

A transferable linear interaction energy model for kinases: derivation and inhibitor discovery.

Peter Kolb^{1,2,†}, Danzhi Huang¹, Fabian Dey¹, and Amedeo Caflisch^{1,‡}

[[†]kolb@blur.compbio.ucsf.edu, [‡]caflisch@bioc.uzh.ch]

¹Department of Biochemistry, University of Zurich, Zurich, Switzerland and ²present address: Department of Pharmaceutical Chemistry, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158, USA



KTI/CTI

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. 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.

The LIE Method

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 [1]. In LIE, the free energy of binding is approximated by

$$\Delta G = \alpha \left(\langle E^{vdW} \rangle_{bound} - \langle E^{vdW} \rangle_{free} \right) + \beta \left(\langle E^{elec} \rangle_{bound} - \langle E^{elec} \rangle_{free} \right) \quad (1)$$

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 or Monte Carlo trajectory.

The LIECE Method

Unfortunately, LIE cannot be used for high-throughput docking because of its computational requirements (the currently fastest implementation needs about six hours for each compound). 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, i.e., the numerical solution of the Poisson equation by the finite-difference technique [2]. The LIECE approach, where the last two letters stand for continuum electrostatics, is about two orders of magnitude faster than previous LIE methods and shows a similar precision on the targets tested.

General LIECE model for kinases

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

$$\Delta G = \alpha \Delta E_{vdW} \quad (2)$$

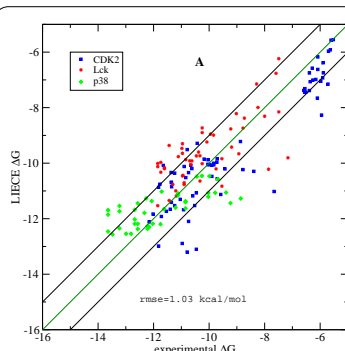
a two-parameter model with continuum electrostatics

$$\Delta G = \alpha \Delta E_{vdW} + \beta \Delta G_{elec} \quad (3)$$

and a three-parameter model with decomposed electrostatics

$$\Delta G = \alpha \Delta E_{vdW} + \beta_1 \Delta E_{coul} + \beta_2 \Delta G_{solvat} \quad (4)$$

where 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_{solvat}).



General LIECE model

We derived a LIECE model that is applicable to kinases in general by combining datasets for three different kinases: CDK2, Lck, and P38 [3]. In total, this training set contained 165 inhibitors with ΔG values spanning 8 kcal/mol.

Panel A: experimental vs. predicted ΔG for the compounds for which the model has been derived. Predicted values have been calculated using equation (3) with the values of α and β listed in Table 1. The 73 inhibitors of CDK2 are depicted as blue squares, the 51 inhibitors of Lck as red stars and the 41 inhibitors of P38 as green diamonds, respectively.

Panels B and C: to demonstrate the transferability of the parameters, we predicted the binding affinity for 37 inhibitors of EphB4 and 128 inhibitors of EGFR.

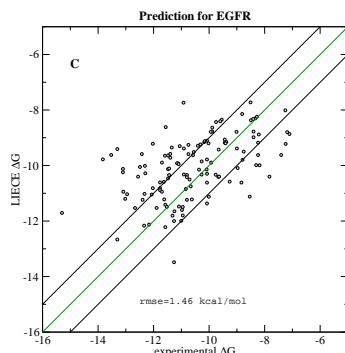
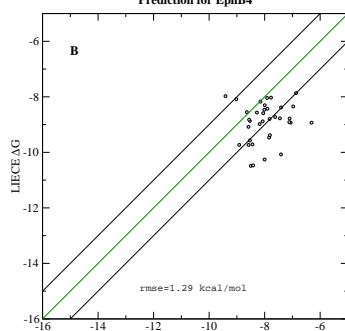
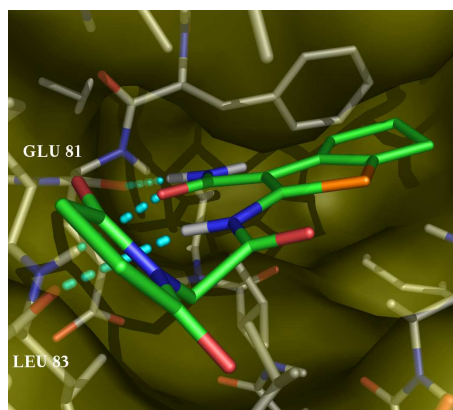


Table 1. LIECE parameters for the three-protein models.

	α	β or β_1	β_2	rms ^a [kcal/mol]	LOO ^b cv q^2
$\alpha \Delta E_{vdW}$	0.2463	-	-	1.13	0.69
standard deviation	± 0.0019	-	-		
$\alpha \Delta E_{vdW} + \beta \Delta G_{elec}$	0.2898	0.0442	-	1.03	0.74
standard deviation	± 0.0075	± 0.0074	-		
$\alpha \Delta E_{vdW} + \beta_1 \Delta E_{coul} + \beta_2 \Delta G_{solvat}$	0.2961	0.0325	0.0454	1.03	0.74
standard deviation	± 0.0089	± 0.0115	± 0.0075		

^aRoot mean square of the error when predicting ΔG values. ^bLeave-one-out cross-validated q^2 .



Virtual screening

We screened a diversity-based subset of the ZINC library containing 40375 compounds against CDK2, which resulted in a hit with an IC_{50} of 7 μM (left). Moreover, in a similar screen against EphB4, three more compounds could be identified, two of which bound with IC_{50} s around 1 μM [4].

Conclusions

The general LIECE model for kinases can predict binding affinities within 1–1.5 kcal/mol for a panel of five kinases. In virtual screening applications, LIECE achieves good enrichments, although the elimination of false positives with simple filters (cutoffs in van der Waals and electrostatic interaction energies) is still necessary (cf. [3]).

References

- [1] ÅGVIK J, MEDINA C, SAMUELSSON J-E. A new method for predicting binding affinity in computer-aided drug design. *Prot. Eng.* 7 (1994), 385.
- [2] HUANG D, CAFLISCH A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.* 47 (2004), 5791.
- [3] KOLB P, HUANG D, DEY F, CAFLISCH A. Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J. Med. Chem.* 51 (2008), 1179.
- [4] KOLB P, BERSSET-KIPOUROUS C, HUANG D, CAFLISCH A. Structure-based tailoring of compound libraries for high-throughput screening: Discovery of novel EphB4 kinase inhibitors. *Proteins: Struct. Funct. Bioinf.* 73 (2008), 11.