University of Zurich
Zurich Open Repository and Archive

Year: 2011

Hydrogen bonding penalty upon ligand binding

Zhao, H; Huang, D

Hydrogen bonding penalty upon ligand binding

Abstract

Ligand binding involves breakage of hydrogen bonds with water molecules and formation of new hydrogen bonds between protein and ligand. In this work, the change of hydrogen bonding energy in the binding process, namely hydrogen bonding penalty, is evaluated with a new method. The hydrogen bonding penalty can not only be used to filter unrealistic poses in docking, but also improve the accuracy of binding energy calculation. A new model integrated with hydrogen bonding penalty for free energy calculation gives a root mean square error of 0.7 kcal/mol on 74 inhibitors in the training set and of 1.1 kcal/mol on 64 inhibitors in the test set. Moreover, an application of hydrogen bonding penalty into a high throughput docking campaign for EphB4 inhibitors is presented, and remarkably, three novel scaffolds are discovered out of seven tested. The binding affinity and ligand efficiency of the most potent compound is about 300 nM and 0.35 kcal/mol per non-hydrogen atom, respectively.

PLoS one

Hydrogen Bonding Penalty upon Ligand Binding

Hongtao Zhao, Danzhi Huang*

Department of Biochemistry, University of Zurich, Zurich, Switzerland

Abstract

Ligand binding involves breakage of hydrogen bonds with water molecules and formation of new hydrogen bonds between protein and ligand. In this work, the change of hydrogen bonding energy in the binding process, namely hydrogen bonding penalty, is evaluated with a new method. The hydrogen bonding penalty can not only be used to filter unrealistic poses in docking, but also improve the accuracy of binding energy calculation. A new model integrated with hydrogen bonding penalty for free energy calculation gives a root mean square error of 0.7 kcal/mol on 74 inhibitors in the training set and of 1.1 kcal/mol on 64 inhibitors in the test set. Moreover, an application of hydrogen bonding penalty into a high throughput docking campaign for EphB4 inhibitors is presented, and remarkably, three novel scaffolds are discovered out of seven tested. The binding affinity and ligand efficiency of the most potent compound is about 300 nM and 0.35 kcal/mol per non-hydrogen atom, respectively.

Introduction

Hydrogen bonding is an exchange reaction whereby the hydrogen bond donors and acceptors of the free protein and ligand break their hydrogen bonds with water and form new ones in the protein-ligand complex [1,2,3]. About thirty years ago, Wilkinson and coworkers found mutation of Cys-35 in Tyrosyl-tRNA synthetase to Ser-35 causes poorer ATP binding and catalysis although the hydroxyl group of serine forms far stronger hydrogen bonds than does the thiol group of cysteine [1]. Analysis of the hydrogen bonding geometry revealed that a hydrogen bond of Ser-35 is at least 0.5 Å longer than the optimum. Accordingly, Ser-35 would have to lose a good hydrogen bond with a bound water molecule to form this weak hydrogen bond with ATP in the enzyme-substrate complex, and thus the mutant shows poorer binding and catalysis. Therefore, enthalpic loss in hydrogen bonding could take place upon ligand binding if not compensated by formation of good hydrogen bonds between the protein and ligand.

Virtual screening has emerged as an efficient tool in drug discovery from lead identification to optimization and beyond [4,5]. However, scoring functions that model the solvent environment as a continuum [6,7] are still grossly inaccurate [8]. The role of individual waters can be critical in predication of binding affinities, and continuum models often provide poor results in treating bound waters in a confined cavity [9]. Glide docks explicit waters into the binding site and measures the exposure of polar/charged groups to the explicit waters. When a polar/charged ligand or protein group is judged to be inadequately solvated, a desolvation penalty is assessed [9,10]. By contrast, most other scoring functions [11] do not properly take into account the enthalpic loss of hydrogen bonding upon ligand binding. Incorporation of bound water molecules into molecular docking was suggested for improvement of accuracy [12]. On the

other hand, in high-throughput molecular docking campaigns a significant part of binding poses are rather unrealistic, e.g. burial of polar atoms in hydrophobic sites, and thus discarding them at an early stage is desirable. Filters such as van der Waals efficiency based on arbitrary cutoff are often used to remove poses that unlikely bind [13]. However, it seems lack of a reliable and efficient filter with transferable cutoff among different proteins.

Protein kinases play an important role in cell-signaling pathways regulating a variety of cellular functions. Dysregulation of kinase activity has been implicated in pathological conditions ranging from neuronal disorders to cellular transformation in leukemia [14]. The tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4) is involved in cancer related angiogenesis [15]. So far, two high-throughput virtual screening campaigns have been reported, with two scaffolds identified in the low micromolar range [13,16]. Highly potent EphB4 inhibitors have been developed via chemical synthesis [17,18,19]. The marketed drug dasatinib, with Abl1 and Src as primary targets, also shows a very high affinity to Eph kinases [20].

Here, we report a new approach to calculate hydrogen bonding penalty (HBP) associated with ligand binding. HBP is further integrated into a binding energy calculation, and the fitted parameter of 1.7 kcal/mol is consistent with the estimate of contribution by formation of one neutral hydrogen bond ranging from 0.5 to 1.5 kcal/mol [21]. Moreover, statistics of HBP in kinase crystal structures and an application in a high-throughput docking campaign is presented.

Methods

Binding of a ligand to a protein involves the breakage of hydrogen bonds with water molecules and formation of new hydrogen bonds between the protein and ligand, which can be described by the following equation [21] by using one pair of

donor (D) and acceptor (A):

$$D \cdots OH_2 + A \cdots HOH \leftrightarrow D \cdots A + HOH \cdots OH_2 \quad (1)$$
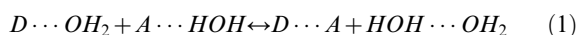
Based on hydrogen bonding being an exchange reaction [1,21,22], its energy can be described using normalized weights:

$$E_{HB-unbound} = [(w_D + w_{O-H_2O}) + (w_A + w_{H-H_2O})] * E_{HB} \quad (2)$$

$$E_{HB-bound} = [f_{hb} * (w_D + w_A) + (w_{O-H_2O} + w_{H-H_2O})] * E_{HB} \quad (3)$$

wherein, w_D and w_A is the hydrogen bonding weight of a donor or acceptor, respectively, f_{hb} stands for the fraction of hydrogen bonding relative to that of an optimum geometry, and E_{HB} is unit hydrogen bonding energy. Hydrogen bonds with water are assumed to be in the optimum geometry. HBP (p_{HB}) associates with ligand binding is then described as

$$p_{HB} = (1 - f_{hb}) * (w_D + w_A) \quad (4)$$

Probing hydrogen bonding status

Oxygen and nitrogen atoms in double or triple bonds are regarded as hydrogen bond acceptors, and hydrogen atoms bonded to oxygen, nitrogen or sulfur atoms are regarded as hydrogen bond donors. The existence of C—H...O hydrogen bonds has been confirmed by neutron diffraction data on organic compounds [23]. Analysis of 100 kinase crystal structures complexed with small molecule inhibitors at a resolution of at least 2.5 Å gives 64 short C—H...O interactions, showing typical hydrogen bonding features (Figure S1).

Each hydrogen bond donor or acceptor at the binding interface is firstly checked whether it forms hydrogen bond with water molecules. For this purpose, an optimum solvation radius (r_{sol}) is defined for each donor/acceptor and if a water molecule can be placed within 0.15 Å of the r_{sol} no penalty is applied. Here, 2.8 and 2.9 Å are used as r_{sol} for any oxygen and nitrogen, respectively, which were derived from an analysis of 397 crystal structures with X-ray resolutions below 1.0 Å (Figure S2). The r_{sol} of polar hydrogen is 1.9 Å (except 2.15 Å for H bonded to sulfur), which is the difference between the r_{sol} of nitrogen and the bond length [24] of N—H. The r_{sol} of other atom types are listed in Figure 1 and the values are mainly adapted based on the van der Waals radii of Bondi [25]. Details of probing hydrogen bonds with water were described in File S1. In case of not forming hydrogen bonds with water, the possibility of forming hydrogen bonds between the protein and ligand (including intra-molecular hydrogen bonds) is further checked and penalty (p_{HB}) is then calculated.

Fraction of hydrogen bonding

Similar to the strategy of evaluating hydrogen bonding energy in LUDI [26], the following equations are used to calculate the fraction of hydrogen bonding (f_{hb}) to that of an optimum geometry.

$$f_{hb}(r,\theta) \quad = \quad f(r) \cdot f(\theta) \quad (5)$$

	Atom types	Hydrogen bonding weight	r_{sol} (Å)
	=O, −O⁻	1.0	
O	−OH	0.5	2.8
	−O−	0.2	
N	=N−, ≡N	1.0	2.9
	−NH⁺	1.0	
	−OH, −NH, −NH₃⁺, −NH₂⁺	0.5	
H	−NH₂	0.3	1.88
	CH=N, HC=O	0	
	−SH	0.2	2.15
S	=S, −S−	-	3.2
C	Unsaturated	-	3.0
	Saturated	-	3.2
H	Apolar	-	2.28
F	-	-	2.99
Cl	-	-	3.27
Br	-	-	3.37
I	-	-	3.50

Figure 1. Hydrogen bonding weights and solvation radii of different atom types.
doi:10.1371/journal.pone.0019923.g001

$$f(r) = \begin{cases} 1 & \text{if} \quad r \leq 2.0 \\ 1 - 0.5 * (r - 2.0) & \text{if} \quad 2.0 < r \leq 2.8 \quad (6) \\ 0 & \text{if} \quad r > 2.8 \end{cases}$$

$$f(\theta) = \begin{cases} 1 & \text{if} \quad \theta \geq 150^o \\ 4.0 * (-\cos\theta)^{0.25} - 2.86 & \text{if} \quad 110^o \leq \theta < 150^o \quad (7) \\ 0 & \text{if} \quad \theta < 110^o \end{cases}$$

wherein, r is the distance between the hydrogen atom and the acceptor and θ is the angle centered at hydrogen among donor, hydrogen and acceptor. The equation to calculate $f(r)$ and $f(\theta)$ as well as the upper and lower limit in r and θ are derived from the calculation using density functional theory [27]. In case of one hydrogen atom is shared by two acceptors or one acceptor interacting with two donors, the f_{hb} for the corresponding donor/acceptor is additive but with 1 as the upper limit.

Hydrogen bonding penalty

The HBP at the protein-ligand interface is summarized over each donor/acceptor as

$$P_{HB} = \sum_{pro,lig} w * (1 - \sum f_{hb}) \qquad (8)$$

However, no penalty is applied for protein atoms which are not water accessible before ligand binding or participate in intramolecular hydrogen bonds. Initial guess of hydrogen bonding weights (w) is based on chemical intuition by considering atomic partial charge and water solubility of a few small molecules (Table S1). Empirical weights as proof-of-principle are then optimized with a trial-and-error procedure according to the fitted parameter in the binding free energy calibration.

Evaluation of binding free energy

The equation used for fitting the calculated energies to the experimental free energies of binding ($\Delta G = RT\ln(K_d)$) is a three-parameter model

$$\Delta G = \alpha \Delta E_{ff} + \beta P_{HB} + \gamma \qquad (9)$$

where, ΔE_{ff} is the interaction energy between the ligand and the protein calculated by the CHARMm force filed [28] and P_{HB} stands for HBP. Three parameters α, β, and γ are generated with fitting. ΔE_{ff} is calculated by the following equation:

$$\Delta E_{ff} = \Delta E_{vdW} + \Delta E_{coul} + \Delta G_{solv} + \Delta E_{strain} \qquad (10)$$

where, ΔE_{vdW} is the intermolecular van der Waals energy, ΔE_{coul} is the intermolecular Coulombic energy in vacuo, ΔE_{strain} is the strain energy of ligand upon binding, and ΔG_{sol} is the change in solvation energy of ligand and protein upon binding.

The van der Waals and Coulombic interaction energy are calculated by subtracting the values of the isolated components from the energy of the complex with CHARMM [29] and the CHARMm22 force filed [28]. The van der Waals energy is calculated using the default nonbonding cutoff of 14 Å. Coulombic energy is calculated using infinite cutoff and a dielectric constant of 2.0. The electrostatic solvation energy was calculated by the finite-difference Poisson approach (FDP) [30] using PBEQ module [31] in CHARMM and a focusing procedure with a final grid spacing of 0.25 Å. The size of the initial grid is determined by considering a layer of at least 12.5 Å around the solute. The dielectric discontinuity surface was delimited by the van der Waals surface. The ionic strength is set to zero and the temperature to 300 K. Two finite-difference Poisson calculations are performed for each of the three systems (protein, ligand, and protein/ligand complex). The exterior dielectric constant was set to 78.5 and 2.0 for the first and second calculation, respectively, while the solute dielectric constant is 2.0 to take polar fluctuations into account. The solvation energy is the difference between the two calculations. The strain energy of the ligand is the energy difference between the bound and global minimum. Here, the global minimum is the one showing the lowest $E_{vdW} + E_{coul} + E_{bonded} + G_{sol}$ among all the poses that have been minimized outside of the protein.

Twenty-three inhibitors [32] of CDK2 (1H0V), 24 inhibitors [18] (**8** to **32**, excluding **30**) of EphB4 (2VWX), and 27 uncharged inhibitors [33] of p38 alpha MAP kinase (3GC7) are used as the training set. Thirty type II inhibitors [34] of Braf (3II5), 14 charged inhibitors [33] of p38 alpha and another 20 p38 alpha inhibitors [35] (1YWR) are used as the test set. Protein structures were taken from the X-ray structure (PDB code indicated in the brackets) and prepared as described below. Some key physio-chemical properties of inhibitors are summarized in Figure S3.

Version 4 of AutoDock [36] was used to generate the binding poses over the conformational search space using the Lamarckian genetic algorithm. The binding site was determined by 4.0 Å away from any atom of the ligand complexed in the respective protein structure. The number of energy evaluations was 2,750,000 and the number of poses was 50. Poses were further clustered using all atom RMSD cutoff of 0.3 Å to remove redundancy and in average 20 cluster representatives were kept. All other parameters were set as default. A few poses for each inhibitor were also generated by manual modification of the scaffold present in the respective crystal structure. All poses were further minimized by CHARMM in the respective proteins. The protein structure was kept rigid in all steps.

Preparation of protein-ligand complexes

One hundred kinase crystal structures (including 15 different classes, File S2) complexed with small molecule inhibitors at a resolution of at least 2.5 Å were downloaded from Protein Data Bank for analysis of HBP. Hydrogen atoms were added according to the protonation states of chemical groups at pH 7. Partial charges were then assigned using MPEOE method [37,38]. The added hydrogen atoms were minimized by the conjugate gradient algorithm to a RMS 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 heavy atoms were fixed.

Preparation of the compounds library for virtual screening

The compounds were selected from Zinc library [39]. Preparation included the assignment of CHARMm atom types, force field parameters [28], and partial charges [37,38], and energy minimization with a distance dependent dielectric function using the program CHARMM [29].

Enzymatic assay

In vitro kinase activity was measured using the Panvera Z'lyte Tyr2 kinase assay PV3191 (Invitrogen) according to the

manufacturer's instructions. The reaction assay (10 μL) contained 7.5 ng of EphB4 kinase (Proqinase, Germany), 30 μM ATP, and 5% DMSO. The reaction was performed at room temperature for 1 h.

Results and Discussion

Statistics of hydrogen bonding penalty in kinase complexes

Small HBPs can be observed for the binding modes of inhibitors in the X-ray structures. One example is c-Kit tyrosine kinase with its apo and holo form in complex with Imatinib (PDB codes 1T45 and 1T46). In the apo conformation, donors/acceptors at the ATP binding site form hydrogen bonds with bound water molecules. While upon ligand binding, as shown in the holo conformation, some water molecules are displaced by Imatinib. HBP on the protein part is close to zero because new hydrogen bonds to the protein are formed to compensate for the replacement of the water molecules. However, one nitrogen atom of the Imatinib pyrimidine ring (N1 of Figure S4) becomes water inaccessible and does not form a new hydrogen bond, leading to a penalty of 1. By contrast, the other nitrogen atom (N2 of Figure S4) remains hydrogen bonding with a nearby bound water molecule and thus has no penalty.

To check the distribution of HBP values in crystal structures, 100 kinase-ligand complexes are investigated. In this data set, all the small molecule inhibitors have molecular weights from 200 to 700 g/mol and number of donors or acceptors from 2 to 11 (File S2). The HBP has been calculated for each of them and the values are in general small, with 62% smaller than 1 and 36% and 2% in the range from 1 to 2 and 2.0 to 2.1, respectively (Figure 2 and File S2). It has also been observed that larger HBPs appear in some X-ray structures, e.g., the structures of PDB code 3KVX and 1JSV, and the large values actually originate from poor fitting of small molecules to the density, a common problem in crystallography [40] which can be manifested by clash of atoms.

Distribution of HBPs for docked poses of small-molecule inhibitors is also evaluated. Here, the 138 molecules used in the binding free

energy calibration are docked into the corresponding protein binding sites with AutoDock. For each molecule, about 20 poses in average are generated. Then the HBPs and binding energies are calculated for all the poses. Firstly, the binding pose with the most favorable binding energy for each molecule (Figure S5) is selected and the distribution of HBPs is plotted. As can be observed from **B** of Figure 2, the distribution is similar to that of the 100 kinase complex structures (**A**). On the other hand, the distribution of all poses (**C**) spreads more widely with the largest HBP being 6.5. Compared with the HBPs in the crystal structures (**A**), 2 is a reasonable threshold, and about 50% of poses with unrealistic binding modes can be filtered out from further evaluations.

Hydrogen bonding penalty improves the accuracy of binding energies calculation

Binding energies can be calculated using equation 9 with the parameters obtained by least-squares fitting on the training data
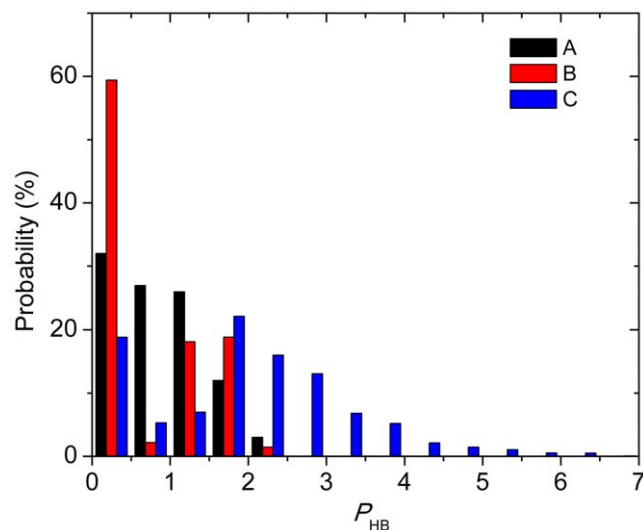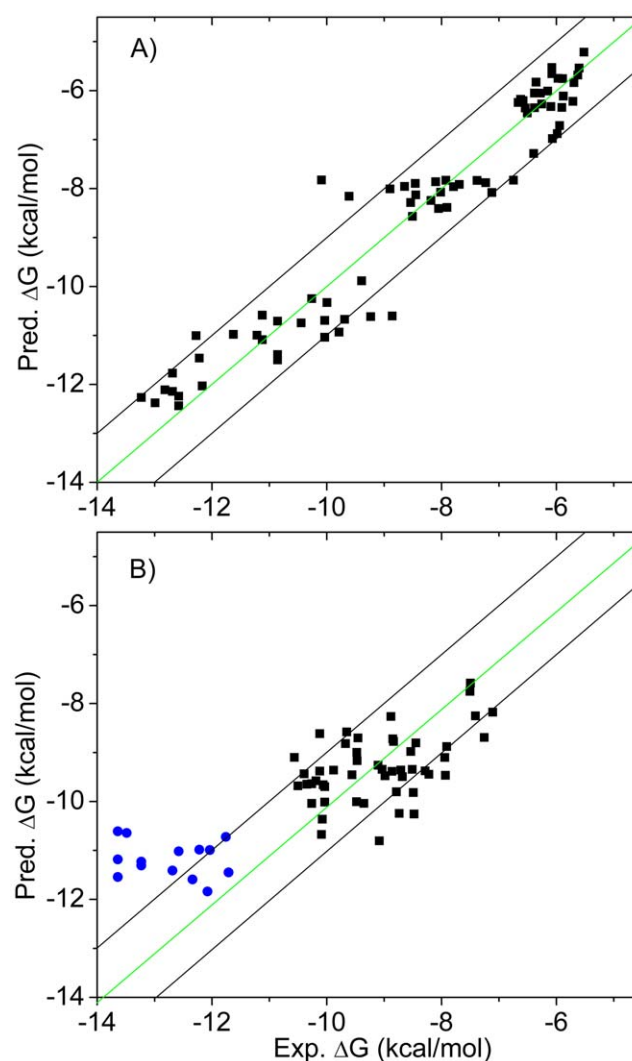

Figure 2. Distribution of hydrogen bonding penalties for: A) the binding modes in crystal structures of the 100 kinase complexes; B) poses with the most favorable calculated binding energies of the 138 molecules used in binding free energy calibration; C) all poses of the 138 molecules.
doi:10.1371/journal.pone.0019923.g002

Figure 3. Comparison of the calculated versus experimental binding energies. A) Training set of 74 inhibitors. $R^2 = 0.92$ and RMS error = 0.69 kcal/mol; B) Validation set of 64 inhibitors. RMS error = 1.12 kcal/mol. The blue dots indicated the 14 p38α inhibitors with one formal charge. The green diagonal line is the ideal line of perfect prediction. The black diagonals delimit the 1 kcal/mol error region.
doi:10.1371/journal.pone.0019923.g003

Table 1. Further validation of the three-parameter model with kinases and aspartic protease.

Protein	PDB code	ΔE_{ff} (kcal/mol)	P_{HB}	ΔG_{pred} (kcal/mol)	ΔG_{exp} (kcal/mol)
Abl	1OPJ	−64.80	1.24	−12.45	−10.81
Braf	1UWH	−57.61	1.27	−10.91	−10.45
JAK2	3E63	−30.18	0.00	−7.41	−7.91
Lck	2OFV	−59.13	0.53	−12.51	−13.23
JNK3	1PMV	−30.16	0.17	−7.12	−9.31
Ret	2X2L	−26.67	0.07	−6.58	−7.20
EGFR	1XKK	−66.60	2.30	−11.00	−10.91
CSrc	3G5D	−52.34	1.64	−9.19	−12.82
HIV-1 protease	1HIH	−65.71	1.49	−12.21	−11.01
	1HPX	−65.44	1.43	−12.26	−12.46
	1HXB	−61.24	0.95	−12.21	−13.49
	1HXW	−72.66	1.41	−13.78	−14.71
BACE-1	2QMF	−73.62	1.56	−13.72	−11.63
	2QP8	−68.76	0.47	−14.59	−11.05
	2XFI	−71.10	2.36	−11.83	−10.67

set of the 74 CDK2, EphB4, and p38α inhibitors as following:

$$\Delta G = 0.207 * \Delta E_{ff} + 1.72 * P_{HB} - 1.17 \quad (11)$$

The calculated binding energies show high correlation with the experimental values (R-square of 0.92) and a small RMS error of 0.69 kcal/mol (Figure 3A). Here, the parameter β corresponds to the unit hydrogen bonding energy. Notably, the fitted value 1.72 kcal/mol is in agreement with the experimental value, e.g., breakage of a neutral hydrogen bond resulting in loss of energy from 0.5 to 1.5 kcal/mol [21]. Moreover, a charged primary amine or carboxyl group has a hydrogen bonding weight of 1.5 or 2.0, which can lead to a maximal penalty of 2.58 or 3.44 kcal/mol upon loss of the hydrogen bond/salt bridge. This value also agrees well with the experimental data (up to 4 kcal/mol) [21]. Hydrogen bonding weights were further used to rank the strength of individual hydrogen bonds in DNA base pairs, exhibiting good compatibility with the previously reported results (File S3).

The fitted model has been validated on a test set including 14 charged p38α inhibitors and 30 type II Braf inhibitors, with an RMS error of 1.12 kcal/mol (Figure 3B). Moreover, validation with different kinases shows general transferability of this model (Table 1). Transferability can be also seen for aspartic protease, e.g., HIV-1 protease and β-secretase, although a shift of 2.0 kcal/mol can be observed for the latter. Previously, we reported a two-parameter LIECE model for kinase inhibitors [13], which is not transferable for type II kinase inhibitors, HIV-protease or β-secretase inhibitors. The binding affinities predicted by the two-parameter LIECE on the 24 type I EphB4 inhibitors show about −5.0 kcal/mol shift compared with the experimental values (Table S2). Clearly, the incorporation of HBP into the scoring function improves the general transferability besides the role of ligand reorganization energy [41].

The derived model includes calculation of solvation energy by FDP which requires about 6 min on a single Intel 2.8 GHz CPU. Replacing the FDP approach with a distance-dependent dielectric model for solvation energy calculation gives similar accuracy for the neutral inhibitors at a much fast speed (10 seconds). However, distance-dependent dielectric model can only apply for non-charged compounds due to inaccurate treatment of the solvation effect, and also more false positives in a high-throughput virtual screening are observed. This comparison indicates that accurate calculation of solvation energies in prediction of binding affinities is necessary.

Virtual screening for EphB4 inhibitors

In a recent high throughput docking study for EphB4 inhibitors, ZINC "leads-now" library of about 20 million compounds
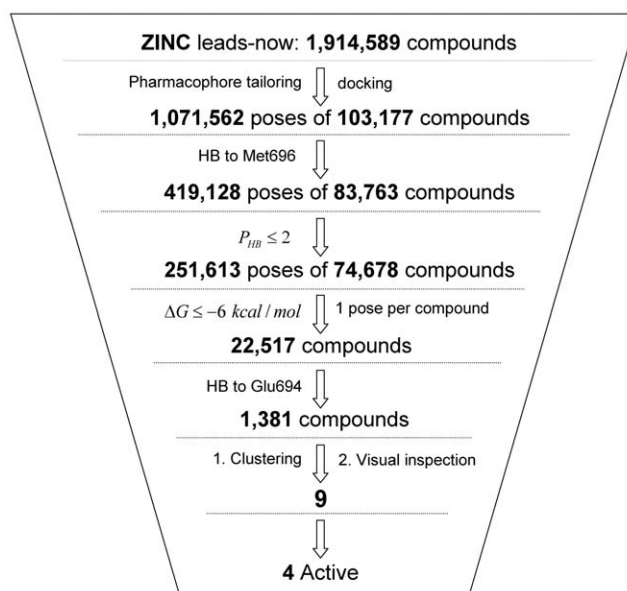

Figure 4. Schematic picture of the high throughput docking approach. HB stands for hydrogen bond. Met696 and Glu694 are the two key residues of the hinge loop (see also Figure 6).
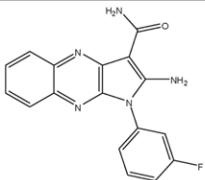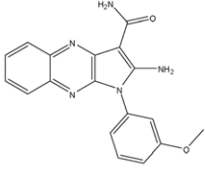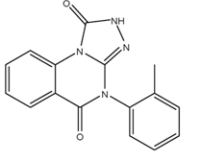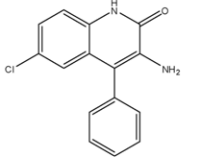doi:10.1371/journal.pone.0019923.g004

Compound	Mw (g/mol)	P_{HB}	Pred. ΔG (kcal/mol)	IC_{50} (µM)[a]
1	321	0.22	-7.9	0.38
2	333	0.74	-8.0	0.30
3	292	1.06	-6.6	15.7
4	270	0.61	-6.4	42% at 20 µM

Figure 5. Identified EphB4 inhibitors by high throughput docking. [a] All IC_{50} values are means of two to four dose-response measurements.
doi:10.1371/journal.pone.0019923.g005

(Mw≤350 and cLogP≤3.5) was first tailored by a pharmacophore model to generate a focused library of 103,177 compounds. This pharmacophore model was specifically designed for EphB4 type I inhibitors, consisting of a bi-dentate hydrogen bonding pattern and a conjugate hydrophobic group to be located in the deep ATP back pocket as well as geometric constraints thereof (H. Zhao, unpublished results). To our best knowledge, all known type I EphB4 inhibitors [13,16,17,18] can fulfill this model.

The focused library was docked by AutoDock 4 and about 1 million poses were generated by clustering with a RMSD cutoff of 1.0 Å. The cluster representatives which do not form a hydrogen bond to NH of Met696 were further filtered out. The HBP (≤2) was then used to remove unrealistic poses (about 40%). The remaining poses were further ranked by the predicted binding energy, and the top about 30% compounds (22,517) with calculated binding energy smaller than −6 kcal/mol (~50 µM) were kept. Among them, 1381 compounds forming a hydrogen bond to Glu694 were selected and can be classified into 80 structural scaffolds. Finally, 7 scaffolds (9 compounds) of them were purchased for experimental measurements based on visual inspection of the binding modes, commercial availability and structural novelty. The procedures used in the virtual screening are shown in Figure 4. Comparison of the performances between the proposed and AutoDock 4 scoring function is shown in Figure S6.

Notably, 4 of the 9 tested compounds show inhibitory activity at micro-molar to high nano-molar range, with the most active compound showing IC_{50} at 300 nM (Figure 5). Interestingly, the two compound also show a high ligand efficiency [42] of −0.35 kcal/mol per non-hydrogen atom. The predicted binding mode of compound **3** (Figure 6) is further confirmed by the preliminary X-ray crystallography (J. Dong, unpublished results).
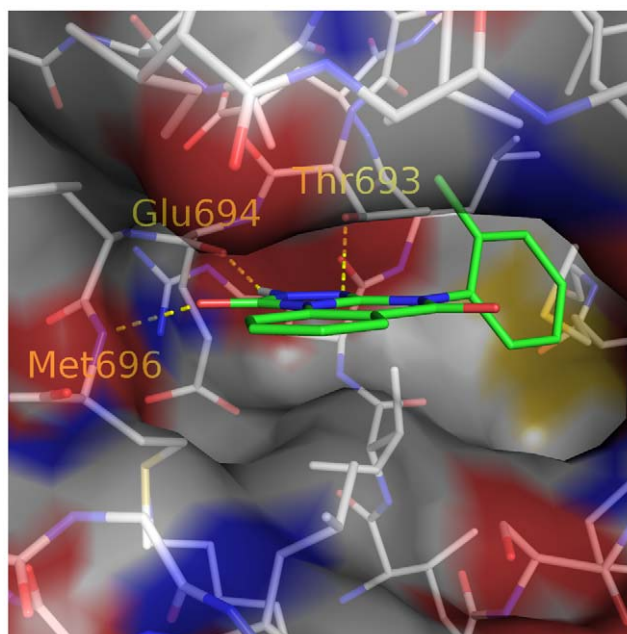

Figure 6. Binding mode of compound 3 (carbon atoms in green) predicted by docking. The intermolecular hydrogen bonds to the residues at the hinge loop (Glu694 and Met696) and the gatekeeper (Thr693) are shown by yellow dashed lines. The protein surface is colored based on atom types with carbon in white, oxygen in red, and nitrogen in blue. This figure was prepared using PyMOL (Delano Scientific, San Carlos, CA).
doi:10.1371/journal.pone.0019923.g006

Conclusion

Hydrogen bonding in biological system is a complex phenomenon as water competes with ligand for the hydrogen bonding sites. Removal of a group that forms a hydrogen bond in unfavorable geometry actually improves binding [21]. In view of hydrogen bonding being an exchange reaction [1,21,22], a new approach is proposed to evaluate the HBP upon ligand binding. Analysis of the 100 crystal structures indicates the penalty in general is low, predominantly smaller than 2 for inhibitors. A high throughput docking case shows HBP can function as an efficient filter to remove poses that unlikely bind. Incorporation of HBP into binding free energy calculation can significantly improve the predictive accuracy and transferability. The fitted parameter of 1.72 kcal/mol means loss of a neutral hydrogen bond would result in a penalty of from 0.34 to 1.72 kcal/mol in binding energy, consistent with the experimental data from 0.5 to 1.5 kcal/mol [21]. Four inhibitors of three scaffolds were discovered out of nine tested, and the binding affinity and ligand efficiency of the most potent compound is about 300 nM and 0.35 kcal/mol per non-hydrogen atom, respectively.

Supporting Information

Figure S1   Scatter plot of C—H⋯O angles against H⋯O distances in short C—H⋯O interactions between ligands and proteins.
(DOC)

Figure S2   Distribution of distances between crystal water oxygen and oxygen or nitrogen atoms of proteins.
(DOC)

Figure S3   Distribution of some key properties of the inhibitors used in the training and test set.
(DOC)

Figure S4   2D plot of the binding mode of Imatinib. Upon ligand binding, one nitrogen atom of the Imatinib pyrimidine ring (N1) becomes water inaccessible and does not form a new hydrogen bond, leading to a penalty of 1. By contrast, the other nitrogen atom (N2) remains hydrogen bonding with a nearby bound water molecule and thus has no penalty.
(DOC)

Figure S5   Poses with the most favorable binding energy of inhibitors of CDK2 (A), EphB4 (B), p38 α (C), Braf (D)

and another set of p38α inhibitors (E). The molecules with bonds in red are the binding modes of the corresponding scaffolds in the crystal structures.
(DOC)

Figure S6   Distribution of predicted binding affinities by Autodock4 (black) and the proposed scoring function (red) on 74,678 compounds passing the first two filters (HB to Met696 and $P_{HB} \leq 2$ kcal/mol). Bin size: 0.1 kcal/mol.
(DOC)

Table S1   MPEOE partial charge and water solubility of model small molecules used to generate initial guess of hydrogen bonding weights.
(DOC)

Table S2   Two-parameter LIECE energy and hydrogen bonding penalty on the 24 EphB4 inhibitors.
(DOC)

File S1   Probing hydrogen bonds formed with implicit water.
(DOC)

File S2   Hydrogen bonding penalty of the 100 kinase complex structures.
(DOC)

File S3   Ranking the strength of individual hydrogen bonds in DNA base pairs.
(DOC)

Acknowledgments

Author Contributions

Conceived and designed the experiments: DH HZ. Performed the experiments: HZ. Analyzed the data: HZ DH. Contributed reagents/materials/analysis tools: HZ. Wrote the paper: HZ DH.

References

1. Wilkinson AJ, Fersht AR, Blow DM, Winter G (1983) Site-directed mutagenesis as a probe of enzyme structure and catalysis: tyrosyl-tRNA synthetase cysteine-35 to glycine-35 mutation. Biochemistry 22: 3581–3586.
2. Winter G, Fersht AR, Wilkinson AJ, Zoller M, Smith M (1982) Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. Nature 299: 756–758.
3. Wilkinson AJ, Fersht AR, Blow DM, Carter P, Winter G (1984) A large increase in enzyme-substrate affinity by protein engineering. Nature 307: 187–188.
4. Bajorath J (2002) Integration of virtual and high-throughput screening. Nat Rev Drug Discov 1: 882–894.
5. Langer T, Hoffmann RD (2001) Virtual screening: an effective tool for lead structure discovery? Curr Pharm Des 7: 509–527.
6. Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268: 1144–1149.
7. Feig M, Onufriev A, Lee MS, Im W, Case DA, et al. (2004) Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 25: 265–284.
8. Schneider G (2010) Virtual screening: an endless staircase? Nat Rev Drug Discov 9: 273–276.
9. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, et al. (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 49: 6177–6196.
10. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47: 1739–1749.
11. Wang R, Lu Y, Wang S (2003) Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 46: 2287–2303.
12. Thilagavathi R, Mancera RL (2010) Ligand-protein cross-docking with water molecules. J Chem Inf Model 50: 415–421.
13. Kolb P, Huang D, Dey F, Caflisch A (2008) Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. J Med Chem 51: 1179–1188.
14. Hunter T (1998) The role of tyrosine phosphorylation in cell growth and disease. Harvey Lect 94: 81–119.
15. Pennisi A, Ling W, Li X, Khan S, Shaughnessy JD, Jr., et al. (2009) The ephrinB2/EphB4 axis is dysregulated in osteoprogenitors from myeloma patients and its activation affects myeloma bone disease and tumor growth. Blood 114: 1803–1812.
16. Zhou T, Caflisch A (2010) High-throughput virtual screening using quantum mechanical probes: discovery of selective kinase inhibitors. ChemMedChem 5: 1007–1014.
17. Miyazaki Y, Nakano M, Sato H, Truesdale AT, Stuart JD, et al. (2007) Design and effective synthesis of novel templates, 3,7-diphenyl-4-amino-thieno and furo-[3,2-c]pyridines as protein kinase inhibitors and in vitro

evaluation targeting angiogenetic kinases. Bioorg Med Chem Lett 17: 250–254.

18. Bardelle C, Cross D, Davenport S, Kettle JG, Ko EJ, et al. (2008) Inhibitors of the tyrosine kinase EphB4. Part 1: Structure-based design and optimization of a series of 2,4-bis-anilinopyrimidines. Bioorg Med Chem Lett 18: 2776–2780.

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

20. Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, et al. (2008) A quantitative analysis of kinase inhibitor selectivity. Nat Biotechnol 26: 127–132.

21. Fersht AR, Shi JP, Knill-Jones J, Lowe DM, Wilkinson AJ, et al. (1985) Hydrogen bonding and biological specificity analysed by protein engineering. Nature 314: 235–238.

22. Hine J (1972) Structural Effects on Rates and Equilibria .15. Hydrogen-Bonded Intermediates and Stepwise Mechanisms for Proton-Exchange Reactions between Oxygen-Atoms in Hydroxylic Solvents. J Am Chem Soc 94: 5766–&.

23. Steiner T, Saenger W (1993) Role of C-H…O Hydrogen-Bonds in the Coordination of Water-Molecules - Analysis of Neutron-Diffraction Data. J Am Chem Soc 115: 4540–4547.

24. Cordero B, Gomez V, Platero-Prats AE, Reves M, Echeverria J, et al. (2008) Covalent radii revisited. Dalton Trans. pp 2832–2838.

25. Bondi A (1964) Van Der Waals Volumes+Radii. Journal of Physical Chemistry 68: 441–&.

26. Bohm HJ (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 8: 243–256.

27. Morozov AV, Kortemme T, Tsemekhman K, Baker D (2004) Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. Proc Natl Acad Sci U S A 101: 6946–6951.

28. Momany FA, Rone R (1992) Validation of the General-Purpose Quanta(R)3.2/Charmm(R) Force-Field. J Comput Chem 13: 888–900.

29. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comput Chem 4: 187–217.

30. Warwicker J, Watson HC (1982) Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. J Mol Biol 157: 671–679.

31. Im W, Beglov D, Roux B (1998) Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Computer Physics Communications 111: 59–75.

32. Gibson AE, Arris CE, Bentley J, Boyle FT, Curtin NJ, et al. (2002) Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O(6)-substituted guanine derivatives. J Med Chem 45: 3381–3393.

33. Stelmach JE, Liu L, Patel SB, Pivnichny JV, Scapin G, et al. (2003) Design and synthesis of potent, orally bioavailable dihydroquinazolinone inhibitors of p38 MAP kinase. Bioorg Med Chem Lett 13: 277–280.

34. Berger DM, Torres N, Dutia M, Powell D, Ciszewski G, et al. (2009) Non-hinge-binding pyrazolo[1,5-a]pyrimidines as potent B-Raf kinase inhibitors. Bioorg Med Chem Lett 19: 6519–6523.

35. Koch P, Jahns H, Schattel V, Goettert M, Laufer S (2010) Pyridinylquinoxalines and pyridinylpyridopyrazines as lead compounds for novel p38 alpha mitogen-activated protein kinase inhibitors. J Med Chem 53: 1128–1137.

36. Goodsell DS, Olson AJ (1990) Automated docking of substrates to proteins by simulated annealing. Proteins 8: 195–202.

37. No KT, Grant JA, Scheraga HA (1990) Determination of Net Atomic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method .1. Application to Neutral Molecules as Models for Polypeptides. Journal of Physical Chemistry 94: 4732–4739.

38. No KT, Grant JA, Jhon MS, Scheraga HA (1990) Determination of Net Atomic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method .2. Application to Ionic and Aromatic-Molecules as Models for Polypeptides. Journal of Physical Chemistry 94: 4740–4746.

39. Irwin JJ, Shoichet BK (2005) ZINC–a free database of commercially available compounds for virtual screening. J Chem Inf Model 45: 177–182.

40. Hawkins PC, Warren GL, Skillman AG, Nicholls A (2008) How to do an evaluation: pitfalls and traps. J Comput Aided Mol Des 22: 179–190.

41. Yang CY, Sun H, Chen J, Nikolovska-Coleska Z, Wang S (2009) Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction. J Am Chem Soc 131: 13709–13721.

42. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9: 430–431.