Can Relative Binding Free Energy Predict Selectivity of Reversible Covalent Inhibitors?

Payal Chatterjee[a,$], **Wesley M. Botello-Smith**[a,$], **Han Zhang**[a], **Li Qian**[a], **Abdelaziz Alsamarah**[a], **David Kent**[a], **Jerome J. Lacroix**[b], **Michel Baudry**[b], and **Yun Luo**[a]

[a]Western University of Health Sciences, College of pharmacy, Pomona CA 91766

[b]Western University of Health Sciences, Graduate College of Biomedical Sciences, Pomona CA 91766

Abstract

Reversible covalent inhibitors have many clinical advantages over noncovalent or covalent drugs. However, apart from selecting a warhead, substantial efforts in design and synthesis are needed to optimize noncovalent interactions to improve target-selective binding. Computational prediction of binding affinity for reversible covalent inhibitors presents a unique challenge since the binding process consists of multiple steps, which are not necessarily independent of each other. In this study, we lay out the relation between relative binding free energy and the overall reversible covalent binding affinity using a two-state binding model. To prove the concept, we employed free energy perturbation (FEP) coupled with λ-exchange molecular dynamics method to calculate the binding free energy of a series of α-ketoamide analogs relative to a common warhead scaffold, in both noncovalent and covalent bond states, and for two highly homologous proteases, calpain-1 and calpain-2. We conclude that covalent binding affinity alone, in general, can be used to predict reversible covalent binding selectivity. However, exceptions may exist. Therefore, we also discuss the conditions under which the noncovalent binding step is no longer negligible and propose a novel approach that combines the relative FEP calculations with a single QM/MM calculation of warhead to predict the binding affinity and binding kinetics for a large number of reversible covalent inhibitors. Our FEP calculations also revealed that covalent and noncovalent states of an inhibitor do not necessarily exhibit the same selectivity. Thus, investigating both binding states, as well as the kinetics will provide extremely useful information for optimizing reversible covalent inhibitors.

Graphical abstract

Introduction

The advantages of covalent over non-covalent inhibitors include long residence time, higher potency, and decreased drug resistance[1-2]. In the past two years, a number of covalent inhibitors such as carfilzomib, telaprevir, abiraterone, and afatinib have been approved by the FDA for various clinical indications, ushering in a new era for covalent modifiers[3-4]. From a lead optimization perspective, covalent inhibitor design is not restricted by the maximum binding affinity of 1.5 kcal/mol per nonhydrogen atom limitation[5], which has been hampering noncovalent drug design for decades. The main hurdle for covalent inhibitor development is the lack of specificity or selectivity. The risk of toxic events occurring due to the use of covalent inhibitors can be lessened through modulation of electrophilic warhead reactivity and optimization of noncovalent interactions, which may improve target receptor recognition and increase the selectivity of covalent inhibitors. A recent review highlighted the progress in quantum mechanics/molecular mechanics (QM/MM) methods for predicting warhead reactivity and mechanism in the binding site[6]. However, once an ideal electrophilic warhead is found for a specific target, substantial efforts in design and synthesis are needed to optimize the noncovalent interactions to improve the selectivity of covalent inhibitors. Computational prediction of covalent inhibitor binding affinity presents a unique challenge since the binding process consists of multiple steps, which are not necessarily independent of each other. Because of these associated difficulties, computational tools for optimizing covalent drugs are far less developed than for noncovalent drugs. The majority of tools that exist for use in pursuing a covalent inhibitor design are implemented within various molecular docking programs in which the searching algorithms and scoring functions have been adjusted from noncovalent docking to suit covalent docking[6]. A QM-based scoring function was also developed and shown improved correlation with IC_{50} for irreversible covalent inhibitors[7]. Engels and coworkers successfully developed covalent reversible inhibitors from irreversible inhibitors using a QM/MM and docking combined protocol[8].

Free energy calculation approaches, such as free energy perturbation (FEP), have been considered as most rigorous approach for predicting the binding affinity of noncovalent drugs and has became a standard protocol in pharmaceutical industry to rank molecule candidates at later stage of lead optimization[9-12]. However, its application in covalent binder is scarce. Kuhn et al. has recently performed a pioneering work of prioritizing covalent inhibitors using FEP on covalent binding state[13]. In the current study, we focus on investigating the following fundamental question: for a given reversible covalent inhibitor, is the binding affinity determined solely by the noncovalent binding state (*Michaelis* complex analog), the covalent binding state, or from both states? Such question is, foremost, important for understanding the fundamental concepts and limitations of applying FEP method to covalent binding processes, which is critical for the emerging field of covalent inhibitor design.

As a proof of concept, we investigated α-ketoamide analogs, which covalently bind to the catalytic site of calcium-dependent cysteine proteases, calpain-1 and calpain-2, in a reversible manner (Figure 1) [14-16]. Capain-1 and calpain-2 are two members of the calpain family, which are ubiquitously present in mammalian brains. Strikingly, despite their 71% sequence identity in their proteolytic core, they play opposite functions in both synaptic plasticity and neuroprotection/neurodegeneration, with calpain-1 being neuroprotective and calpain-2 being predominantly neurodegenerative[17-18]. The differential functions of two calpain isoforms underscore the critical need to design inhibitors that can selectively target calpain-2 but not calpain-1, as indicated in a recent review[19].

α-ketoamide peptidomimetics are among the reversible covalent inhibitors that have become increasingly popular in drug design as they show advantages in cytosolic stability and improved selectivity over irreversible compounds[6, 8, 20]. They bear weak to moderate electrophilic warheads, such as ketoamide or acrylamide, which must first be properly oriented within the target protein before it can react efficiently with an appropriate nucleophilic residue. However, even when a reversible warhead is employed, achieving selectivity can still be challenging. This is especially true when the molecule targets a catalytic cleft that is highly conserved among protein isoforms. In addition, the S1′ binding site of calpains consist of two flexible loops. Such flexible binding site is difficult to model using covalent docking algorithms, which further emphasize the need of FEP methodology that allows protein, ligand and solvent dynamics. Thus the calpain systems represent an ideal benchmark system to investigate whether the FEP methodology could be a reliable computational tool for predicting the binding selectivity of reversible covalent inhibitors.

We employed a free energy perturbation (FEP) coupled with λ-exchange molecular dynamics method to calculate binding free energies of a series of α-ketoamide analogs relative to a common warhead scaffold, in both noncovalent and covalent bond states, for calpain-1 and calpain-2, respectively. We obtained a strong correlation between relative binding free energy of covalent states and experimental binding selectivity. We discussed the relation between the covalent and noncovalent binding states and proved mathematically that the noncovalent state becomes negligible in predicting the overall binding selectivity only when the covalent binding is at least 5.5 kcal/mol stronger than that noncovalent binding. If this condition is not satisfied, we proposed an approach that combines the relative FEP

calculations described here with a single QM/MM calculation of the warhead core structure to estimate the thermodynamics and kinetics of reversible covalent inhibitors. To our knowledge, this is also the only FEP study thus far which revealed that for a reversible covalent binder, the covalent step and noncovalent step can have different binding selectivity. Furthermore, we speculated that there exists an intrinsic relation between noncovalent binding affinity and activation barrier due to the unique kinetics of some reversible covalent binders. Thus, investigating both binding states, as well as the kinetics will provide extremely useful information for optimizing reversible covalent inhibitors.

Theory

Relation between relative binding free energy and covalent binding affinity

$$\text{E+I} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \text{E} \cdot \text{I} \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} \text{EI} \qquad \text{Eq. 1}$$
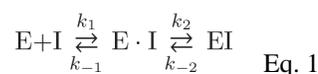
$$K_1 = \frac{k_{-1}}{k_1} = \frac{[E][I]}{[E \cdot I]} = e^{\Delta G_{\text{dm}}/\text{RT}}$$
$$K_2 = \frac{k_{-2}}{k_2} = \frac{[E \cdot I]}{[EI]} = e^{\Delta G_{\text{mc}}/\text{RT}} \qquad \text{Eq. 2}$$

$$1/K_{\text{d}} = \frac{[E \cdot I] + [EI]}{[E][I]}$$
$$= \frac{1 + K_2}{K_1 K_2} = \frac{1 + e^{\Delta G_{\text{mc}}/\text{RT}}}{e^{(\Delta G_{\text{dm}} + \Delta G_{\text{mc}})/\text{RT}}}$$
$$= \frac{1 + e^{\Delta G_{\text{mc}}/\text{RT}}}{e^{\Delta G_{\text{dc}}/\text{RT}}}$$
$$= e^{-\Delta G_{\text{dc}}/\text{RT}}$$
$$+ e^{(\Delta G_{\text{mc}} - \Delta G_{\text{dc}})/\text{RT}}$$
$$= e^{-\Delta G_{\text{dc}}/\text{RT}}$$
$$+ e^{-\Delta G_{\text{dm}}/\text{RT}} \qquad \text{Eq. 3}$$

Eq. 1 describes a general mechanism for the binding of a covalent inhibitor consisting of two steps (Figure 2). In the first step, an initial noncovalent binding of the inhibitor with the target protein positions the warhead close to the nucleophile. This complex is analogous to the Michaelis complex in the enzyme-substrate binding process. The subsequent step is the formation of a covalent bond, which generates the inhibited complex. The inhibitor binds irreversibly if $k_{-2} = 0$, whereas the inhibitor binds noncovalently if $k_2 = 0$. Reversible covalent inhibitors have finite values for both k_{-2} and $k_2 \cdot K_1$ and K_2 are the equilibrium dissociation constants of step1 and step2, which are thermodynamic quantities that are determined by the free energy difference between the two states, as illustrated in a simplified free energy profile of two-states binding in Figure 1. Namely, K_1 is determined by the free

energy difference between the dissociation and Michaelis states, G_{dm}. K_2 is determined by the free energy difference between the Michaelis and covalent states, G_{mc} (Eq. 2). The overall association constant of the reversible covalent inhibitor is defined by both K_1 and K_2, which depends on the binding free energy of both the covalent G_{dc} and noncovalent complexes G_{dm} (Eq. 3).

According to Eq 3, binding free energies for both the noncovalent and covalent bound states need to be calculated in order to estimate the overall binding free energy. However, direct calculation of the binding free energy of the covalent complex G_{dc} requires computing the energy for the formation of a covalent bond, along with multiple steps of proton transfer. Such computations require QM/MM calculations, which would become too expensive for high throughput applications, such as screening a large number of compounds to determine potential drug candidates. In this study, we investigate whether relative binding free energy calculation could serve as an alternative to computationally intensive QM/MM calculations, if modifications of the noncovalent binding interactions can be shown to have a negligible effect on the reactivity of the warhead. In this case, the binding free energy of the two states could be represented as the sum of the binding free energy of the common core structure, G^{core}, and the relative binding free energy between the full ligand and the core G (Eq 4).

$$\Delta G_{dc} = \Delta\Delta G_{dc} + \Delta G_{dc}^{core}$$
$$\Delta G_{dm} = \Delta\Delta G_{dm} + \Delta G_{dm}^{core} \qquad \text{Eq. 4}$$

Combining Eq. 4 with Eq. 3 provides the relation between the overall association constant $1/K_d$ and the relative binding free energies of the covalent state G_{dc} and the noncovalent state G_{dm}. For two ligands with the same warhead G^{core}, absolute binding free energy of the warhead at covalent and noncovalent states can be simplified as constants A and B, as shown in Eq. 5. In the current work, only G_{dc} and G_{dm} are calculated.

$$1/K_d = e^{-(\Delta\Delta G_{dc} + \Delta G_{dc}^{core})/RT} + e^{-(\Delta\Delta G_{dm} + \Delta G_{dm}^{core}))/RT}$$
$$\text{let } e^{-\Delta G_{dc}^{core}/RT} = A;\ e^{-\Delta G_{dm}^{core}/RT} = B$$
$$1/K_d = Ae^{-\Delta\Delta G_{dc}/RT} + Be^{-\Delta\Delta G_{dm}/RT} \qquad \text{Eq. 5}$$

Methods

System Preparation

The covalent binding complexes of ketoamide compounds in calpain-1 were built based on a co-crystal structure (PDB 2R9C) of a ketoamide compound GRD and calpain-1 (Figure 1). In the crystal structure, calpain-1 is in a Ca^{2+}-bound active state. The catalytic cysteine forms a hemithioketal complex with the ketoamide, which resembles the transition state of a peptide substrate cleavage. To prepare covalent complexes in calpain-2, a crystal structure of calpain-2 in complex with calpastatin (PDB 3BOW) was used in which calpain-2 is also in a Ca^{2+}-bound active state. The hemithioketal ligand is transferred from the calpain-1 catalytic

site to the structurally-aligned calpain-2 catalytic site using *Molecular Operating Environment* (MOE)[21]. To prepare the noncovalent binding complex, the covalent bond was cleaved using MOE. The catalytic cysteine was then converted back to anion to form the Cys-His ion pair, and the ligand was converted back to ketoamide. The catalytic mechanism of cysteine proteases has been previously studied using semiempirical quantum mechanics and QM/MM[22-23], which supports the existence of the Cys-His ion pair in the active state. For both calpain-1 and calpain-2, the entire catalytic domain (domain I-II) and the adjacent domain III, which is in contact with P3 site of ligand, were used for simulations. The missing domain III of calpain-1 in the crystal structure was modeled based on the crystal structure of calpain-2 and equilibrated for 700 ns (Supporting Information Methods and Figure S1). Both calpain-1 and calpain-2 structures remained in active conformation (catalytic triad in close contact) with two calcium ions remaining bound in the EF-hands of domain I and II during all simulations.

Force Field

CHARMM36 force field[24] was used for all simulations. CHARMM36 force fields for deprotonated cysteine (resname CYM) and protonated histidine (resname HSP) were used. Recent pKa calculations using a replica-exchange thermodynamic integration approach have shown that the CHARMM36 force field reproduces the pKa of papain cysteine at 4.4, close to the experimental pKa value of 3.3[25]. All ligands were parameterized using CHARMM General Force Field (CGenFF)[26-27]. Ketoamide warhead was re-parameterized using Force Field Toolkit plugin[28] of the VMD 1.9.3 together with Gaussian 09 revision E.01 package[29] (Supporting Information Methods, Figure S2, and Table S1). The covalent-linked ligand was parameterized as an artificial cysteine by combining the cysteine backbone parameters from protein force field with the ligand parameters from CGenFF.

Simulation Protocol

All the systems for MD simulation in explicit solvent were prepared by using CHARMM-GUI[30-31]. Each system was solvated into a rectangular water box composed of CHARMM TIP3P water molecules and 150 mM KCl, with an edge distance of 10 Å. All the simulations of ligand-receptor complex, including FEP simulations, were performed with NAMD 2.10[32] using periodic boundary conditions at constant temperature and pressure (NPT ensemble) of 300 K and 1 atm using Langevin thermostat[33] and Andersen-Hoover barostat[34]. Particle Mesh Ewald (PME) method was used for long-range electrostatic interactions[35]. A smoothing function was applied to van der Waals forces between 10 Å and 12 Å. The dynamics were propagated using Langevin dynamics with Langevin damping coefficient of 1 ps^{-1} and a time step of 2 fs. The SHAKE algorithm was applied to all hydrogen atoms[36]. The solvated enzyme-ligand complexes were minimized and equilibrated using a stepwise procedure setup by CHARMM-GUI. The homology model of calpain-1 domain III was simulated together with the catalytic domain for 700 ns using AMBER16[37] GPU version with the same force field.

FEP/λ-REMD for noncovalent and covalent binding states

To calculate the relative binding free energy using FEP, we applied free energy perturbation/ λ-exchange molecular dynamics (FEP/λ-REMD) implemented in NAMD program using

generalized scalable multiple copy algorithms[38]. It applies the staging simulation protocol to decouple the shifted Weeks-Chandler-Anderson repulsive and dispersive components of the Lennard-Jones potential[39], electrostatic contribution, and restraining potential[40-42]. The highly scalable lambda-exchange in FEP has achieved good convergence for the calculation of absolute binding affinity[43-45]. Here we adopted this algorithm for relative binding free energy calculation, in which only R groups, instead of the whole ligand, are decoupled using the same staging protocol. The rest of the ligand structure, which is called the core structure, has a neutral charge. As illustrated in Figure 3a, the relative noncovalent binding free energy between a ligand and a core is calculated from four separate FEP steps: the solvation free energy difference between the core structure and a particular ligand is calculated by decoupling the electrostatics and Lennard-Jones interactions between the R group and water molecules (ΔG1). To decouple the R group of a bound ligand from its environment, the electrostatics and Lennard-Jones interactions between ligand and its solvated protein environment are turned off gradually (ΔG2). All systems were first equilibrated prior to FEP/λ-REMD. The coordinates and velocities of the last equilibration step from NAMD were used to start the replica-exchange simulations. A thermodynamic coupling parameter λ was used to perturb the R groups from a fully coupled (λ=1) to a fully decoupled state (λ=0). A total number of 32 replicas (20 $\lambda_{dispersive}$, 6 $\lambda_{repulsive}$, 6 $\lambda_{electrostatic}$) were used to perturb the R-groups. The length of each FEP simulation varied between 2.0 to 3.0 ns per replica and showed good convergence (Figure S3). The λ-swapping attempts between neighboring replicas were performed frequently during each simulation (0.2 ps^{-1}). For noncovalent binding state, a harmonic restraint potential was applied between the sulfur of the catalytic cysteine and the warhead carbonyl carbon to avoid any possible dissociation. The free energy contribution of applying this restraint (ΔG6 and ΔG7) was calculated by using a stepwise decrement of the force constant over 13 windows (100 ps per window) to gradually reduce it from 20 to 0 kcal/mol/Å^2 using the thermodynamic integration framework within the *colvars* module in NAMD. For covalent binding free energy, since there is no extra distance restraint required for covalent ligand, the relative free energy between core and covalent ligand was calculated from two decoupling steps ΔG1 and ΔG2, using the same protocol as the one in noncovalent binding state.

Results and Discussion

Seven ketoamide analogs, which have reported inhibitory constants for calpain-1 and calpain-2 were used in this study as examples (Figure 4 top)[15]. Each compound was subjected to the relative free energy calculations in both the noncovalent and covalent states (see Methods), for calpain-1 and calpain-2, respectively. The experimental binding selectivity data and each calculated values are shown in Table S2. The correlations between experimental and computational binding selectivity of calpain-2 over calpain-1 are plotted in Figure 4 along with the statistics.

Covalent binding selectivity correlates well with overall binding selectivity, but exceptions may exist

Overall, we found fairly strong correlation from covalent binding state, but weak anti-correlation from the noncovalent binding state. The strong Pearson correlation coefficient

0.87 and rank correlation coefficients (Spearman's ϱ 0.85 and Kendall's τ 0.71) indicate that, for current system, we can use the relative binding free energy difference between the covalently bound core structure and a covalently bound ligand ($\Delta\Delta G_{dc}$) to predict and rank the experimental binding selectivity. Referring back to Eq. 5, such correlation will only exist when the second term ($Be^{-\Delta\Delta G_{dm}/RT}$) is sufficiently smaller than the first term ($Ae^{-\Delta\Delta G_{dc}/RT}$). When this is the case, Eq. 5 may be approximated to a single term form, hence the strong correlation between the overall binding affinity and the $\Delta\Delta G_{dc}$. This indicates that good prediction power can be attained using solely covalent binding state FEP, confirming what has been reported in previous literature on covalent binder[13]. However, what we learn from Eq. 5 is that such simplification is not always true. Mathematically, according to Eq. 6, this correlation only occurs when $e^{-(\Delta G_{mc}^{core}+\Delta\Delta G_{mc})/RT} \ggg 1$; or, more intuitively, if $\Delta G_{mc}^{core}+\Delta\Delta G_{mc} > 5.5\,\mathrm{kcal/mol}$, then $\dfrac{Ae^{-\Delta\Delta G_{dc}/RT}}{Be^{-\Delta\Delta G_{dm}/RT}} > 10^4$. Hence, for each ligand, the effects of the noncovalent intermediate can be neglected for the purposes of computing the overall dissociation constant K_d, only when the binding free energy of covalent state is at least 5.5 kcal/mol more favorable then the noncovalent state.

$$Ae^{-\Delta\Delta G_{dc}/RT}+Be^{-\Delta\Delta G_{dm}/RT} \approx Ae^{-\Delta\Delta G_{dc}/RT}$$
$$\text{when:} Ae^{-\Delta\Delta G_{dc}/RT} \ggg Be^{-\Delta\Delta G_{dm}/RT}$$

$$\therefore$$
$$\frac{Ae^{-\Delta\Delta G_{dc}/RT}}{Be^{-\Delta\Delta G_{dm}/RT}} = \frac{e^{-\Delta G_{dc}^{core}/RT} \times e^{-\Delta\Delta G_{dc}/RT}}{e^{-\Delta G_{dm}^{core}/RT} \times e^{-\Delta\Delta G_{dm}/RT}}$$
$$= e^{-(\Delta G_{dc}^{core}-\Delta G_{dm}^{core})/RT} \times e^{-(\Delta\Delta G_{dc}-\Delta\Delta G_{dm})/RT}$$
$$= e^{-(\Delta G_{mc}^{core})/RT} \times e^{-(\Delta\Delta G_{mc})/RT}$$
$$= e^{-(\Delta G_{mc}^{core}+\Delta\Delta G_{mc})/RT} \ggg 1 \qquad \text{Eq. 6}$$

The binding of the covalent state is generally more stable than the noncovalent state, since the former involves the free energy contributions of the formation of a covalent bond. However, if the noncovalent binding is strong and the reversible warhead is relative weak, the 5.5 kcal/mol rule aforementioned may not hold. In such case, the correlation between $\Delta\Delta G_{dc}$ and experimental K_d will not exist because the full Eq. 5 must be used to incorporate the non-negligible role of the noncovalent state. The broad usage of Eq 5 is that, even in such case, it is not necessary to calculate the absolute covalent binding for each ligand. One only needs to calculate ΔG_{dc}^{core} and ΔG_{dm}^{core} once for the core structure to get the constant value A and B, then Eq. 5 can be applied for a large number of ligands using the same FEP approaches described here, without the need to perform QM/MM calculation for each ligand.

Calpain-2 selective at covalent binding step but calpain-1 selective at noncovalent binding step

As discussed above, the strong correlation from covalent binding state indicates $Be^{-\Delta\Delta G_{dm}/RT}$ is sufficiently smaller than $Ae^{-\Delta\Delta G_{dc}/RT}$. Therefore, the contribution of the

noncovalent binding state is negligible for the overall binding selectivity, thus the correlation with overall binding free energy is not necessary. However, the anti-correlation between the experimental binding selectivity and the relative binding free energy of the noncovalent bound state obtained in this study is quite interesting. Such anti-correlation tells us that the compounds tested here that are calpain-2 selective at covalent binding step are more calpain-1 selective in noncovalent binding step. This points out an important message that improving the selectivity at noncovalent binding step will not necessarily improve the overall selectivity of a reversible covalent binder. The anti-correlation observed here could possibly be explained by the unique kinetics of α-ketoamide inhibitors. The binding kinetics of a series of peptidyl-α-ketoamide inhibitors were reported for the Hepatitis C NS3 protease[46]. It was shown that the binding selectivity is solely controlled by the association constant k_{on}. The dissociation constant k_{off} was insensitive to the structural changes remote from the ketoamide warhead. It was also reported that the formation of noncovalent complex is rapid comparing with the covalent bond formation[46-47]. Thus, it might be reasonable to assume that the activation barrier is much higher than the first noncovalent binding free energy barrier. k_{on} and k_{off} can thus be approximated to k_2 and k_{-2}, which is proportional to the activation energy ΔG_{on} and the deactivation energy ΔG_{off}, as illustrated in Figure 5a (green curve). As seen from our FEP results, both covalent and noncovalent binding free energies are changed simultaneously when there is a structural change located away from the warhead. If the structural change remote from the warhead stabilizes the covalent-bound state (negative ΔG_{dc}), the activation free energy ΔG_{on} may decrease, similar as described in Bell-Evans-Polanyi principle[48]. Likewise, if the same structural change destabilizes the noncovalent-bound state (positive ΔG_{dm}), the activation free energy ΔG_{on} may decrease (Figure 5a orange curve). Therefore, ΔG_{dc} and ΔG_{dm} have opposite effects on the kinetic parameter ΔG_{on}. In fact, as we mentioned in the introduction, for weak to moderate electrophilic warheads, the noncovalent binding step is important to properly orient the warhead towards the catalytic nucleophilic residue in the target protein. However, if the noncovalent binding is too stable, it may slow down the activation process, which requires structural reorganization to convert the sp2 hybridization carbon to sp3 hybridization. Our FEP calculations of both states pointed out that investigating both covalent and noncovalent binding states, as well as the binding kinetics would provide detailed insight for optimizing reversible covalent inhibitors.

$$y_1 = a_1 x^2 \quad \text{Eq. 7}$$

$$y_2 - \Delta G_{mc} = a_2 (x - d)^2 \quad \text{Eq. 8}$$

at the point of crossing between the parabolas,

$$a_1 x^2 - \Delta G_{mc} = a_2(x-d)^2 = a_2 x^2 - 2da_2 x + d^2 a_2 (a_2 - a_1)x^2 - 2da_2 x + d^2 a_2 + \Delta G_{mc} = 0$$

apply the quadratic formula,

$$x = \frac{2da_2 \pm \overline{)(2da_2)^2 - 4(a_2 - a_1)(d^2a_2 + \Delta G_{mc})}}{2(a_2 - a_1)}$$

assume $a_1 < a_2$

$$\Delta G^{\neq}_{on} = a_1 \left(\frac{2da_2 - \overline{)(2da_2)^2 - 4(a_2 - a_1)(d^2a_2 + \Delta G_{mc})}}{2(a_2 - a_1)} \right)^2 = a_1 \frac{\left(da_2 - \overline{)d^2a_2a_1 + (a_1 - a_2)\Delta G_{mc}} \right)^2}{(a_2 - a_1)^2}$$

Eq. 9

Relation between ΔG_{mc} and activation energy barrier ΔG_{on}

To estimate the relation between relative binding free energy and kinetics, we assume a two-state free energy profile represented by two parabolas along a reaction coordinate that defines the lowest-energy route (Figure 5b). ΔG_{on} can be approximated as the free energy barrier at the crossing point between two parabolas with quadratic function Eq. 7 and Eq. 8. To avoid confusion with kinetic rate constant, we indicate the force constants of two parabolas as a_1 and a_2. The binding free energy difference between two states is ΔG_{mc} and the distance between the two energy minima is d. Such a simplistic diagram bares similarity to Marcus rate theory, but differs in that covalent bond formation is being considered (as opposed to an electron transfer event) and that the two parabolas have different shapes, which is similar to proton transfer model Koepple and Kresge proposed[49-50]. Alternatively, other functional forms such as cubic function may be used to represent a more realistic two-state free energy profile. This needs further investigation and the best choice may depend on the system. Because only the second parabola involves a covalent bond formation, one can assume $a_1 < a_2$. Thus, the relation between ΔG_{on} and ΔG_{mc} can be obtained using Eq. 9, which depends on the shape of the two parabolas a_1 and a_2. a_1 and a_2 can be calculated based on the chosen reaction coordinate. For example, if the reaction coordinate is the distance between warhead and catalytic sulfur, a_1 and a_2 can be estimated using umbrella sampling simulation at noncovalent bound state and the covalent bound state. Same as Eq. 5, the broad usage of Eq. 9 lays in the fact that only one QM/MM calculation needs to be performed to obtain ΔG_{on} for the core. The ΔG_{on} for the inhibitors that share the same core can be estimated simply by adding $\Delta\Delta G_{mc}$ to ΔG_{mc} in Eq. 9. Since $\Delta\Delta G_{mc} = \Delta G_{dc} - \Delta G_{dm,}$ it is calculated from our FEP method. A workflow from QM/MM to FEP calculations for designing highly selective reversible covalent inhibitors is currently under development.

Conclusions

In conclusion, using a two-state binding model, the relation between relative binding free energy and the overall reversible covalent binding affinity is laid out. We evaluated the correlation between the calculated and experimental binding selectivity of reversible covalent inhibitors and discussed the potential applications and limitations of applying relative binding free energy calculations on reversible covalent inhibitors. We proved mathematically that when the covalent binding is >5.5 kcal/mol stronger that noncovalent binding, relative binding free energy of covalent state can be used to predict the binding selectivity. If such condition is not satisfied, we proposed an approach that can utilize relative FEP calculations described here to minimize the amount of QM/MM calculation needed to estimate the thermodynamics and kinetics of reversible covalent inhibitors. To satisfy Eq. 5 and Eq. 9, structural modifications remote from the warhead should not affect the electrophilicity of the warhead, both in bulk and in binding site. This precondition can be checked using various methods, such as single point QM/MM calculation using snapshots from MD trajectories. We also noticed from our calculations that the strong correlation was only obtained for binding selectivity, which is $\Delta\Delta G^{calpain2} - \Delta\Delta G^{calpain1}$, not with individual binding affinity $\Delta\Delta G_{dc}$ (Figure S4). It is well known that error cancelation exists between different ligands when calculating the relative solvation free energy and binding free energies. In our case, error cancelation between calpain-1 and calpain-2 may originate from using MM to treat the catalytic triad and bound states or may also originate from experimental measurements. Therefore, the method described here is particularly advantageous for predicting binding selectivity among protein isoforms, which is the major challenge in covalent inhibitors design.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Mah R, Thomas JR, Shafer CM. Bioorg Med Chem Lett. 2014; 24:33–9. [PubMed: 24314671]

2. Backus KM, Correia BE, Lum KM, Forli S, Horning BD, Gonzalez-Paez GE, Chatterjee S, Lanning BR, Teijaro JR, Olson AJ, Wolan DW, Cravatt BF. Nature. 2016; 534:570–4. [PubMed: 27309814]

3. Singh J, Petter RC, Baillie TA, Whitty A. Nat Rev Drug Discov. 2011; 10:307–17. [PubMed: 21455239]

4. Mehdi, S. Covalent Enzyme Inhibition in Drug Discovery and Development. In: Yang, HC.Y, WK., McCarthy, JR., editors. Enzyme Technologies: Pluripotent Players in Discovering Therapeutic Agents. John Wiley and Sons; Hoboken, New Jersey: 2014. p. 81-129.

5. Kuntz ID, Chen K, Sharp KA, Kollman PA. Proc Natl Acad Sci U S A. 1999; 96:9997–10002. [PubMed: 10468550]

6. Awoonor-Williams E, Walsh AG, Rowley CN. Biochim Biophys Acta. 2017 doi:10.1016/j.bbapap. 2017.05.009.

7. Fanfrlik J, Brahmkshatriya PS, Rezac J, Jilkova A, Horn M, Mares M, Hobza P, Lepsik M. J Phys Chem B. 2013; 117:14973–82. [PubMed: 24195769]

8. Schirmeister T, Kesselring J, Jung S, Schneider TH, Weickert A, Becker J, Lee W, Bamberger D, Wich PR, Distler U, Tenzer S, Johe P, Hellmich UA, Engels B. J Am Chem Soc. 2016; 138:8332–5. [PubMed: 27347738]

9. Jorgensen WL, Ravimohan C. J Chem Phys. 1985; 83:3050–3054.

10. Kollman PA. Chem Rev. 1993; 93:2395–2417.

11. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, Murcko M, Frye L, Farid R, Lin T, Mobley DL, Jorgensen WL, Berne BJ, Friesner RA, Abel R. J Am Chem Soc. 2015; 137:2695–703. [PubMed: 25625324]

12. Homeyer N, Stoll F, Hillisch A, Gohlke H. J Chem Theory Comput. 2014; 10:3331–44. [PubMed: 26588302]

13. Kuhn B, Tichy M, Wang L, Robinson S, Martin RE, Kuglstatter A, Benz J, Giroud M, Schirmeister T, Abel R, Diederich F, Hert J. J Med Chem. 2017; 60:2485–2497. [PubMed: 28287264]

14. Qian J, Cuerrier D, Davies PL, Li Z, Powers JC, Campbell RL. J Med Chem. 2008; 51:5264–70. [PubMed: 18702462]

15. Li Z, Ortega-Vilain AC, Patil GS, Chu DL, Foreman JE, Eveleth DD, Powers JC. J Med Chem. 1996; 39:4089–98. [PubMed: 8831774]

16. Li Z, Patil GS, Golubski ZE, Hori H, Tehrani K, Foreman JE, Eveleth DD, Bartus RT, Powers JC. J Med Chem. 1993; 36:3472–80. [PubMed: 8230139]

17. Wang Y, Zhu G, Briz V, Hsu YT, Bi X, Baudry M. Nat Commun. 2014; 5:3051. [PubMed: 24394804]

18. Baudry M, Bi X. Trends Neurosci. 2016; 39:235–45. [PubMed: 26874794]

19. Ono Y, Saido TC, Sorimachi H. Nat Rev Drug Discov. 2016; 15:854–876. [PubMed: 27833121]

20. Bradshaw JM, McFarland JM, Paavilainen VO, Bisconte A, Tam D, Phan VT, Romanov S, Finkle D, Shu J, Patel V, Ton T, Li X, Loughhead DG, Nunn PA, Karr DE, Gerritsen ME, Funk JO, Owens TD, Verner E, Brameld KA, Hill RJ, Goldstein DM, Taunton J. Nat Chem Biol. 2015; 11:525–31. [PubMed: 26006010]

21. Inc., C. C. G. Molecular Operating Environment (MOE), 2013.08. Chemical Computing Group Inc; Montreal, QC, Canada: 2016.

22. Ma S, Devi-Kesavan LS, Gao J. J Am Chem Soc. 2007; 129:13633–13645. [PubMed: 17935329]

23. Arad D, Langridge R, Kollman PA. J Am Chem Soc. 1990; 112:491–502.

24. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD Jr. J Chem Theory Comput. 2012; 8:3257–3273. [PubMed: 23341755]

25. Awoonor-Williams E, Rowley CN. J Chem Theory Comput. 2016; 12:4662–73. [PubMed: 27541839]

26. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr. J Comput Chem. 2010; 31:671–90. [PubMed: 19575467]

27. Yu W, He X, Vanommeslaeghe K, MacKerell AD Jr. J Comput Chem. 2012; 33:2451–68. [PubMed: 22821581]

28. Mayne CG, Saam J, Schulten K, Tajkhorshid E, Gumbart JC. J Comput Chem. 2013; 34:2757–2770. [PubMed: 24000174]

29. Frisch, MJ., Trucks, GW., Schlegel, HB., Scuseria, GE., Robb, MA., Cheeseman, JR., Scalmani, G., Barone, V., Petersson, GA., Nakatsuji, H., Li, X., Caricato, M., Marenich, AV., Bloino, J., Janesko, BG., Gomperts, R., Mennucci, B., Hratchian, HP., Ortiz, JV., Izmaylov, AF., Sonnenberg, JL., Williams, Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, VG., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, JA., Jr, Peralta, JE., Ogliaro, F., Bearpark, MJ., Heyd, JJ., Brothers, EN., Kudin, KN., Staroverov, VN., Keith, TA., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, AP., Burant, JC., Iyengar, SS., Tomasi, J., Cossi, M., Millam, JM.,

Klene, M., Adamo, C., Cammi, R., Ochterski, JW., Martin, RL., Morokuma, K., Farkas, O., Foresman, JB., Fox, DJ. Gaussian 09, Revision E 01. Wallingford, CT: 2009.

30. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, Wei S, Buckner J, Jeong JC, Qi Y, Jo S, Pande VS, Case DA, Brooks CL 3rd, Mac Kerell AD Jr, Klauda JB, Im W. J Chem Theory Comput. 2016; 12:405–13. [PubMed: 26631602]

31. Jo S, Kim T, Iyer VG, Im W. J Comput Chem. 2008; 29:1859–65. [PubMed: 18351591]

32. James CP, Rosemary B, Wei W, James G, Emad T, Elizabeth V, Christophe C, Robert DS, Laxmikant K, Klaus S. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]

33. Adelman S, Garrison B. J Chem Phys. 1976; 65:3751–3761.

34. Hoover WG. Phys Rev A. 1985; 31:1695–1697.

35. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. J Chem Phys. 1995; 103(19):8577–8593.

36. Ryckaert JP, Ciccotti G, Berendsen HJ. J Comput Phys. 1977; 23:327–341.

37. Case, DA., Betz, RM., Botello-Smith, W., Cerutti, DS., Cheatham, TE., I, Darden, TA., Duke, RE., Giese, TJ., Gohlke, H., Goetz, AW., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, TS., Le Grand, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, KM., Monard, G., Nguyen, H., Nguyen, HT., Omelyan, I., Onufriev, A., Roe, DR., Roitberg, A., Sagui, C., Simmerling, CL., Swails, J., Walker, RC., Wang, J., Wolf, RM., Wu, X., Xiao, L., York, DM., Kollman, PA. AMBER 2016. University of California; San Francisco: 2016.

38. Jiang W, Phillips JC, Huang L, Fajer M, Meng Y, Gumbart JC, Luo Y, Schulten K, Roux B. Comput Phys Commun. 2014; 185:908–916. [PubMed: 24944348]

39. Weeks JD, Chandler D, Andersen HC. J Chem Phys. 1971; 54:5237–5247.

40. Wang J, Deng Y, Roux B. Biophys J. 2006; 91:2798–2814. [PubMed: 16844742]

41. Deng Y, Roux B. J Chem Theory Comput. 2006; 2:1255–1273. [PubMed: 26626834]

42. Deng Y, Roux B. J Phys Chem. 2004; 108:16567–16576.

43. Lin YL, Meng Y, Huang L, Roux B. J Am Chem Soc. 2014; 136:14753–62. [PubMed: 25243930]

44. Lin YL, Meng Y, Jiang W, Roux B. Proc Natl Acad Sci U S A. 2013; 110:1664–9. [PubMed: 23319661]

45. Alsamarah A, LaCuran AE, Oelschlaeger P, Hao JJ, Luo Y. Plos One. 2015; 10:e0132221. [PubMed: 26133550]

46. Liu Y, Stoll VS, Richardson PL, Saldivar A, Klaus JL, Molla A, Kohlbrenner W, Kati WM. Arch Biochem Biophys. 2004; 421:207–16. [PubMed: 14984200]

47. Narjes F, Brunetti M, Colarusso S, Gerlach B, Koch U, Biasiol G, Fattori D, De Francesco R, Matassa VG, Steinkuhler C. Biochemistry. 2000; 39:1849–61. [PubMed: 10677236]

48. Dill, Ken, Bromberg, S. Molecular Driving Forces. Garland Science; New York: 2010.

49. Marcus RA. J Phys Chem. 1968; 72:891.

50. Koeppl GW, Kresge AJ. Journal of the Chemical Society, Chemical Communications. 1973:371–373.
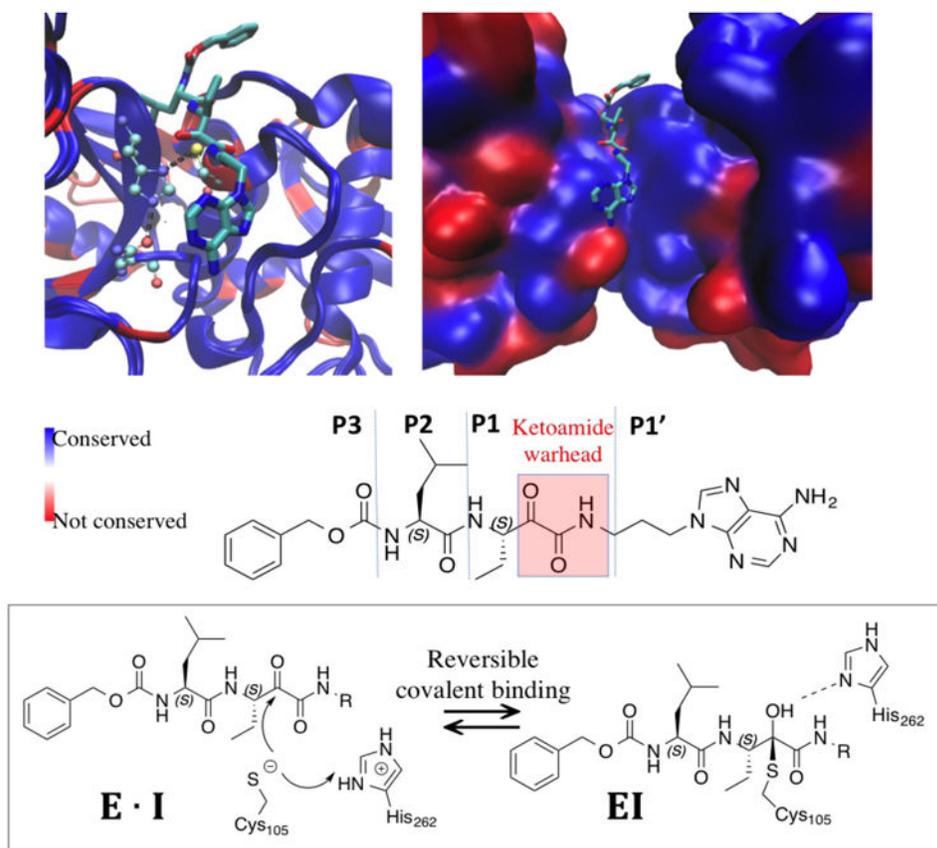
Figure 1. Binding pose of a ketoamide ligand to calpain-1 catalytic site (PDB 2R9C)
Active conformations of calpain-1 and calpain-2 (PDB 3BOW) backbone are aligned. Blue
color indicates identical residues and red color indicates non-identical residues between
calpain-1 and calpain-2. **Left:** protein backbone are shown in new cartoon mode with ligand
shown in licorice and catalytic triad C105/H262/N286 (calpain2 numbering) shown in CPK
mode in atom color code: cyan carbon, blue nitrogen, and red oxygen. All hydrogen atoms
are omitted for clarity. **Right:** protein backbones are shown in surface mode with the same
color code. **Bottom:** the structure and nomenclature of the ketoamide ligand, as well as the
reversible covalent binding scheme.
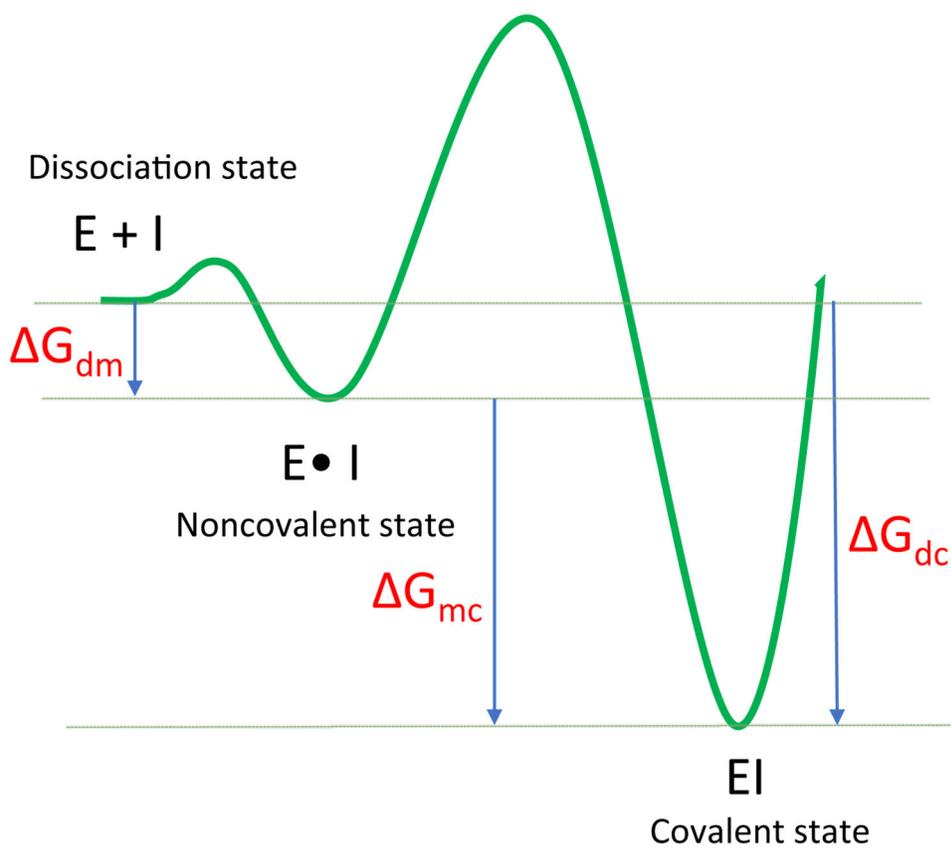
Figure 2. Simplified free energy profile of two-state binding for a reversible covalent binder
Equations 1-3 show the relation between the association constant, $1/K_d$, and the binding free energy of the covalent (ΔG_{dc}), and noncovalent complexes (ΔG_{dm}).
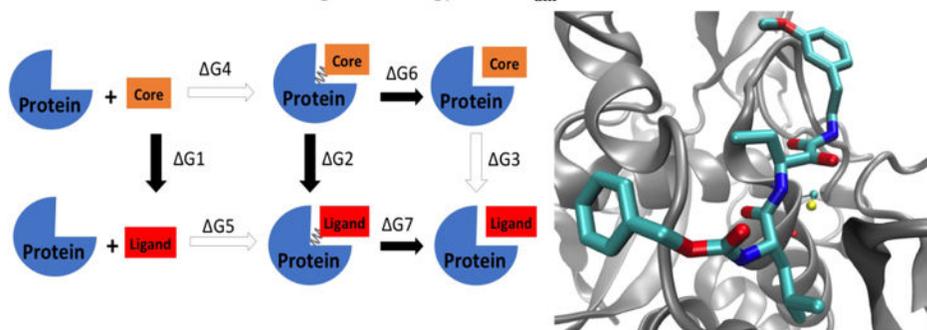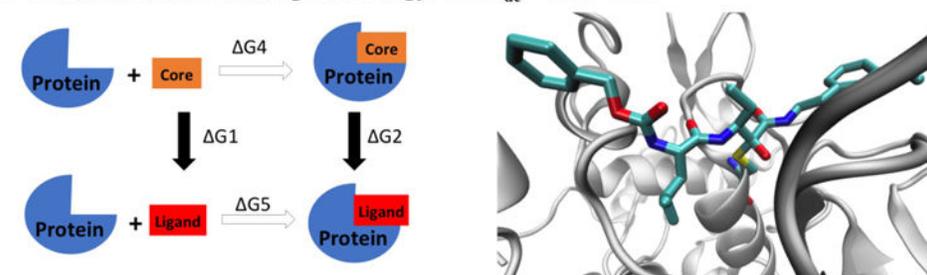
Figure 3. Thermodynamics cycle for calculating the relative binding free energy of the noncovalent (a) and covalent (b) states

The structure on the right shows the scaffold of the binding complexes used for FEP/λ-REMD simulations. Protein backbones are shown in silver new cartoon mode. In noncovalent complex, the catalytic cysteine is shown in CPK mode with atom color code: yellow sulfur, red oxygen, blue nitrogen, and cyan carbon. Ligand 4 is shown in licorice mode with the same color code. In covalent complex, the catalytic cysteine is covalently linked to the ligand warhead. All hydrogen atoms are omitted for clarity.
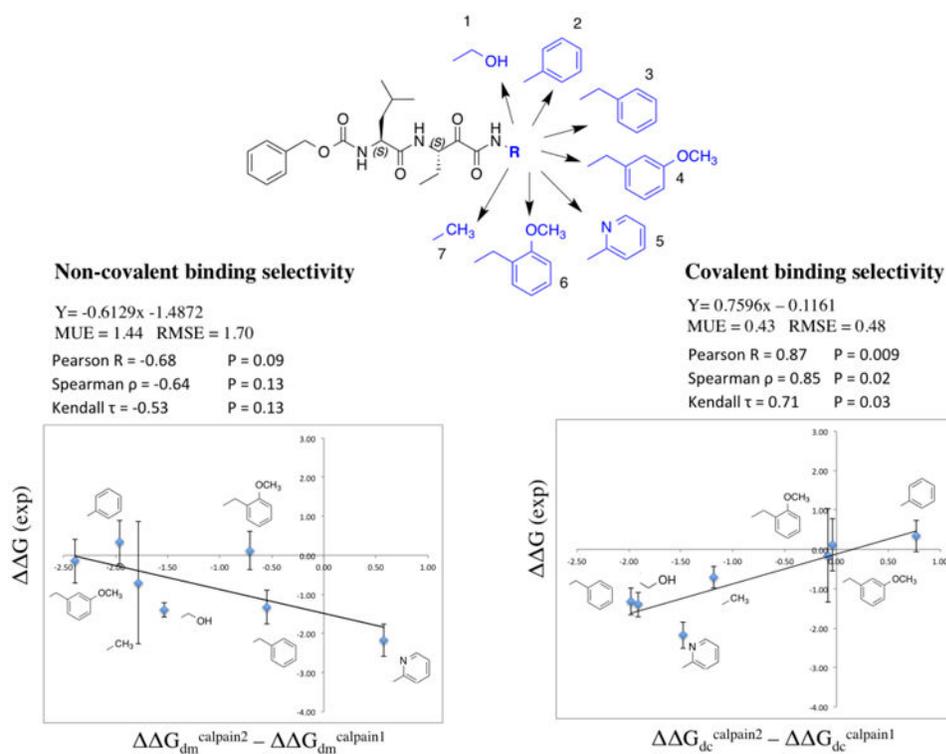
Figure 4. Correlations between experimental and calculated selectivity

Chemical structures of the common core shown in black and the functional groups of the seven compounds shown in blue are drawn at the top. All free energy data, MEU and RMSE are in kcal/mol unit. Experimental selectivity data were obtained using $\Delta\Delta G$ (exp) = RT ln($K_i^{Calpain2}$/ $K_i^{Calpain1}$), in which $K_i^{Calpain2}$ and $K_i^{Calpain1}$ are the inhibitory constants for each compound in calpain-2 and calpain-1 from ref [15]. Calculated selectivity data were obtained using $\Delta G^{calpain2}$ – $\Delta G^{calpain1}$, in which ΔGs are the relative binding free energy difference between the core structure and each ligand in calpain-1 or calpain-2, attained via the FEP calculations. The correlation with noncovalent binding state is shown on the left, and that with covalent binding state is shown on the right. The error bars of $\Delta G^{capain2}$ – $\Delta G^{capain1}$ are calculated using $\sqrt{SD(x)^2 + SD(y)^2 - Cov(x,y)}$, in which x is $\Delta G^{capain2}$ and y is $\Delta G^{capain1}$ from the last three data points. Each data point was calculated from 200 ps/replica. SD is standard deviation and Cov is covariance (see Table S2 for data table).
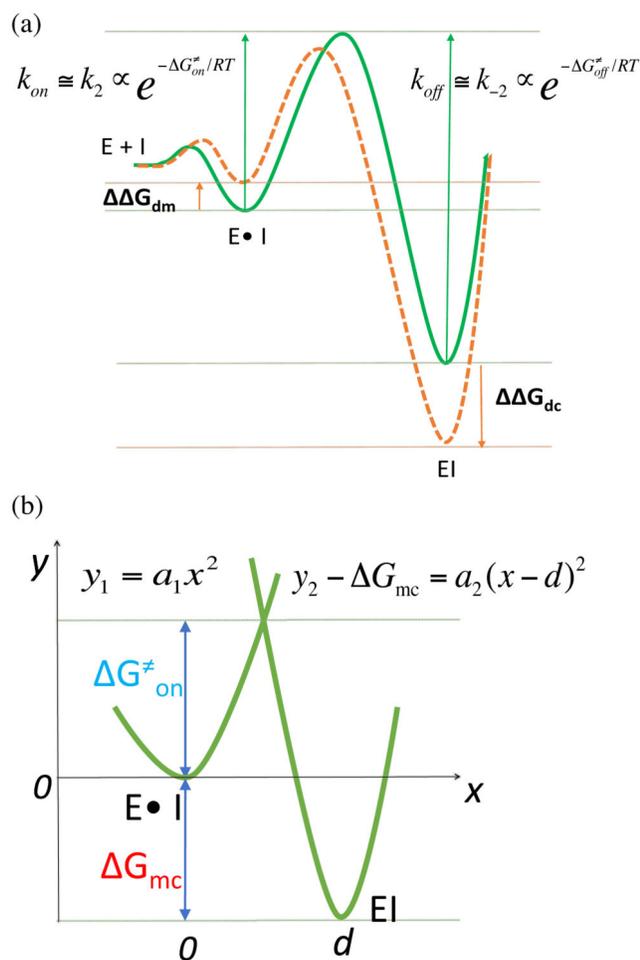
Figure 5. Relationship between the activation energy and the binding free energy
(a) Hypothetical scenarios to reduce ΔG $_{on}$: destabilizing the noncovalent-bound state and stabilizing the covalent-bound state. The green curve represents the free energy profile of a common core structure. The orange curve represents the free energy profile of an individual ligand. (b) The two-state potential surfaces are represented by two parabolas with different shapes.