGFR (kcal/mol)	rms EphB4 (kcal/mol)
$\alpha\Delta E_{vdW}$		
CDK2	1.76	0.83
Lck	1.37	1.50
p38	1.78	0.82
CDK2 + Lck	1.50	1.01
CDK2 + p38	1.77	0.82
Lck +p38	1.49	1.02
CDK2 + Lck + p38	1.58	0.92
$\alpha\Delta E_{vdW} + \beta\Delta G_{elec}$		
CDK2	1.63	1.10
Lck	1.37	1.52
p38	1.54	1.08
CDK2 + Lck	1.48	1.40
CDK2 + p38	1.63	0.97
Lck + p38	1.40	1.55
CDK2 + Lck + p38	1.46	1.29
$\alpha\Delta E_{vdW} + \beta_1\Delta E_{coul} + \beta_2\Delta G_{solv}$		
CDK2	1.89	0.84
Lck	1.98	1.70
p38	1.73	1.01
CDK2 + Lck	1.47	1.48
CDK2 + p38	1.84	0.82
Lck + p38	1.46	1.58
CDK2 + Lck + p38	1.46	1.36

The usefulness of these filters is illustrated by high-throughput docking into CDK2 using the 1KE5 structure.[31] A diverse set of 40 375 compounds has been selected from the ZINC database[51] on the basis of mutual dissimilarity calculated by the program DAIM.[52] By use of a fragment-based high-throughput docking approach,[10,52–54] a total of 690 530 poses were generated, with an average of 17.1 ± 19.7 poses per compound. Since it is computationally expensive to evaluate the LIECE energy for all poses, the electrostatic solvation free energy was calculated only for the 171 898 poses (of 14 701 unique compounds) with a van der Waals interaction energy more favorable than −35 kcal/mol and a van der Waals efficiency more favorable than −0.1 kcal/g. These cutoffs were determined by choosing values close to the peaks of histograms of the respective properties of the diverse set of 40 735 compounds (Supporting Information, Figure S2). Note that the 23 inhibitors of Gibson et al.[32] do not pass this filter because of their poor van der Waals interaction energy. This observation is consistent with their relatively poor binding affinity, which ranges from −6.6 to −5.5 kcal/mol, while the binding-affinity range of the compounds of Bramson et al. is −12.2 to −7.7 kcal/mol.[31]

Upon application of the aforementioned filters, the LIECE two-parameter CDK2-model ranks most of the known inhibitors before most of the docked compounds as shown by the receiver-operating characteristic (ROC[55]) plots (Figure 4). In other words, the combined use of filters and LIECE model generates few false positives. It is important to note that by keeping the poses with a very favorable van der Waals term there is an enrichment of compounds with favorable ΔG evaluated according to the van der Waals dominated LIECE model. Consequently, the ROC plots shown in Figure 4 underestimate the enrichments that can be achieved by using a combined filter/LIECE ranking. In light of this, it is worth noting that by use of the filters mentioned above, half of the 32 known inhibitors (which passed the filters) are ranked among the first 300 compounds of an initial library of 40 375 diverse compounds. Furthermore, upon ranking with the LIECE two-parameter CDK2-model, the 5% enrichment factors range from 4.4 to 11.2 depending on the filter (Table 5,

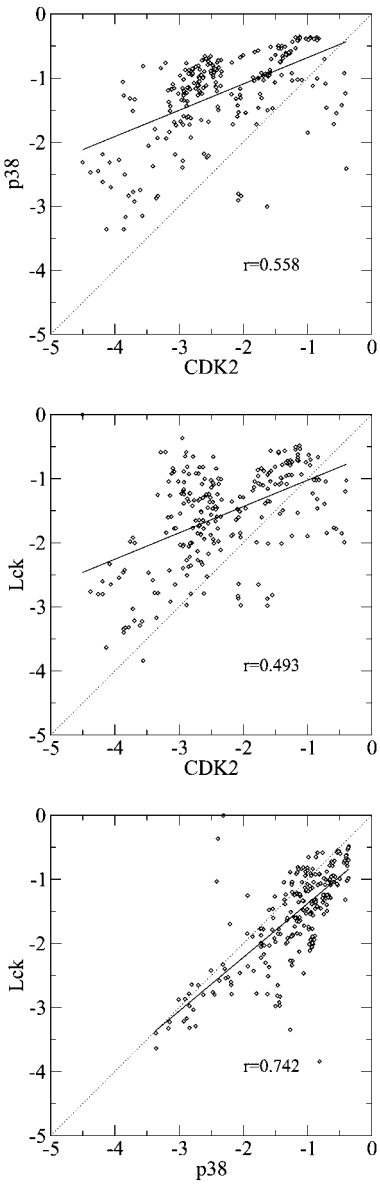


Figure 3. Scatter plot of the van der Waals interaction energy between a C═O group and a kinase at 81 grid points in the ATP-binding site. Regression lines are solid black, and the values of *r* are the correlation coefficients. All values are in kcal/mol.

last column). Interestingly, the enrichment factors are quite insensitive to the choice of the cutoff value as shown in the Supporting Information (Figure S3). In particular, the enrichment is almost constant for a van der Waals efficiency cutoff between −0.12 and −0.10 kcal/g or a Coulombic efficiency cutoff between −0.06 and −0.01 kcal/g.

The ranking obtained by the LIECE one-parameter CDK2-model yielded significantly worse enrichment factors than the two-parameter model (Supporting Information). This result indicates that while van der Waals energy alone is sufficient for self-prediction, it is not possible to neglect electrostatic interactions for ranking large libraries of compounds upon high-throughput docking. For the kinase used in this study, the most accurate filter in terms of final enrichment is the orientational filter, which requires at least one hydrogen bond with the hinge region. Importantly, the effect of each of the filters is specific because they tend to eliminate all poses of few compounds rather than few poses of many compounds (Table 5).

Table 5. Enrichment Factors (EF)[a]

filter	N_{inh}[b]	N_{poses}[c]	N_{cpds}[d]	N_{random}[e]	$N_{inh5\%}$[f]	EF[g]
E_{coul}/MW ≤ −0.01 kcal/g	48	62141	7793	11951.8 ± 3.6	26	7.1
H-bond to hinge region	48	22046	6716	8457.8 ± 4.3	41	11.2
E_{vdW}/MW ≤ −0.11 kcal/g	32	121880	11842	13859.8 ± 2.4	16	4.4
all filters	32	7895	2654	4914.0 ± 3.1	32	8.8

[a] Enrichment factors. The initial set of active molecules consists of 73 known inhibitors of CDK2 with a unique binding mode. The initial database of inactive molecules consists of 40 375 ZINC compounds. Therefore, the composite library consists of 40 448 molecules. The LIECE two-parameter CDK2 model was evaluated for the 171 898 poses, of 14 701 ZINC compounds, with $E_{vdw} \leq -35$ kcal/mol and $E_{vdw}/MW \leq -0.1$ kcal/g. Only the pose with the most favorable LIECE energy was taken into account for the ranking used to calculate the values in the last two columns. [b] Number of known inhibitors that pass the filter. [c] Number of poses of docked compounds that pass the filter. [d] Number of unique docked compounds that pass the filter. [e] Number of compounds ± standard error when randomly picking N_{poses} poses of the 14 701 compounds (100 separate random pickings). Note that each cutoff filters out a significant amount of compounds and not just a few poses of every compound, as can be deduced from the fact that N_{random} is significantly larger than N_{cpds}. [f] Number of known inhibitors among the first 5% of the LIECE ranking. [g] Enrichment factors calculated as $(N_{inh5\%}/N_{cpds,5\%})(73/40448)^{-1}$, where $N_{cpds,5\%} = 2022$ (5% of 40 448).

Experimental Validation. In the search for EphB4 inhibitors, a two-step procedure has recently been adopted to rank compounds upon fragment-based high-throughput docking. The van der Waals efficiency and presence of at least one hydrogen bond with the hinge region were first used to filter out unlikely poses. Then the poses that passed these filters were ranked according to the two-parameter CDK2-Lck-p38-model (Table 3). In this way, two inhibitors of EphB4 with an IC_{50} below 10 μM in two different enzymatic assays and a compound with an IC_{50} of 76 μM (tested in one assay) could be identified (Table 6). Notably, only 43 compounds were tested in enzymatic assays. To exclude unspecific inhibition effects due to aggregation, compound **1** was also tested in an assay with and without nonionic detergent. Encouragingly, very similar values of IC_{50} were obtained under both conditions in the Omnia Tyr recombinant kit KNZ4051 (Biosource). Such behavior is generally interpreted as strong evidence for specific inhibition, i.e., for a nonpromiscuous and nonaggregating compound.[49] The same result can also be expected for compound **2**, since the two compounds share the same scaffold and have 24 of 26 heavy atoms in common. It is important to underline that the "general" LIECE model can successfully be used to identify inhibitors of kinases that were not used to derive it.

To further validate the LIECE model for scoring, it was recently used in a docking campaign to identify CDK2 inhibitors. Again, filters were applied to the van der Waals efficiency and the presence of at least one key hydrogen bond. Thereafter, the two-parameter CDK2-model was used to rank the remaining poses. In this docking campaign only 30 compounds were tested in enzymatic assays and compound **4** emerged with an IC_{50} below 10 μM. Its predicted binding mode with three hydrogen bonds to the hinge region is depicted in Figure 5.

Computational Requirements. The LIECE approach requires about 5 min (mainly for the finite-difference Poisson calculations) of CPU time on a single Athlon 2.1 GHz for each kinase inhibitor. It is about 2 orders of magnitude faster than the most efficient LIE method reported (about 6 h for each inhibitor[23]). For a protein of 250 residues, the memory requirement for the finite-difference Poisson calculations is about 0.3 GB.

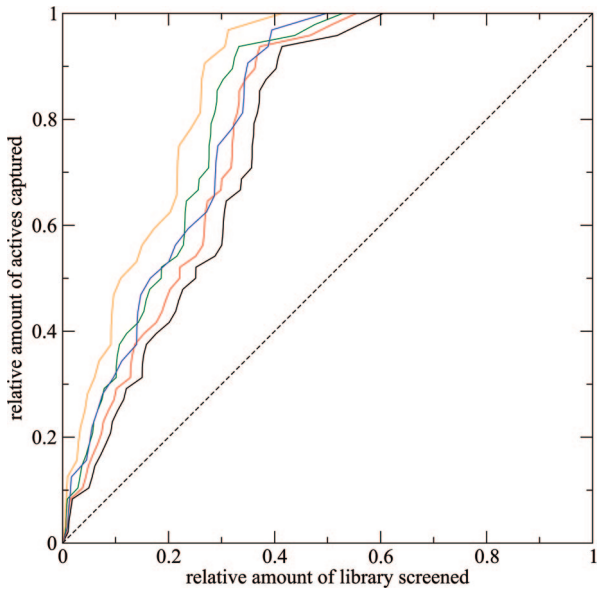


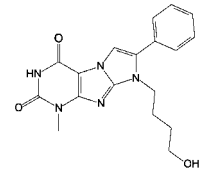
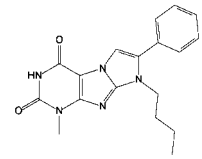
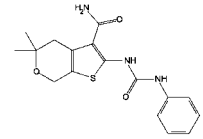
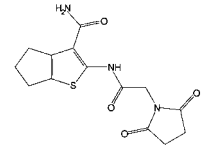
Figure 4. ROC plot for CDK2 showing improvement of the retrieval rate of known inhibitors by using filters. A set of 40 375 compounds from the ZINC library was used as starting database of inactive molecules. This set of 40 375 compounds was seeded with 73 known inhibitors, and the filters described below were applied to both ZINC compounds and inhibitors. Poses were first filtered according to an $E_{vdW} \leq -35$ kcal/mol and an $E_{vdw}/MW \leq -0.1$ kcal/g, which yielded 171 898 poses of 14 701 ZINC compounds. These 14 701 compounds and the 48 inhibitors that passed the above filter make up the set to which all subsequent filters (see also Table 5) were applied. (Black curve) The LIECE two-parameter CDK2-model was used to rank the combined library consisting of 48 known inhibitors of CDK2 and the pose with the most favorable LIECE energy of each of the 14 701 compounds. The area under this curve is $A_{ROC} = 0.76$. (Red curve) Forty-eight known inhibitors and 7793 compounds with Coulombic efficiency more favorable than −0.01 kcal/g, $A_{ROC} = 0.79$. (Green curve) Forty-eight known inhibitors and 6716 compounds with at least one hydrogen bond to the hinge region, $A_{ROC} = 0.82$. (Blue curve) Thirty-two known inhibitors and 11 842 compounds with van der Waals efficiency more favorable than −0.11 kcal/g, $A_{ROC} = 0.81$. (Orange curve) Thirty-two known inhibitors and 2654 compounds that satisfy all aforementioned filters, $A_{ROC} = 0.86$. The dashed line is the random model.

4. Conclusions

The main goal of this work was to evaluate the usefulness of LIECE in structure-based high-throughput docking. Particular emphasis was put on parameter robustness and transferability whose ultimate validation was based on the application of a LIECE model, fitted using inhibitors of CDK2, Lck, and p38, to rank poses of a large library of compounds and identify low-micromolar inhibitors of a kinase (EphB4) not used to derive the model.

In Silico Validation. Three main results emerge from the computational part of this study. First, a LIECE two-parameter model (van der Waals and electrostatics) based on three kinases shows good predictive ability for a set of 165 inhibitors of two distantly related kinases not used to generate the model. In particular, the binding free energy of 37 inhibitors of EphB4 and 128 inhibitors of EGFR is predicted with an average error of 1.3 and 1.5 kcal/mol, respectively. These results compare favorably with the extended linear response (ELR) approach with Monte Carlo simulations,[21] which has been applied to the same three kinases used as training set in the present work. A good predictive ability has been reported for a five-parameter ELR model, but the cross-validation did not include a completely independent test set of inhibitors and was based solely on the leave-one-out procedure.[21] A leave-one-out cross-validated correlation coefficient of 0.67 was reported for the 148 inhibitors used to derive the five-parameter ELR model,[21] while in the present study a value of 0.74 is obtained for the 165 inhibitors used to derive the two-parameter LIECE model. It has to be mentioned that the main objective of the ELR study was to

Table 6. Experimental Data on Kinase Inhibitors Discovered by High-Throughput Docking and LIECE Rankings

compound	MW [g/mol]	IC_{50}[a] [μM]	IC_{50}[b] [μM]	predicted ΔG[c] [kcal/mol]
EphB4 inhibitors				
1 	353	1.6, 1.8	6.8	−11.0
2 	337	1.5, 1.5	8.3	−10.7
3 	349	76	n.d.	−8.8
CDK2 inhibitor				
4 	321	n.d.	7.5, 8.0	−10.1

[a] Panvera FRET-based enzymatic assay. Compounds **1** and **2** were tested twice. Experiments performed using 5 concentrations (**1**, **3**) or 3 and 10 concentrations (**2**). [b] FRET-based enzymatic assay performed at Cerep (Celle l'Evescault, France). Duplicate measurements. [c] LIECE two-parameter CDK2-Lck-p38 model for EphB4 ranking and two-parameter CDK2 model for CDK2 ranking.

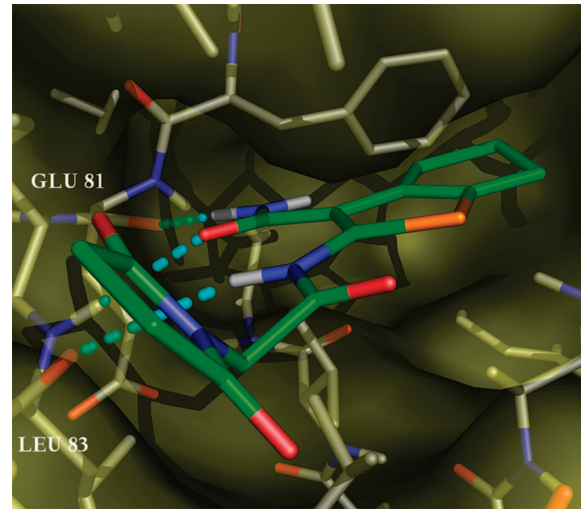


Figure 5. Predicted binding mode of compound **4**. Intermolecular hydrogen bonds to the hinge region are shown by dashed cyan lines. The pose shown is the one with the lowest binding energy according to the two-parameter CDK2 model after docking with DAIM, SEED, and FFLD and minimization with CHARMm22 in CDK2 (1KE5). Figure was prepared with PyMOL (DeLano Scientific, San Carlos, CA).

investigate the significance of individual energetic and structural descriptors (e.g., steric and electrostatic complementarity between kinase and inhibitor were present in all ELR models[21]). In fact, in contrast to the LIECE model, the ELR approach cannot be used for high-throughput docking because of the very high computational cost required for Monte Carlo sampling with explicit water treatment (22 Å water sphere for the ATP-binding site of kinases).

Second, the LIECE parameter transferability among kinases is mainly due to the dominance of the van der Waals interaction energy, whose multiplicative parameter is between 5 and 10 times larger than the one of the electrostatic free energy. A possible explanation is that the shape of the ATP-binding site is highly conserved among the different kinases. In a previous work it was found that LIECE parameters are not transferable between human β-secretase and HIV-1 protease,[22] despite the fact that both enzymes are aspartic proteases. This lack of generality originates from the significant differences between the substrate-binding site of mammalian and viral aspartic proteases which bind different polypeptide substrates. Moreover, the electrostatic interactions play a more important role in the substrate-binding site of β-secretase than HIV-1 protease.[22] Although the van der Waals term seems to dominate in the LIECE models of kinases, the ranking of large libraries seeded with a few known inhibitors shows significantly better enrichment factors using LIECE models with both van der Waals and electrostatics. A pure van der Waals model generates many false positives because it favors large compounds.

Third, upon high-throughput docking, it is essential to weed out compounds with unlikely binding modes to decrease the number of false positives in a LIECE ranking. In this work, the application of filters based on the van der Waals efficiency, the Coulombic efficiency, and the presence of key hydrogen bonds between ligand and protein proved useful. Being developed with information of known inhibitors, a LIECE model alone cannot identify all of the unlikely binding modes. Moreover, only energetic criteria are used by LIECE, namely, van der Waals and electrostatics. In this context, it is not surprising that the orientational filter requiring at least one hydrogen bond with the hinge region is most efficient in terms of enrichment of known inhibitors, as it uses information that is not already contained in the LIECE models.

In Vitro Validation. The ultimate test of any computational approach has to involve an experimental validation. In this work, the efficacy of the combined filter/LIECE approach was demonstrated by scoring the poses resulting from two recent high-throughput docking campaigns for discovering ATP-binding site inhibitors of the kinases EphB4 and CDK2. Overall, three novel inhibitors with IC_{50} values below 10 μM were identified with only 73 compounds tested. This shows that the combined use of filters and LIECE is capable of correctly selecting strong binders. It must be stressed again that the three inhibitors of EphB4 were discovered upon ranking with the LIECE two-parameter CDK2-Lck-p38-model. This result provides strong evidence on the transferability of the model. Most importantly, the usefulness of a general LIECE model for kinases is evident if one considers that it is impossible to derive a LIECE model based on the known EphB4 inhibitors, which are few and span a small range of activity values.

Acknowledgment. We thank Stjepan Jelakovic for the generation of the EphB4 model structure, and we thank Stephan Audétat, Catherine Berset, and Julia Tietz for activity measurements of the EphB4 compounds. We also thank Dariusz Ekonomiuk, Stefanie Muff, and Pietro Alfarano for comments on the manuscript. The calculations were performed on Matterhorn, a Beowulf Linux cluster at the Informatikdienste of the University of Zurich, and we thank C. Bolliger, T. Steenbock, and A. Godknecht for installing and maintaining the Linux cluster. We thank A. Widmer (Novartis Pharma, Basel, Switzerland) for providing a program for multiple linear regression and the molecular modeling program Wit!P, which was used for preparing the structures. This work was supported by the KTI (Kommission Technologie and Innovation) and the National Center of Competence in Research "Neural Plasticity and Repair" (NCCR Neuro).

Supporting Information Available: Depiction of the predicted binding mode of compound **14f**, histograms of the van der Waals interaction energy and the van der Waals efficiency, all 2D structures of known inhibitors, and enrichment factors for the three-protein one-, two-, and three-parameter models. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Holloway, M. K.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M. D.; Vacca, J. P.; Dorsey, B. D.; Levi, R. B.; Thompson, W. J.; Chen, L. J.; deSolms, S. J.; Gaffin, N.; Ghosh, A. K.; Giuliani, E. A.; Graham, S. L.; Guare, J. P.; Hungate, R. W.; Lyle, T. A.; Sanders, W. M.; Tucker, T. J.; Wiggins, M.; Wiscount, C. M.; Woltersdorf, O. W.; Young, S. D.; Darke, P. L.; Zugay, J. A. A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J. Med. Chem.* **1995**, *38*, 305–317.

(2) Apostolakis, J.; Caflisch, A. Computational ligand design. *Comb. Chem. High Throughput Screening* **1999**, *2*, 91–104.

(3) Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* **2002**, *45*, 2213–2221.

(4) Glen, R. C.; Allen, S. C. Ligand−protein docking: cancer research at the interface between biology and chemistry. *Curr. Med. Chem.* **2003**, *10*, 763–777.

(5) Walters, W. P.; Namchuk, M. Designing screens: how to make your hits a hit. *Nat. Rev. Drug Discovery* **2003**, *2*, 259–266.

(6) Alvarez, J. C. High-throughput docking as a source of novel drug leads. *Curr. Opin. Chem. Biol.* **2004**, *8*, 365–370.

(7) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.

(8) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.

(9) Huang, D.; Lüthi, U.; Kolb, P.; Edler, K.; Cecchini, M.; Audetat, S.; Barberis, A.; Caflisch, A. Discovery of cell-permeable non-peptide inhibitors of β-secretase by high-throughput docking and continuum electrostatics calculations. *J. Med. Chem.* **2005**, *48*, 5108–5111.

(10) 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.

(11) Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798–2814.

(12) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391.

(13) Hansson, T.; Åqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137–1144.

(14) 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.* **1996**, *40*, 1539–1549.

(15) Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27–35.

(16) Wang, J.; Dixon, R.; Kollman, P. Ranking ligand binding affinities with avidin: a molecular dynamics-based interaction energy study. *Proteins: Struct., Funct., Bioinf.* **1999**, *34*, 69–81.

(17) Carlson, H. A.; Jorgensen, W. L. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* **1995**, *99*, 10667–10673.

(18) Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding constants of neuraminidase inhibitors: an investigation of the linear interaction energy method. *J. Med. Chem.* **1999**, *42*, 5142–5152.
(19) Zhou, R.; Friesner, R. A.; Ghoshs, A.; Rizzo, R. C.; Jorgensen, W. J.; Levy, R. M. New linear interaction method for binding affinity calculations using a continuum solvent model. *J. Phys. Chem. B* **2001**, *102*, 10388–10397.
(20) Tounge, B. A.; Reynolds, C. H. Calculation of the binding affinity of β-secretase inhibitors using the linear interaction energy method. *J. Med. Chem.* **2003**, *46*, 2074–2082.
(21) Tominaga, Y.; Jorgensen, W. L. General model for estimation of the inhibition of protein kinases using Monte Carlo simulations. *J. Med. Chem.* **2004**, *47*, 2534–2549.
(22) Huang, D.; Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatic solvation. *J. Med. Chem.* **2004**, *47*, 5791–5797.
(23) Stjernschantz, E.; Marelius, J.; Medina, C.; Jacobsson, M.; Vermeulen, N. P. E.; Oostenbrink, C. Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? An evaluation of the linear interaction energy (LIE) method. *J. Chem. Inf. Model.* **2006**, *46*, 1972–1983.
(24) Ersmark, K.; Nervall, M.; Hamelink, E.; Janka, L. K.; Clemente, J. C.; Dunn, B. M.; Blackman, M. J.; Samuelsson, B.; Åqvist, J.; Hallberg, A. Synthesis of malarial plasmepsin inhibitors and prediction of binding modes by molecular dynamics simulations. *J. Med. Chem.* **2005**, *48*, 6090–6106.
(25) 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.
(26) Dancey, J.; Sausville, E. A. Issues and progress with protein kinase inhibitors for cancer treatment. *Nat. Rev. Drug Discovery* **2003**, *2*, 296–313.
(27) Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Science* **2004**, *303*, 1800–1805.
(28) Smith, C. M.; Shindyalov, I. N.; Veretnik, S.; Gribskov, M.; Taylor, S. S.; Ten- Eyck, L. F.; Bourne, P. E. The protein kinase resource. *Trends Biochem. Sci.* **1997**, *22*, 444–446.
(29) Cohen, P. Protein kinases—the major drug targets of the twenty-first century? *Nat. Rev. Drug Discovery* **2002**, *1*, 309–315.
(30) Aparna, V.; Rambabu, G.; Panigrahi, S. K.; Sarma, J. A. R. P.; Desiraju, G. R. Virtual screening of 4-anilinoquinazoline analogues as EGFR kinase inhibitors: importance of hydrogen bonds in the evaluation of poses and scoring functions. *J. Chem. Inf. Model.* **2005**, *45*, 725–738.
(31) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T.; Harris, P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.* **2001**, *44*, 4339–4358.
(32) Gibson, A. E.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Curtin, N. J.; Davies, T. G.; Endicott, J. A.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Whitfield, H. J. Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O6-substituted guanine derivatives. *J. Med. Chem.* **2002**, *45*, 3381–3393.
(33) Chen, P.; Norris, D.; Iwanowicz, E. J.; Spergel, S. H.; Lin, J.; Gu, H. H.; Shen, Z. Q.; Wityak, J.; Lin, T. A.; Pang, S. H.; de Fex, H. F.; Pitt, S.; Shen, D. R.; Doweyko, A. M.; Bassolino, D. A.; Roberge, J.; Poss, M. A.; Chen, B. C.; Schieven, G. L.; Barrish, J. C. Discovery and initial SAR of imidazoquinoxalines as inhibitors of the src-family kinase p56(Lck). *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1361–1364.
(34) Chen, P.; Iwanowicz, E. J.; Norris, D.; Gu, H. H.; Lin, J.; Moquin, R. V; Das, J.; Wityak, J.; Spergel, S. H.; de Fex, H.; Pang, S. H.; Pitt, S.; Shen, D. R.; Schieven, G. L.; Barrish, J. C. Synthesis and SAR of novel imidazoquinoxaline-based Lck inhibitors: improvement of cell potency. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3153–3156.
(35) Stelmach, J. E.; Liu, L. P.; Patela, S. B.; Pivnichny, J. V.; Scapin, G.; Singh, S.; Hop, C. E. C. A.; Wang, Z.; Strauss, J. R.; Cameron, P. M.; Nichols, E. A.; O'Keefe, S. J.; O'Neill, E. A.; Schmatz, D. M.; Schwartz, C. D.; Thompson, C. M.; Zaller, D. M.; Doherty, J. B. Design and synthesis of potent, orally bioavailable dihydroquinazolinone inhibitors of p38 MAP kinase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 277–280.
(36) Berset, C.; Audétat, S.; Tietz, J.; Gunde, T.; Barberis, A.; Schumacher, A.; Traxler, P. Protein Kinase Inhibitors. WO/2005/120513, 2005 (Oncalis AG).
(37) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
(38) Momany, F.; Rone, R. Validation of the general purpose QUANTA 3.2/CHARMm force field. *J. Comput. Chem.* **1992**, *13*, 888–900.
(39) No, K.; Grant, J.; Scheraga, H. 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.* **1990**, *94*, 4732–4739.
(40) No, K.; Grant, J.; Jhon, M.; Scheraga, H. 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.* **1990**, *94*, 4740–4746.
(41) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
(42) Šali, A.; Blundell, T. L. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
(43) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Šali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.
(44) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
(45) Holm, L.; Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **1993**, *233*, 123–138.
(46) Laskowski, R. A.; MacArthur, M. W.; Moss, D.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
(47) Im, W.; Beglov, D.; Roux, B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the poisson-boltzmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59–75.
(48) Feng, B. Y.; Simeonov, A.; Jadhav, A.; Babaoglu, K.; Inglese, J.; Shoichet, B. K.; Austin, C. P. B. High-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.* **2007**, *50*, 2385–2390.
(49) Shoichet, B. K. Screening in a spirit haunted world. *Drug Discovery Today* **2006**, *11*, 607–615.
(50) Rockey, W. M.; Elcock, A. H. Rapid computational identification of the targets of protein kinase inhibitors. *J. Med. Chem.* **2005**, *48*, 4138–4152.
(51) Irwin, J. J.; Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
(52) Kolb, P.; Caflisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.* **2006**, *49*, 7384–7392.
(53) Budin, N.; Majeux, N.; Caflisch, A. Fragment-based flexible ligand docking by evolutionary opimization. *Biol. Chem.* **2001**, *382*, 1365–1372.
(54) Cecchini, M.; Kolb, P.; Majeux, N.; Caflisch, A. Automated docking of highly flexible ligands by genetic algorithms: A critical assessment. *J. Comput. Chem.* **2004**, *25*, 412–422.
(55) Zweig, M. H.; Campbell, G. Receiver-operating characteristic (ROC) plots—a fundamental evaluation tool in clinical medicine. *Clin. Chem.* **1993**, *39*, 561–577.

JM070654J