Published in final edited form as:
Mol Biosyst. 2014 July 3; 10(7): 1890–1904. doi:10.1039/c3mb70620f.

Computational Delineation of Tyrosyl-Substrate Recognition and Catalytic Landscapes in the Epidermal Growth Factor Receptor Tyrosine Kinase Domain

Yingting Liu[1] and **Ravi Radhakrishnan**[1,2,3,*]

[1]Department of Bioengineering, University of Pennsylvania, 240 Skirkanich, 210 S. 33rd Street, Philadelphia, PA 19104, USA

[2]Department of Chemical and Biomolecular Engineering, University of Pennsylvania, 240 Skirkanich, 210 S. 33rd Street, Philadelphia, PA 19104, USA

[3]Department of Biochemistry and Biophysics, University of Pennsylvania, 240 Skirkanich, 210 S. 33rd Street, Philadelphia, PA 19104, USA

Abstract

The epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase (RTK), which catalyzes protein phosphorylation reactions by transferring the γ-phosphoryl group from an ATP molecule to the hydroxyl group of tyrosine residues in protein substrates. EGFR is an important drug target in the treatment of cancers and a better understanding of the receptor function is critical to discern cancer mechanisms. We employ a suite of molecular simulation methods to explore the mechanism for substrate recognition and to delineate the catalytic landscape of the phosphoryl transfer reaction. Based on our results, we propose that a highly conserved region, corresponding to Val852-Pro853-Ile854-Lys855-Trp856 in the EGFR tyrosine kinase domain (TKD) is essential for substrate binding. We also provide a possible explanation for the established experimental observation that protein tyrosine kinases (including EGFR) select substrates with a glutamic acid at the P−1 position and a large hydrophobic amino acid at the P+1 position. Furthermore, our mixed quantum mechanics/molecular mechanics (QM/MM) simulations show that the EGFR protein kinase favors the dissociative mechanism, although an alternative channel through the formation of an associative transition state is also possible. Our simulations establish some key molecular rules in operation for substrate-recognition and for phosphoryl transfer in the EGFR TKD.

Keywords

Molecular Dynamics; Molecular Docking; Quantum Mechanics Molecular Mechanics; Phosphoryl Transfer; Molecular Mechanics Poisson-Boltzmann Surface Area

*Address correspondence to: rradhak@seas.upenn.edu.

Introduction

The epidermal growth factor receptor (EGFR) is one of the four members of the ErbB/HER family receptor tyrosine kinases (RTKs), which collectively regulate essential cellular functions, including proliferation, survival, migration and differentiation [1]. Each member of the ErbB family of RTKs is composed of an extracellular domain, a transmembrane segment, an intracellular protein tyrosine kinase domain, and a cytoplasmic tail (or C-terminal segment) harboring tyrosine phosphorylation sites. Under physiological conditions, ligand binding promotes homo- or heterodimerization of the receptors and activation of their cytoplasmic domains. Dimerization results in auto- or trans-phosphorylation of tyrosine residues in the RTK C-terminal segments, which serve as docking sites for signaling molecules containing SH2 or PTB domains [1]. Aberrant activation of the ErbB network due to the deregulation of the kinase domain (either due to over expression or due to mutations) is frequently associated with cellular transformation and clinical malignancies such as lung, gastric, and breast cancers. Hence, ErbB RTKs in general, and EGFR in particular have been identified as important drug targets in the treatment of cancer [2, 3]. Small molecule inhibitors of EGFR, namely, gefitinib, erlotinib, and lapatinib, are canonical examples of RTK inhibitors in targeted cancer therapy.

The deregulation of RTK signaling in human cancers can occur through several mechanisms: (1) increased ligand production through enhanced local autocrine activation; (2) specific gene translocations to produce kinase-domain fusions with altered signaling profiles; (3) RTK overexpression at the cell surface; (4) mutation of the RTK protein to modulate (enhance) activity; and (5) deregulation of phosphatase and endocytosis mechanisms to increase RTK signal propagation. The results form Catalog of Somatic Mutations In Cancer (COSMIC) has identified the prevalence of EGFR TKD mutations throughout many human cancers. In EGFR, the most common mutation is L858R, (sometimes also called L834R due to an alternate numbering convention which ignores the presence of a cleaved signal peptide). Even though this mutation accounts for over half of the mutations that are listed in COSMIC for EGFR in lung cancer, there are other mutations (DEL, S768I, L861Q, E709G, G719S, L833V, T790M) clinically related to enhanced drug-response [4–6]. However, it is rather striking to note that more than half of the residues in the kinase domain have been observed to be mutated in lung cancer samples; to be precise, 343 different mutations in 153 unique amino acid positions in the EGFR TKD has been clinically documented till date. Yet, only the eight mutations (L858R, DEL, S768I, L861Q, E709G, G719S, L833V, T790M) have been mechanistically studied for kinase activity or inhibition. Since the clinical data clearly indicate that the therapeutic response depends on the type of mutation, there is an imminent need to characterize the remaining mutations from a clinical perspective in order to determine the optimal treatment strategy for cancer patients harboring these other (i.e. outside of the eight) mutations. The recent availability of several crystal structures of the EGFR kinase domain [7–10] have brought to focus the importance of molecular details in the structure-function relationship, including the mechanisms for substrate recognition and phosphoryl transfer.

Tyrosine kinases catalyze protein phosphorylation reactions by transferring the γ-phosphate from an ATP molecule to the hydroxyl group of tyrosine residues in protein substrates [11].

Tyrosine kinase domains (TKDs) share a highly conserved two-lobed catalytic domain [11–13] implying a conserved catalytic mechanism [11–13]. The ATP binds into the cleft between the N- and the C- lobes and typically, two divalent metal (magnesium) ions are thought to be involved in the catalytic reaction [11, 12, 14, 15] along with a number of highly conserved regions in the catalytic domain, including the glycine-rich loop, the DFG motif of the activation loop as well as the catalytic loop. Collectively, they serve to position the phosphate of ATP and the two Mg^{2+} ions. Then, by recognizing [16, 17], recruiting, and phosphorylating a variety of peptide substrates, the kinase regulates several essential signaling pathways. Even though a complete elucidation of specificity in tyrosine phosphorylation by TKDs continues to be elusive [18], it is believed that the specificity of signaling cascades are in part determined by the sequence identity of phospho-tyrosines on substrates [19]. The specificity, in part, depends on how the kinase recognizes the local sequence of amino acids proximal to the binding site (shown in Figure 1) [20, 21]; the nomenclature for these resides are: P0 denotes the tyrosine, P−1, P−2, P−3, etc. are the amino acid labels in sequence going toward the N-terminus from the P0 site, and P+1, P+2, P+3, etc. are the amino acids in sequence going toward the C-terminus from the P0 site. It has been suggested that receptor tyrosine kinases including EGFR TKD show preference for substrates with glutamic acid at the P−1 position and large hydrophobic amino acids at the P +1 position [22–25]. The precise relationship of such a sequence identity to substrate specificity and the catalytic mechanism is still not understood.

Experimental and theoretical studies have demonstrated that, in phosphoryl-transfer catalyzed by kinases and polymerases, the nucleophilic attack on the target phosphate proceeds via a conformation that resembles a trigonal-bipyramidal transition state [26, 27]. However, the phosphoryl-transfer itself can occur through either an associative or a dissociative mechanism [14, 26, 28, 29]. In the dissociative mechanism, the cleavage of the bond between the phosphorus center and the leaving group oxygen precedes the bond formation with the attacking nucleophile. On the contrary, in the associative mechanism, the bond formation between the attacking nucleophile and the phosphorus center precedes the bond cleavage. Although it is well accepted that the dissociative mechanism is the preferred mechanism for the reaction in solution for a monoester [30] (which is also the case most relevant for the reaction catalyzed by TKDs), discerning the transition state of protein kinase catalyzed reaction is complex, resulting in controversial experimental interpretations [14, 31–34]. Besides, it is also important to understand the catalytic roles of specific residues around catalytic site of a protein kinase, in particular, the role of the conserved aspartic acid residue (D813 in EGFR) in the catalytic loop [14, 35, 36]. To address this question in other well-characterized kinase systems such as PKA, CDK2, insulin kinase, and MAPK, computational methods have been successfully used to investigate the phosphoryl transfer reaction [28, 29, 37–42]. The challenge in extending such studies to EGFR kinase is that, while the structure for EGFR crystallized with a bi-substrate (ATP-peptide analog) is available [8], providing molecular resolution to the ternary complex of the kinase/substrate/ATP system, the covalent constraint between the ATP analog and the substrate, and the fact that the peptide sequence in the bi-substrate ligand is a c-Src substrate [8, 43] rather than a native EGFR substrate, introduces an important challenge in resolving the structural details of the catalytically productive ternary complex [44]. This brings to focus a

related but equally important question on how EGFR kinase recognizes its native peptide substrates.

In this article, we employ methods of molecular modeling and simulation to: (1) investigate how the EGFR TKD recognizes the local amino acid sequence of substrates proximal to the binding site, and (2) characterize the EGFR TKD-catalyzed mechanism of phosphoryl transfer. Specifically, to address (1) we employ a computational protocol (see Figure 2) combining molecular dynamics simulations, ensemble molecular docking [45, 46], and Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) [47–53] to investigate the bound conformations of peptide substrates derived from two of the prominent autophosphorylation sites of the cytoplasmic tail region of the EGFR TKD [23], namely Y1068 (VPEYINQ) and Y1173 (NAEYLRV). We find that in our predicted bound conformations, the substrates bound into the same pocket as the bi-substrate peptide and that the P−1 glutamic acid residue and the hydrophobic amino acid at the P+1 position situate to interact with a highly conserved motif in the protein kinase family, namely Val852-Pro853-Ile854-Lys855-Trp856 in EGFR. This motif creates a binding surface to form hydrophobic contacts as well as hydrogen bonds to residues close to the tyrosine in the substrate and therefore helps to orient the tyrosine residue towards the active site. Using our predicted binary complex of EGFR TKD/substrate, we modeled the ternary complex structure of EGFR TKD/substrate/ATP and two Mg^{2+} ions, and addressed the phosphoryl transfer mechanisms in EGFR TKD using mixed quantum mechanics/molecular mechanics (QM/MM) simulations. Our QM/MM simulations showed that the EGFR protein kinase favors the dissociative mechanism, although an alternative channel through the formation of an associative transition state is also possible. Our potential energy scans suggested that the kinase directs the reaction through the dissociative channel primarily using a hydrogen bond between the conserved aspartic acid residue in the catalytic loop and the hydroxyl group of the substrate. The alternative channel can sustain weak activity, but can become significant in non-cognate environments where the catalytic aspartic acid is mutated or the enzyme is locked in a constitutively inactive conformation. Collectively, our simulations establish some key molecular mechanisms for the substrate binding and the phosphoryl transfer in the EGFR TKD.

Results

Tyrosyl-peptide substrate recognition by EGFR TKD

Following the computational protocols described in Figure 2 and the methods and materials section, 500 bound conformations were predicted using the ensemble docking protocol for binding of the EGFR TKD to a seven-residue native EGFR peptide with P0 corresponding to the autophosphorylation site Y1068 (VPEYINQ). Since a large number of the resulting structures showed peptide conformations in which the tyrosine residue was far from the active site, we constructed a histogram of D (the distance between the tyrosyl-oxygen and the $O\delta_2$ oxygen of the conserved catalytic aspartic acid residue D813, see Figure 3A), and only considered those conformations with D < 6 Å; this choice for the cut-off is intended to cover situations where the hydroxyl group of the peptide forms a hydrogen bond with $O\delta_2$ of D813, either directly, or through a water mediated interaction. As an additional filter,

conformations where the peptide substrate overlaps with the position of ATP (quantified using the bi-substrate conformation in the PDBID:2GS6 structure) were also discarded. Thus, among the 500 predicted conformations, 18 structures of the TKD-substrate complex satisfied our constraints, which we categorized into 5 clusters based on root-mean-squared deviations (RMSD) of the backbone heavy atoms.

The five conformations with the lowest score in each cluster are depicted in Figure 3B, along with the bi-substrate conformation from the PDBID:2GS6 structure. The computed MMPBSA binding free energies ΔGs (see methods and materials section) of the five conformations are provided in Table 1. As would be expected, the binding free energy is dominated by the difference between the unfavorable polar solvation and the favorable molecular mechanics terms. Interestingly, the conformation 3, which corresponds to the lowest MMPBSA score, also aligns the closest to the bi-substrate conformation as shown in Figure 3B and Table 1. A 9 ns MD simulation (see methods and materials section) performed using conformation 3 in the presence of ATP and two Mg^{2+} ions reached steady state after 3 ns of simulation (Figure 4). The peptide backbone is stable during the entire simulation as monitored by the backbone RMSD relative to the initial conformation (Figure 4). Therefore, our predicted substrate conformation emerges to be in close alignment with the bi-substrate structure in PDBID:2GS6 structure as shown in Figure 2B. The side-chain residues, however adjust conformations to result in stable hydrogen bonds and hydrophobic interactions in the final predicted bound structure of the ternary complex (EGFR TKD/tyrosyl-substrate/ATP) as shown in Figure 3C,D.

The predicted conformation (Figure 3C,D) provides new insights into the substrate recognition mechanism of EGFR TKD: (1) with respect to the P0 tyrosine, the C-terminus region of the peptide is anchored into a binding pocket formed by residues in the C-terminus region of the EGFR activation loop, and the N-terminus of the peptide extends to the front of the C-lobe of EGFR TKD (Figure 3C); (2) The hydroxyl oxygen of the P0 tyrosine is 2.9 Å away from the $O\delta_2$ oxygen atom of residue D813, i.e. within a hydrogen bonding distance (Figure 3D); (3) the P−1 glutamic acid residue forms a salt-bridge with a highly conserved lysine residue (K855); and (4) the isoleucine in the P+1 position situates in a hydrophobic pocket formed by Val852, Pro853, Ile854, and Trp856 (Figure 3D).

To further validate the robustness of this predicted mode of tyrosyl-substrate interaction, we also predicted the bound conformation of a different EGFR-native peptide sequence (NAEYLRV) derived from the auto-phosphorylation site Y1173 of EGFR. 500 bound conformations were predicted using the ensemble docking protocol for binding of EGFR TKD peptide Y1173. 22 structures of the TKD-substrate complex were found to satisfy our constraint of D < 6 Å, which we categorized into 6 clusters based on root-mean-squared deviations (RMSD) of the backbone heavy atoms, shown in Figure 5A. The computed MMPBSA binding free energies ΔGs of the six conformations are provided in Table 2. The conformation 5, which corresponds to the lowest MMPBSA score, also aligns the closest to the bi-substrate conformation as shown in Figure 5A and Table 2. A MD simulation performed using conformation 5 in the presence of ATP and two Mg^{2+} ions reached steady state, resulting in stable hydrogen bonds and hydrophobic interactions yielding a final predicted bound structure of the ternary complex of the EGFR TKD/tyrosyl-substrate/ATP.

The conformation predicted for Y1173 is very similar to what we describe for Y1068 (shown in Figure 5, see also Table 2). In particular, the hydrogen bond between the hydroxyl of tyrosine and D813, the salt-bridge between E(P−1) and K855, and the hydrophobic contact between L(P+1) and VAL852-Pro853-Ile854 are all preserved. However, we note that the best MMPBSA score for the Y1173 peptide is lower than that for Y1068, implying that the binding affinity of the Y1173 peptide to the active state of the kinase is lower than that for the Y1068, as predicted by the calculations. This is consistent with the observations of Gibbes-Johnson et. al., who report that the experimental binding affinity of the Y1173 peptide is indeed lower than that of the Y1068 peptide in EGF-stimulated receptors [54]. Hence, while the salient features for peptide substrate recognition might involve just the P −1, P0, and P+1 sites, we note that interactions with more distal sites can be important in fine tuning the overall binding affinity of the peptide substrate to the kinase. Indeed, it has been experimentally known that a non-native substrate (called the optimal substrate in the literature) for EGFR TKD with sequence EEEYFEL has a binding affinity that is 8-fold stronger than that for the native Y1068 substrate [54].

Pre-catalytic conformational landscape orchestrates multiple reaction channels

During the course of the MD simulation of the ternary complex involving the Y1068 peptide (Figure 4), the conformational equilibration of ATP and the two Mg^{2+} ions relative to the substrate tyrosine resulted in small rearrangements of the phosphate tail of ATP within the active site. Hence, we chose to explore the free energy landscape for the pre-organization of the nucleophilic attack distance λ_a (Y(P0):Oη - ATP:P_γ distance) as this distance is most relevant to catalysis as well as the most sensitive to the phosphate tail conformation, see Figure 6. Conformation 1 in the figure 6 is obtained from energy minimization of our model structure, $\lambda_a \approx 4$ Å. The system settles into another metastable state with $\lambda_a < 3$ Å (conformation 2 in figure 6), surpassing a nominal free energy barrier of 3.5 k_BT (2 Kcal/mol). This subtle pre-catalytic conformational transition corresponds to a conformational rearrangement of the γ-phosphate of ATP, see discussion below.

An analysis of the conformations within the window corresponding to conformation 2 revealed that that the proton of the tyrosine-hydroxyl (Hη) is primed for transfer to two possible positions owing to two mutually exclusive hydrogen bonds: (1) hydrogen bonded to $O\delta_2$ oxygen of D813 (indicated by a small value of δ_I), and (2) hydrogen bonded to the O1γ oxygen of ATP (indicated by a small value of δ_{II}), see Figure 7. We hypothesize that the bimodal distributions of δ_I and δ_{II} values represented in the normalized histograms in Figure 7A,B serve to select either of two competing catalytic routes, labeled as pathways I and II in Figure 8 for proton transfer and concomitant phosphoryl transfer. We explore this hypothesis further by investigating the possible mechanisms of phosphoryl transfer in EGFR TKD following the protocols for QM/MM minimization and QM/MM MD simulations (see methods and materials section). Our simulations (see Figure 9 and Figure 10) reveal that the phosphoryl transfer via pathway I, in which the proton is abstracted through the catalytic aspartic acid residue, (highlighted in red, see Figure 8), can occur via both the dissociative (see Figure 9A), and associative mechanisms (see Figure 9B), while pathway II (highlighted in turquoise, see Figure 8) supports only the associative mechanism (Figure 9C). In the dissociative reaction channel, the sum of the nucleophilic attack distance λ_a (between Y-OH

and ATP Pγ) and the bond-cleavage distance λ_c (between ATP Pγ and ATP O2/3β) in the transition state is greater than the corresponding sum in the reactant and product states; this scenario is reflected in Figure 9A where the $\lambda_a+\lambda_c$ values for transition state (TS) are greater than those for Reactant (R) or Product (P) states, i.e., the TS region lies above the line joining the R and P states in the λ_a versus λ_c plot. In contrast, a nucleophilic attack through an associative mechanism is evident in Figure 9B where the $\lambda_a+\lambda_c$ values for TS are smaller than those for R or P, i.e., TS lies below the line joining the R and P in the λ_a versus λ_c plot. Pathway II is characterized by the migration of the proton from the hydroxyl group of the tyrosine substrate to the O1γ oxygen of ATP and the subsequent transfer of the proton to the O3β atom of ATP. The phosphoryl transfer takes place concomitant with the proton transfer steps but only through an associative mechanism, see Figure 9C; interestingly, for this channel, our attempts to drive the system through a dissociative mechanism yielded an associative mechanism with the O2β atom of ATP playing the surrogate (as the leaving group) to the O3β atom of ATP.

Depictions of the snapshots highlighting key conformations of the active site along the reaction pathways are provided in a movie format in the accompanying supplementary material. In Figure 9(D–F), we depict the computed energy changes along the reaction pathways in active EGFR TKD, which suggest that for EGFR, the dissociative mechanism of phosphoryl transfer through pathway I corresponds to the lowest activation energy, E_a=16 kcal/mol (Figure 9D), while energy profiles for the associative mechanisms through pathways I and II yield barriers of 24 kcal/mol (Figure 9E) and 26 kcal/mol (Figure 9F), respectively.

Effect of mutation on tyrosyl-peptide substrate recognition

The kinase-activating mutation L834R locates closely to the catalytic site of EGFR TKD. To further investigate whether mutating a hydrophobic residue LEU to a positively charged residue ARG will affect the Y1068 peptide binding affinity, we employed the free energy perturbation (FEP) method to calculate the binding free energy difference between the wildtype and mutant systems. The FEP results (Figure 10 and Table 3) showed that L834R mutation resulted in a binding free energy change (ΔΔF) of 4.26 ± 0.32 Kcal/mol for the Y1068 peptide. Our results imply that the mutation does not increase (but rather decreases) the substrate binding affinity. Though experimental results probing the change in KM for the Y1068 peptide in L834R mutant and wildtype EGFR TKD systems are not directly available, Mulloy et al. has described the overall phosphorylation rate of Y1068 to increase modestly (fold increase in k_{cat} of 2.5 in EGF stimulated cells) [55]. This suggests that the mutation rather relies on a shift in the equilibrium between the inactive and active states of the kinase (by favoring the active state) for its activation mechanism.

Discussion and Conclusion

Crystal structures of multiple TKDs have already been solved, which has led to important insights in understanding the kinase function. However, most of these crystal structures are solved in the presence of ATP analogs and small molecule inhibitors. Due to the relatively low binding affinity of the interaction between TKD and peptide substrates, crystallization of TKD-peptide substrate complexes is still a challenge using experimental techniques. In

this work, we suggest that a rich landscape exists for the binding of tyrosyl-substrate and subsequent pre-catalytic and catalytic events in the ternary complex of EGFR TKD/ATP/tyrosyl-substrate, from which we outline new molecular rules, which may, in part, contribute to specificity in phosphorylation reactions.

Our ensemble molecular docking protocol is successful in predicting bound conformations of a 7-residue peptide sequence derived from the cytoplasmic tail region of EGFR to EGFR TKD, with predicted bound conformation very similar to that found in a crystal structure of a bi-substrate analog co-crystalized with EGFR TKD [8] (PDBID: 2GS6). In our predicted complex, the aspartic residue D813 forms a hydrogen bond with the hydroxyl oxygen the P0 tyrosine of the peptide substrate. The aspartic residue has been shown to hydrogen bond with the hydroxyl group of serine/threonine/tyrosine in other kinase systems [34, 56, 57]. In the only existing EGFR TKD/bi-substrate analog structure (i.e. PDBID: 2GS6), the distance between the aniline nitrogen and the $O\delta_{1/2}$ oxygen of D813 is 6.37 Å, which makes it difficult to ascribe a role for the aspartic residue. The conformation in our modeled ternary complex clearly places the hydroxyl oxygen the P0 tyrosine 2.9 Å away from the $O\delta_2$ oxygen atom D813, i.e., within hydrogen bonding distance, which is consistent with the conceived role of D813 as a facilitator of proton abstraction.

Moreover, it has been suggested that protein tyrosine kinase (including EGFR) select substrates with glutamic acid at the P−1 position and large hydrophobic amino acids at the P+1 position [22–25, 58, 59]. Our predicted peptide bound ternary complex suggests that a highly conserved region, corresponding to Val852-Pro853-Ile854-Lys855-Trp856 in EGFR TKD is essential for substrate binding: Lys855 forms a salt bridge with the highly preferred glutamic acid residue at the P−1 site, and the hydrophobic pocket formed by Val852, Pro853, Ile854 and Trp856 serves to anchor a large hydrophobic residue at P+1 position. This observed binding mode for our top-ranked configuration is consistent with the hypothesis that the substrate-binding surface formed through hydrophobic contacts and hydrogen bonds help orient the tyrosine residue for catalysis [24]. Collectively, this extended region, i.e. Val852-Pro853-Ile854-Lys855-Trp856 in EGFR, has been identified as a highly conserved motif in the protein tyrosine kinase family [44], which further adds significance to our predicted conformation for the ternary complex. This allows us to conclude that a conserved binding surface on the kinase sandwiching the P0 tyrosine and making hydrophobic contact to the P+1 and hydrogen bond with P−1 residues, respectively, helps orient the tyrosine residue for catalysis [24]. Interestingly, the corresponding motif in serine/theronine kinase, with a sequence of Gly200-Thr201-Pro202-Glu203-Tyr204 in PKA, has been named as P+1 binding pocket and shown to be critical for substrate specificity [24, 60, 61]. In this case, the glutamic acid residue (E203) in this motif has been proposed to interact with preferred positively charged residues of the N-terminus site in PKA substrates [24] and residues Gly200, Pro 202, Thr201 and Tyr204 favor a hydrophobic residue in the P+1 position [21]. The similarity with the serine/threonine kinases raises the intriguing question as to whether an evolutionary path can be seen across species from a potential minimal and very inefficient transfer route to the much more efficient and Glu catalyzed route. This question may particularly be pertinent for the very unusual and highly conserved ability of MAP Kinasess to catalyze both, threonine and tyrosine phosphorylation.

Two of the main points of contention in kinase-catalyzed phosphoryl transfer reactions are whether the phosphoryl-transfer reaction occurs through an associative or a dissociative mechanism [14, 26, 28, 29] and whether the conserved aspartic (D813 in EGFR) functions as an acceptor for proton transfer from the hydroxyl group of the substrate. With the predicted peptide bound conformation, we were able to build the ternary structure of EGFR TKD, peptide with presence of ATP and Mg^{2+} ions and investigate the mechanistic details of EGFR TKD catalytic reaction mechanism using QM/MM simulations. Based on a pre-catalytic energy landscape of γ-phosphate of ATP we identify two alternative preferred proton positions of the tyrosine hydroxyl based on which we hypothesize two possible pathways: Pathway I involves proton abstraction from the substrate tyrosine hydroxyl group by D813 of EGFR, and pathway II involves proton migration to the $O1_{\gamma}$ oxygen of ATP. Indeed, our QM/MM simulations support both pathways, with the phosphoryl transfer via pathway I sustained via both dissociative and associative mechanisms, while that via pathway II sustained by just the associative mechanism. The dissociative mechanism through pathway I has the most favorable energy barrier of 16 kcal/mol, while energy profiles for the associative mechanisms through pathways I and II each yield barriers of 24 kcal/mol. Combined with the pre-catalytic reaction energy barrier, our estimate for the net energy barrier $\Delta E_{barrier}$ for activation through the most favorable reaction channel (i.e. for pathway I through dissociative mechanism) in the active EGFR TKD is 16+2=18 kcal/mol. This value is compares favorably with the experimental value for the free energy barrier for EGF activated receptor system in the cellular context for which $\Delta G_{barrier}=k_BTln[k_BT/(k_{cat}h)]$=18.5 kcal/mol estimated from the reported k_{cat} of 17.4 min^{-1} by Gibbes-Johnson et. al. [54]. Our estimate for the net energy barrier $\Delta E_{barrier}$ for activation through pathway II for active EGFR TKD is 24+2=26 kcal/mol. This translates to a k_{cat} of $\leq 10^{-6}$ s^{-1} or a reduction by at least a factor 10^{-6} relative to the experimental k_{cat} of EGF activated receptor or a factor of 10^{-5} relative to experimental k_{cat} in the absence of EGF [62, 63]. The reduction in k_{cat} manifests itself as little or very weak residual activity through pathway II, which is consistent with recent experiments [64, 65].

Given that the protonation states of titratable side-chains in the active site can have considerable effect on the stability of the ground state as well as on the energies of the reaction landscape, we explored the different protonation states of the conserved catalytic residue D813 in EGFR. Multiple protonation states of the conserved catalytic residue (D831 of the DFG motif) were not considered because the location is relatively removed from the atoms involved in the catalytic reaction pathway. Although recent studies of the Abl kinase have reported that the protonation state of this residue is important in the flipping of the DFG motif toward the active conformation [66], such a flip for EGFR is already facilitated, even in the unprotonated state for D831, by kinase dimerization [67]. For the D813-protonated system, the phosphoryl transfer step occurs through an associative mechanism and proceeds with the simultaneous migration of the proton from the hydroxyl group of the tyrosine substrate to the $O\delta_1$ oxygen of D813 and the migration of $H\delta_2$ of the protonated $O\delta_2$ oxygen of D813 to O3γ atom of ATP (data not shown). Our attempts to drive the system through a dissociative mechanism yielded an associative mechanism with the O2β atom of ATP playing the surrogate (as the leaving group) to the O3β atom of ATP. These data imply that protonation of D813 is unlikely prior to the phosphoryl transfer.

ErbB3/HER3 is another member of the human epidermal growth factor receptor (EGFR/HER) or ErbB receptor tyrosine kinase family. ErbB3 is considered to be an inactive pseudokinase because it lacks several key conserved (and catalytically important) kinase domain residues – including the catalytic base aspartate. Indeed using QM/MM simulations we recently reported that phosphoryl-transfer can occur in an inactive conformation of ErbB3 TKD without abstraction of the tyrosine –OH proton by the catalytic base aspartate, which may explain the experimentally observed robust *trans*-autophosphorylation sustained in ErbB3, although this activity is substantially less than that observed in EGFR [68]. Pathway I is not sustainable in ErbB3 owing to the crucial substitution of Asn815 (analogous to Asp 813 in EGFR). However, like active EGFR TKD, the proton migrates to the $O1_{\gamma}$ oxygen of ATP, through pathway II in Figure 8. Phosphoryl-transfer concomitant with this pathway can be catalyzed by ErbB3 in the inactive-like conformation, although it is predicted to be several orders of magnitude slower than the most favorable reaction channel in EGFR (*i.e.* for pathway I through a dissociative mechanism) – consistent with experimental observations. The finding also implies that this mechanism can operate in EGFR further suggesting that mutating D813 in this receptor may not completely abolish kinase activity. Indeed, a D813A-mutated variant of EGFR has been reported to retain its ability to promote EGF-dependent DNA synthesis and MAP kinase activation despite sustaining greatly reduced receptor autophosphorylation [69].

Members of the ErbB family of receptors — the epidermal growth factor receptor (EGFR/ErbB1/HER1), ErbB2 (HER2), ErbB3, and ErbB4 — activate a multi-layered signaling network mediating crucial pathways leading to cell proliferation and differentiation [1], in response to activation of the receptors by the epidermal growth factor (EGF), transforming growth factor-α, and several other related peptide growth factors [1]. Altered expression and mutations of ErbB receptors have been implicated in molecular mechanisms of oncogenesis, drug sensitivity, as well as resistance [67, 70–73]. In particular, altered or downstream signaling in cancer lines harboring ErbB mutations leading to preferential activation of certain downstream pathways over others have been discussed as the basis for the differences in cell phenotypic decisions in such cancer cells [74, 75]. While the molecular basis for such a selective activation of downstream pathways in still unclear, altered trafficking has been discussed as a possible mechanism leading to this state [76, 77]. The results we have presented in this article describe another possible mechanism that leads to preferential downstream signaling. Namely, the identity specific tyrosyl recognition (including the free energy landscape of pre-catalytic conformational rearrangements) at the catalytic site can be altered differentially by ErbB mutations, leading to differences in phospho-tyrosyl kinetics and adaptor binding downstream of receptor activation [67, 70–73]. The molecular mechanisms of tyrosine binding and the catalytic landscape of phosphoryl transfer discussed here, establish a direct link between receptor activation and downstream signaling. The results from our molecular model can help delineate mechanisms of preferential activation and help link them to drug sensitivity and resistance [78, 79].

Methods and Materials

We employ a suite of computational methods including molecular dynamics and molecular docking (see Figure 2), umbrella sampling, and QM/MM simulations to investigate

molecular mechanisms of substrate recognition (i.e. bound conformation of a tyrosyl-peptide) by EGFR TKD and the catalytic mechanism associated with the subsequent phosphoryl transfer reaction.

Molecular dynamics (MD)

All the molecular dynamics simulations in this work were initiated by preparing the protein or protein complex structure by adding hydrogen atoms and missing residues using the CHARMM [80] biomolecular simulation package. The protonation states of histidine (HIS) residues were chosen according to recommendations from the WHATIF web interface (http://swift.cmbi.kun.nl/WIWWWI). Protonation states for all titratable groups distal to the catalytic site were chosen based on their individual pKa values in aqueous solution at a pH 7.0 [81]. The systems were then placed in a water (TIP3P model) box with 15Å padding. The ions were placed randomly to neutralize the systems using the Solvate and the Autoionize Plugins in the VMD package [82]. The systems were minimized for 50000 steps with heavy atoms fixed and further relaxed for another 50000 steps without any restraints. The systems were then gradually heated from 0K to 300K followed by 500 ps equilibrating runs at constant pressure (1 atm.) and temperature (300 K) to allow the system to reach the proper density. A productive trajectory was then generated at 300K and 1atm. The simulations were performed using NAMD with the CHARMM force field [83] with periodic boundaries enforced and long-range electrostatics taken into consideration.

Molecular docking

All the docking simulations in this work were performed with AUTODOCK 3.05 [84]. Grid maps were constructed as 126×126×126 points with a grid spacing of 0.184 Å to cover a pre-defined binding pocket. The Lamarckian Genetic Algorithm (LGA) [84] was applied to explore the conformational space of the substrate. In each docking run, the initial population was set to 100 individuals, the maximum number of energy evaluations was set to 10^8, and the generation of GA run was set to 30,000.

Docking of substrate tyrosyl-peptides Y1068 and Y1173 to EGFR TKD

The active conformation of the wildtype EGFR TKD domain was modeled based on the 2GS6 structure from the Protein Data Bank [8]. A 10 ns dynamics trajectory was generated by molecular dynamics simulation using NAMD (see above) [85] and an ensemble of 100 protein conformations was derived by sampling at uniform intervals from the trajectory. Two 7-residue peptides derived from the cytoplasmic tail region of EGFR (corresponding to Y1068: namely, VPEYINQ and to Y1173: namely, NAEYLRV) were docked to each snapshot in the 100-member ensemble by molecular docking using Autodock 3.05 (see above) [84]. 5 possible bound conformations were predicted for each protein conformation, which in total yielded 500 bound conformations of tyrosyl-peptide/EGFR TKD for each peptide.

In order to further sort the binding conformations generated by the ensemble docking protocol, we employed Molecular Mechanics Possion Boltzmann Surface Area (MMPBSA) [48–53, 86] to calculate the binding free energy. For each of the predicted bound conformations, a 1 ns trajectory was generated using molecular dynamics. The MMPBSA

energy of each modeled complex was calculated as the average of the single point MMPBSA scores of 500 snapshots derived from the 1 ns simulation; the water molecules and mobile ions were removed before the MMPBSA score calculation. In the MMPBSA calculation, the molecular mechanics energy, U_{MM}, was evaluated using an infinite cutoff for non-bonded interactions. The electrostatics contribution to the solvation free energy, W_{PB}, was calcuated with Poisson-Boltzmann solver in CHARMM [87]. The reference system was assigned a dielectric constant of 1 and 0 M salt concentration and the solvated system was assigned a dielectric constant of 80.0 and 100 mM salt concentration. The nonpolar contribution to the solvation free energy, W_{SA}, was approximated with the suface area model ΔW_{SA}=[0.00542 kcal/mol/Å^2]×SASA +0.92 kcal/mol [88], where the molecular surface area (SA) was estimated with a 1.4 Å slovent probe radius. The translational and rotional entropy loss upon binding was estimated based on the fluctuations of the center of mass and that of orientations (Euler angles), by following the procedure described by Swanson et. al. [49]. Other (vibrational) entropy changes were neglected in this work.

The top ranked conformation of peptide Y1068 with EGFR TKD identified based on the MMPBSA score (Table 1) was chosen as the initial bound structure for further analysis of the catalytic mechanism. The ATP was placed in homology with the bi-substrate structure (PDB ID: 2GS6) [8] and two Mg^{2+} ions were placed in homology with those resolved in the crystal structure of PKA (PDB ID:1ATP) [37]. A 10 ns molecular dynamics simulation was performed on the ternary complex of ATP/substrate-Y/EGFR TKD.

Free energy perturbation (FEP)

To calculate the binding affinity change of EGFR TKD and its peptide upon the clinical mutation L834R, we designed a thermodynamic cycle shown in Figure 11. Based on this cycle, the binding free energy change $\Delta\Delta G_{W\to M}$ is related to the free energy changes of the mutation in the bound state (ΔG_B) and the unbound state (ΔG_U):

$$\Delta\Delta G_{W\to M}=\Delta G_M-\Delta G_W=\Delta G_U-\Delta G_B$$

The free energy changes due to the particular mutation in the bound state (ΔG_B) and the unbound state (ΔG_U) were calculated by using the free energy perturbation (FEP) method. The free energy procedure was performed using the alchemical FEP method in NAMD 2.7b2 with the dual-topology construction. For each state (bound or unbound), we performed FEP calculations in both the forward and backward directions to ensure convergence, and to estimate the error. For each direction, the perturbation was divided into 72 windows (λ= 0, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.005, 0.01, 0.015, 0.02, 0.03 to 0.1 with an interval of 0.01, 0.1 to 0.9 with a interval of 0.02, 0.9 to 0.98 with an interval of 0.01, 0.098, 0.0985, 0.099, 0.995, 0.999, 0.9999, 0.99999, 0.999999, 1). In each window, the system was equilibrated for 20 ps and run for another 100 ps for data collection. Larger window sizes and longer simulations have also been tested and the current setup gives us a reasonable convergence in the final binding affinity. To avoid the “end-point catastrophes”, the soft-core potential was used to gradually scale the unbounded interaction potential. The van der Waals and electrostatic interactions were scaled separately; that is, the vdW interactions were linearly coupled to the simulation from λ=0 (fully decoupled) to λ=1 (fully coupled) and the

electrostatic interactions were coupled into the simulations over the range of λ=0.5 to λ=1. (For the vanishing particles, the vdW interactions were linearly decoupled from the simulation over the value range of 0 to 1, and the electrostatic interactions were decreased gradually from λ=0 to λ=0.5).

Sampling the pre-catalytic free energy landscape

Since the positions and orientations of the ions and ATP relative to the substrate tyrosines in our initial modeled structures were determined either from a bisubstrate ligand complex or by homology with protein kinase A (see above), prior to QM/MM simulations, we explored the classical free energy landscape along an a priori-chosen reaction coordinate χ_1 (the tyrosyl O-ATP P_γ nucleophilic attack distance). For computing the free energy landscape, the probability distribution $P(\chi_1)$ is calculated by dividing χ_1 into several (6–7) windows. The histograms for each window were collected by harvesting classical molecular dynamics trajectories of length 200 ps per window using a harmonic restraint $0.5\times K\times(\chi_1-\chi_1^{i,0})^2$, $\chi_1^{i,0}$ being the pre-set of the reaction coordinate characterizing the i^{th} sampling window, and K=20 kcal/mol/Å^2 being the force constant for the restraint along χ_1. The sampling data from the different windows were then processed using the weighted histogram analysis method (WHAM) [89].

Quantum mechanics molecular mechanics (QMMM) simulations

The QM/MM approach we adopted was based on an existing interface between GAMESS-UK [90] and CHARMM [80]. For the quantum region, we chose the two Mg^{2+} ions, water molecules within 5 Å of the Mg^{2+} ions, segments of the ATP, the tyrosine residue of the substrate, and two catalytic residues (D813 and D831), yielding a total of 60 atoms. In order to reliably handle the phosphorous chemistry, the quantum region was treated using density functional theory with a B3LYP exchange correlation functional and a 6-31G* basis set [91], and the remaining protein and solvent molecules were treated classically using the CHARMM27 force field. Since the QM/MM boundary cuts across covalent bonds, we applied the single link atom procedure [92] to satisfy the valences of broken bonds in the QM region. The electrostatic interactions between the quantum region and the classical atoms belonging to groups at the QM/MM boundary were discarded as this choice has been shown to enhance the accuracy in the resulting energies and geometries [93]. We have extensively explored this choice of the QM/MM region (including sensitivity to the size of the QM region, and the functional/basis set combination) in our prior studies of closely related systems [94–96]; others have validated the choice of the link atom [93]. Non-bonded van der Waals (VDW) interactions were treated by applying a switching function at 10 Å and truncating the VDW potential energy smoothly at a cutoff distance of 12 Å. The forces in the quantum region were calculated on-the-fly, assuming that the system moves on the Born-Oppenheimer surface defined by the QM/MM Hamiltonian. The system was first subjected to 1200 steps of the adopted basis Newton-Raphson (ABNR) minimization and subsequently QM/MM MD simulations were performed during which the systems were first heated to 300 K and then subjected to constant temperature of 300 K dynamics using a Langevin thermostat for 10 ps using a standard 1 fs time step of integration.

Catalytic reaction pathways

The exact reaction coordinate is complex and multi (high) dimensional and a quantitative estimate of the free energy landscape cost can only be obtained using efficient long-time sampling approaches [97–101], which remain prohibitively expensive for QM/MM applications due to large computational cost. Although we believe that the true reaction coordinate is complex, in the spirit of the a recent study [102], we described the reaction pathways in terms of several simple reaction coordinates, χ. For mechanisms of phosphoryl transfer through associative mechanism, χ includes the reactive distance tyrosine O- ATP P_γ, tyrosine O- ATP O2/3β, and ATP P_γ- ATP O2/3β distances. For mechanisms of phosphoryl transfer through dissociative mechanism, in addition to the three distances, χ includes the coordinate for proton abstraction, namely tyrosine OH- D830:$O\delta_2$. In order to drive the system along a reaction coordinate, we performed restrained minimization as well as restrained sampling simulations using the QMMM Hamiltonian along χ by augmenting the Hamiltonian with a potential bias term $(1/2)\times K\times(|\chi|-\chi_0)^2$; here K is the effective spring constant (ranging between 10 and 30 kcal/mol/Å^2) and χ_0 are reference values for sampling in different windows of χ. Complete pathways were assembled by varying the offset distances χ_0 in increments of ±0.05 Å, where the positive and negative increments represent the directionality of the energy scan. The reaction paths were computed by energy-minimizations in the presence of restraints and recalculating the single-point energies in the absence of the constraints. For our choice of the increments of χ_0, calculations performed in forward and reverse directions satisfied path reversibility in configurations as well as energy values. Following the energy scans, constant temperature restrained dynamics runs of 300 fs-1ps for each choice of χ_0 was also performed to ensure that the reaction path was robust to thermal fluctuations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Mark Lemmon, Fumin Shi, Andrew Shih, Shannon Telesco, and E. Joseph Jordan for helpful discussions. The research leading to these results has received funding from the European Commission grant FP7-ICT-2011-9-600841 and from US National Institutes of Health Grant NIH U01-EB016027. Computational resources were provided in part by XSEDE under the grant MCB060006.

Abbreviations

MD	Molecular Dynamics
QM/MM	Quantum Mechanics/Molecular Mechanics
MMPBSA	Molecular Mechanics Poisson Boltzmann Surface Area
TKD	Tyrosine Kinase Domain
RTK	Receptor Tyrosine Kinase
EGFR	epidermal growth factor receptor

RMSD	root-mean-squared deviation
FEP	free energy perturbation
MMPBSA	molecular mechanics Poisson Boltzmann surface area

References

1. Citri A, Yarden Y. EGF-ERBB signalling: towards the systems level. Nat Rev Mol Cell Biol. 2006; 7(7):505–16. [PubMed: 16829981]
2. Mendelsohn J, Baselga J. The EGF receptor family as targets for cancer therapy. Oncogene. 2000; 19(56):6550–6565. [PubMed: 11426640]
3. Sharma SV, et al. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer. 2007; 7(3):169–81. [PubMed: 17318210]
4. Carey KD, et al. Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. Cancer Research. 2006; 66(16):8163–8171. [PubMed: 16912195]
5. Lynch TJ, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004; 350(21):2129–39. [PubMed: 15118073]
6. Paez JG, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004; 304:1497–1500. [PubMed: 15118125]
7. Stamos J, Sliwkowski MX, Eigenbrot C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J Biol Chem. 2002; 277(48): 46265–72. [PubMed: 12196540]
8. Zhang X, et al. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell. 2006; 125(6):1137–49. [PubMed: 16777603]
9. Beard WA, et al. Enzyme-DNA interactions required for efficient nucleotide incorporation and discrimination in human DNA polymerase beta. J Biol Chem. 1996; 271:12141–12144. [PubMed: 8647805]
10. Yun CH, et al. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity. Cancer Cell. 2007; 11(3):217–227. [PubMed: 17349580]
11. Adams JA. Kinetic and catalytic mechanisms of protein kinases. Chemical Reviews. 2001; 101(8): 2271–2290. [PubMed: 11749373]
12. Hanks SK, Quinn AM, Hunter T. The Protein-Kinase Family - Conserved Features and Deduced Phylogeny of the Catalytic Domains. Science. 1988; 241(4861):42–52. [PubMed: 3291115]
13. Bossemeyer D. Protein-Kinases - Structure and Function. Febs Letters. 1995; 369(1):57–61. [PubMed: 7641885]
14. Kim K, Cole PA. Kinetic analysis of a protein tyrosine kinase reaction transition state in the forward and reverse directions. J Amer Chem Soc. 1998; 120:6851–6858.
15. Taylor SS, et al. Crystal-Structures of the Catalytic Subunit of Camp-Dependent Protein-Kinase Reveal General Features of the Protein-Kinase Family. Receptor. 1993; 3(3):165–172. [PubMed: 8167567]
16. Tanoue TJ, Nishida E. Molecular recognitions in the MAP kinase cascades. Cellular Signalling. 2003; 15(5):455–462. [PubMed: 12639708]
17. Schulman BA, Lindstrom DL, Harlow E. Substrate recruitment to cyclin-dependent kinase 2 by a multipurpose docking site on cyclin A. Proceedings of the National Academy of Sciences of the United States of America. 1998; 95(18):10453–10458. [PubMed: 9724724]
18. Mok J, et al. Deciphering Protein Kinase Specificity Through Large-Scale Analysis of Yeast Phosphorylation Site Motifs. Science Signaling. 3(109)

19. Schulze WX, Deng L, Mann M. Phosphotyrosine interactome of the ErbB-receptor kinase family. Mol Syst Biol. 2005; 1 Article number: 2005.0008.
20. Kennelly PJ, Krebs EG. Consensus Sequences as Substrate-Specificity Determinants for Protein-Kinases and Protein Phosphatases. Journal of Biological Chemistry. 1991; 266(24):15555–15558. [PubMed: 1651913]
21. Ubersax JA, Ferrell JE. Mechanisms of specificity in protein phosphorylation. Nature Reviews Molecular Cell Biology. 2007; 8(7):530–541.
22. Songyang Z, et al. Catalytic Specificity of Protein-Tyrosine Kinases Is Critical for Selective Signaling. Nature. 1995; 373(6514):536–539. [PubMed: 7845468]
23. Hunter T, Cooper JA. Protein-Tyrosine Kinases. Annual Review of Biochemistry. 1985; 54:897–930.
24. Knighton DR, et al. Structural Features That Specify Tyrosine Kinase-Activity Deduced from Homology Modeling of the Epidermal Growth-Factor Receptor. Proceedings of the National Academy of Sciences of the United States of America. 1993; 90(11):5001–5005. [PubMed: 8389462]
25. House C, Baldwin GS, Kemp BE. Synthetic Peptide-Substrates for the Membrane Tyrosine Protein-Kinase Stimulated by Epidermal Growth-Factor. European Journal of Biochemistry. 1984; 140(2):363–367. [PubMed: 6609073]
26. Lahiri SD, et al. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. Science. 2003; 299(5615):2067–2071. [PubMed: 12637673]
27. Steitz TA, Steitz JA. A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci USA. 1993; 90:6498–6502. [PubMed: 8341661]
28. Valiev M, et al. Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations. J Phys Chem B. 2007; 111(47):13455–64. [PubMed: 17983217]
29. Turjanski AG, Hummer G, Gutkind JS. How mitogen-activated protein kinases recognize and phosphorylate their targets: A QM/MM study. J Am Chem Soc. 2009; 131(17):6141–8. [PubMed: 19361221]
30. Cleland WW, Hengge AC. Mechanisms of phosphoryl and acyl transfer. Faseb Journal. 1995; 9(15):1585–1594. [PubMed: 8529838]
31. Mildvan AS, Fry DC. Nmr-Studies of the Mechanism of Enzyme Action. Advances in Enzymology and Related Areas of Molecular Biology. 1987; 59:241–313. [PubMed: 3544711]
32. Schlichting I, Reinstein J. Structures of active conformations of UMP kinase drom Dictyostelium discoideum suggest phosphoryl transfer is associative. Biochemistry. 1997; 36:9290–9296. [PubMed: 9280438]
33. Bossemeyer D, et al. Phosphotransferase and substrate binding mechanism of the cAMP - dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 A structure of the complex with Mn2+ adenylyl imidodiphosphate and inhibitor peptide PKI(5–24). Embo J. 1993; 12:849–859. [PubMed: 8384554]
34. Zheng J, et al. 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. Acta Crystallogr D Biol Crystallogr. 1993; 49(Pt 3):362–5. [PubMed: 15299527]
35. Zhou J, Adams JA. Is there a catalytic base in the active site of cAMP-dependent protein kinase? Biochemistry. 1997; 36(10):2977–2984. [PubMed: 9062128]
36. Madhusudan, et al. Camp-Dependent Protein-Kinase - Crystallographic Insights into Substrate Recognition and Phosphotransfer. Protein Science. 1994; 3(2):176–187. [PubMed: 8003955]
37. Cheng YH, Zhang YK, McCammon JA. How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: An ab initio QM/MM study. Journal of the American Chemical Society. 2005; 127(5):1553–1562. [PubMed: 15686389]
38. Friesner RA, Guallar V. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. Annual Review of Physical Chemistry. 2005; 56:389–427.
39. Cheng YH, Zhang YK, McCammon JA. How does activation loop phosphorylation modulate catalytic activity in the cAMP-dependent protein kinase: A theoretical study. Protein Science. 2006; 15(4):672–683. [PubMed: 16522793]

40. De Vivo M, et al. Computational study of the phosphoryl transfer catalyzed by a cyclin-dependent kinase. Chemistry-a European Journal. 2007; 13(30):8437–8444.
41. Zhou BJ, Wong CF. A Computational Study of the Phosphorylation Mechanism of the Insulin Receptor Tyrosine Kinase. Journal of Physical Chemistry A. 2009; 113(17):5144–5150.
42. Diaz N, Field MJ. Insights into the phosphoryl-transfer mechanism of cAMP-dependent protein kinase from quantum chemical calculations and molecular dynamics simulations. J Am Chem Soc. 2004; 126(2):529–42. [PubMed: 14719950]
43. Levinson NM, et al. A Src-like inactive conformation in the Abl tyrosine kinase domain. Plos Biology. 2006; 4(5):753–767.
44. Bose R, et al. Protein tyrosine kinase-substrate interactions. Current Opinion in Structural Biology. 2006; 16(6):668–675. [PubMed: 17085043]
45. Wong CF, et al. Molecular docking of balanol to dynamics snapshots of protein kinase A. Proteins-Structure Function and Bioinformatics. 2005; 61(4):850–858.
46. Lin JH, et al. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. Biopolymers. 2003; 68(1):47–62. [PubMed: 12579579]
47. Gilson MK, Zhou HX. Calculation of Protein-Ligand Binding Affinities *. Ann Rev Biophys Biomol Struct. 2007; 36(1):21–42. [PubMed: 17201676]
48. Kuhn B, et al. Validation and use of the MM-PBSA approach for drug discovery. Journal of Medicinal Chemistry. 2005; 48(12):4040–4048. [PubMed: 15943477]
49. Swanson JMJ, Henchman R, McCammon JA. Revisiting free energy calculations: One step closer to rigorous scoring functions and one step beyond MM/PBSA. Abstracts of Papers of the American Chemical Society. 2004; 227:U903–U904.
50. Sharp KA, Honig B. Calculating Total Electrostatic Energies with the Nonlinear Poisson-Boltzmann Equation. Journal of Physical Chemistry. 1990; 94 (19):7684–7692.
51. Honig B, Sharp K, Yang AS. Macroscopic Models of Aqueous-Solutions - Biological and Chemical Applications. Journal of Physical Chemistry. 1993; 97 (6):1101–1109.
52. Kollman PA, et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Acc Chem Res. 2000; 33:889–897. [PubMed: 11123888]
53. Wang J, et al. Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. Journal of the American Chemical Society. 2001; 123(22):5221–5230. [PubMed: 11457384]
54. Fan Y-X, et al. Ligand Regulates Epidermal Growth Factor Receptor Kinase Specificity. J Biol Chem. 2004; 279:38143–38150. [PubMed: 15231819]
55. Mulloy R, et al. Epidermal growth factor receptor mutants from human lung cancers exhibit enhanced catalytic activity and increased sensitivity to gefitinib. Cancer Research. 2007; 67(5): 2325–2330. [PubMed: 17332364]
56. Shen K, et al. Protein kinase structure and function analysis with chemical tools. Biochimica Et Biophysica Acta-Proteins and Proteomics. 2005; 1754(1–2):65–78.
57. Hubbard SR. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. Embo J. 1997; 16(18):5572–81. [PubMed: 9312016]
58. Fan, et al. Ligand Regulates Epidermal Growth Factor Receptor Kinase Specificity. J Biol Chem. 2004; 279:38143–38150. [PubMed: 15231819]
59. Fan YX, Wong LL, Johnson GR. EGFR kinase possesses a broad specificity for ErbB phosphorylation sites, and ligand increases catalytic-centre activity without affecting substrate binding affinity. Biochemical Journal. 2005; 392:417–423. [PubMed: 16122376]
60. Brinkworth RI, Breinl RA, Kobe B. Structural basis and prediction of substrate specificity in protein serine/threonine kinases. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(1):74–79. [PubMed: 12502784]
61. Goldsmith EJ, et al. Substrate and docking interactions in serine/threonine protein kinases. Chemical Reviews. 2007; 107(11):5065–5081. [PubMed: 17949044]

62. Fan YX, et al. Ligand regulates epidermal growth factor receptor kinase specificity - Activation increases preference for GAB1 and SHC versus autophosphorylation sites. Journal of Biological Chemistry. 2004; 279(37):38143–38150. [PubMed: 15231819]
63. Brignola PS, et al. Comparison of biochemical and kinetic properties of type 1 receptor tyrosine kinase intracellular domains. J Biol Chem. 2001; 277:1576–1581. [PubMed: 11696537]
64. Prigent SA, Gullick WJ. Identification of c-erbB-3 binding sites for phosphatidylinositol 3′-kinase and SHC using an EGF receptor/c-erbB-3 chimera. Embo J. 1994; 13(12):2831–41. [PubMed: 8026468]
65. Kita YA, et al. NDF/heregulin stimulates the phosphorylation of Her3/erbB3. FEBS Lett. 1994; 349(1):139–43. [PubMed: 8045292]
66. Shan Y, et al. A conserved protonation-dependent switch controls drug binding in the Abl kinase. Proc Natl Acad Sci U S A. 2009; 106(1):139–44. [PubMed: 19109437]
67. Shih AJ, Purvis J, Radhakrishnan R. Molecular systems biology of ErbB1 signaling: Bridging the gap through multiscale modeling and high-performance computing. Mol Biosyst. 2008; 4:1151–1159. [PubMed: 19396377]
68. Shi F, et al. ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. Proc Natl Acad Sci U S A. 2010; 107(17):7692–7. [PubMed: 20351256]
69. Coker KJ, Staros JV, Guyer CA. A kinase-negative epidermal growth factor receptor that retains the capacity to stimulate DNA synthesis. Proc Natl Acad Sci U S A. 1994; 91:6967–6971. [PubMed: 8041731]
70. Liu Y, et al. A multiscale computational approach to dissect early events in the Erb family receptor mediated activation, differential signaling, and relevance to oncogenic transformations. Ann Biomed Eng. 2007; 35(6):1012–25. [PubMed: 17273938]
71. Purvis J, Ilango V, Radhakrishnan R. Role of Network Branching in Eliciting Differential Short-Term Signaling Responses in the Hyper-Sensitive Epidermal Growth Factor Receptor Mutants Implicated in Lung Cancer. Biotechnol Prog. 2008; 24(3):540–553. [PubMed: 18412405]
72. Purvis, J., et al. Efficacy of tyrosine kinase inhibitors in the mutants of the epidermal growth factor receptor: A multiscale molecular/systems model for phosphorylation and inhibition. Proc. Foundations in Systems Biology II; Stuttgart: IRB Verlag; 2007. p. 289-294.
73. Shih, A., et al. The role for molecular modeling in multiscale cancer models. In: Deisboeck, TS.; Stamatakos, G., editors. Multiscale Cancer Modeling of Cancer. 2010. p. 31-43.Chapman & Hall-CRC Mathematical and Computational Biology Series
74. Noro R, et al. Gefitinib (IRESSA) sensitive lung cancer cell lines show phosphorylation of Akt without ligand stimulation. Bmc Cancer. 2006:6. [PubMed: 16401344]
75. Sharma SV, et al. A common signaling cascade may underlie “addiction” to the Src, BCR-ABL, and EGF receptor oncogenes. Cancer Cell. 2006; 10(5):425–435. [PubMed: 17097564]
76. Mosesson Y, Mills GB, Yarden Y. Derailed endocytosis: an emerging feature of cancer. Nat Rev Cancer. 2008; 8(11):835–50. [PubMed: 18948996]
77. Hendriks BS, et al. Decreased internalisation of ErbB1 mutants in lung cancer is linked with a mechanism conferring sensitivity to gefitinib. Systems Biology IEE Proceedings. 2006; 153(6): 457–466.
78. Telesco SE, et al. A multiscale modeling approach to investigate molecular mechanisms of pseudokinase activation and drug resistance in the HER3/ErbB3 receptor tyrosine kinase signaling network. Mol Biosyst. 2011; 7(6):2066–80. [PubMed: 21509365]
79. Telesco SE, Vadigepalli R, Radhakrishnan R. Molecular modeling of ErbB4/HER4 kinase in the context of the HER4 signaling network helps rationalize the effects of clinically identified HER4 somatic mutations on the cell phenotype. Biotechnol J. 2013; 8(12):1452–64. [PubMed: 24318637]
80. Brooks BR, et al. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. Journal of Computational Chemistry. 1983; 4(2):187–217.
81. Radhakrishnan R, Schlick T. Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase beta’s closing. Proc Natl Acad Sci U S A. 2004; 101(16):5970–5. [PubMed: 15069184]

82. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996; 14(1): 33–8. 27–8. [PubMed: 8744570]
83. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998; 102(18):3586–3616. [PubMed: 24889800]
84. Morris GM, et al. Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. J Comput Chem. 1998; 19:1639–1662.
85. Phillips JC, Braun R, Wang W. Scalable molecular dynamics with NAMD. Journal of \ldots. 2005
86. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct. 2007; 36:21–42. [PubMed: 17201676]
87. Im W, Beglov D, Roux B. Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Computer Physics Communications. 1998; 111(1–3):59–75.
88. Sanner MF, Olson AJ, Spehner JC. Reduced surface: An efficient way to compute molecular surfaces. Biopolymers. 1996; 38(3):305–320. [PubMed: 8906967]
89. Roux B. The Calculation of the Potential of Mean Force Using Computer-Simulations. Computer Physics Communications. 1995; 91(1–3):275–282.
90. Schmidt MW, et al. General Atomic and Molecular Electronic-Structure System. Journal of Computational Chemistry. 1993; 14(11):1347–1363.
91. Szabo, A.; Ostlund, NS. Modern Quantum Chemistry. Mineola, New York: Dover Publications; 1996.
92. Field MJ, Bash PA, Karplus M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. J Comput Chem. 2002; 11:700–733.
93. Das D, et al. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. Journal of Chemical Physics. 2002; 117(23):10534–10547.
94. Radhakrishnan R, Schlick T. Fidelity discrimination in DNA polymerase beta: differing closing profiles for a mismatched (G:A) versus matched (G:C) base pair. J Am Chem Soc. 2005; 127(38): 13245–52. [PubMed: 16173754]
95. Radhakrishnan R, Schlick T. Correct and incorrect nucleotide incorporation pathways in DNA polymerase beta. Biochem Biophys Res Commun. 2006; 350(3):521–9. [PubMed: 17022941]
96. Radhakrishnan R. Coupling of fast and slow modes in the reaction pathway of the minimal hammerhead ribozyme cleavage. Biophys J. 2007; 93(7):2391–9. [PubMed: 17545240]
97. Brooks CL, Case DA. Theory and simulation - the control and timescale of structure and reactivity in biological systems: from peptide folding to cellular networks. Curr Opin Struct Biol. 2003; 13:143–145.
98. Elber R, et al. Bridging the gap between long time trajectories and reaction pathways. Adv Chem Phys. 2003; 126:93–129.
99. Chu JW, Trout BL, Brooks BR. A super-linear minimization scheme for the nudged elastic band method. The Journal of Chemical Physics. 2003; 119(24):12708–12717.
100. Bolhuis PG, et al. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. Annual Review of Physical Chemistry. 2002; 53:291–318.
101. Ren W, et al. Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. Journal of Chemical Physics. 2005; 123(13)
102. Rosta E, et al. Artificial reaction coordinate “tunneling” in free-energy calculations: the catalytic reaction of RNase H. J Comput Chem. 2009; 30(11):1634–41. [PubMed: 19462398]

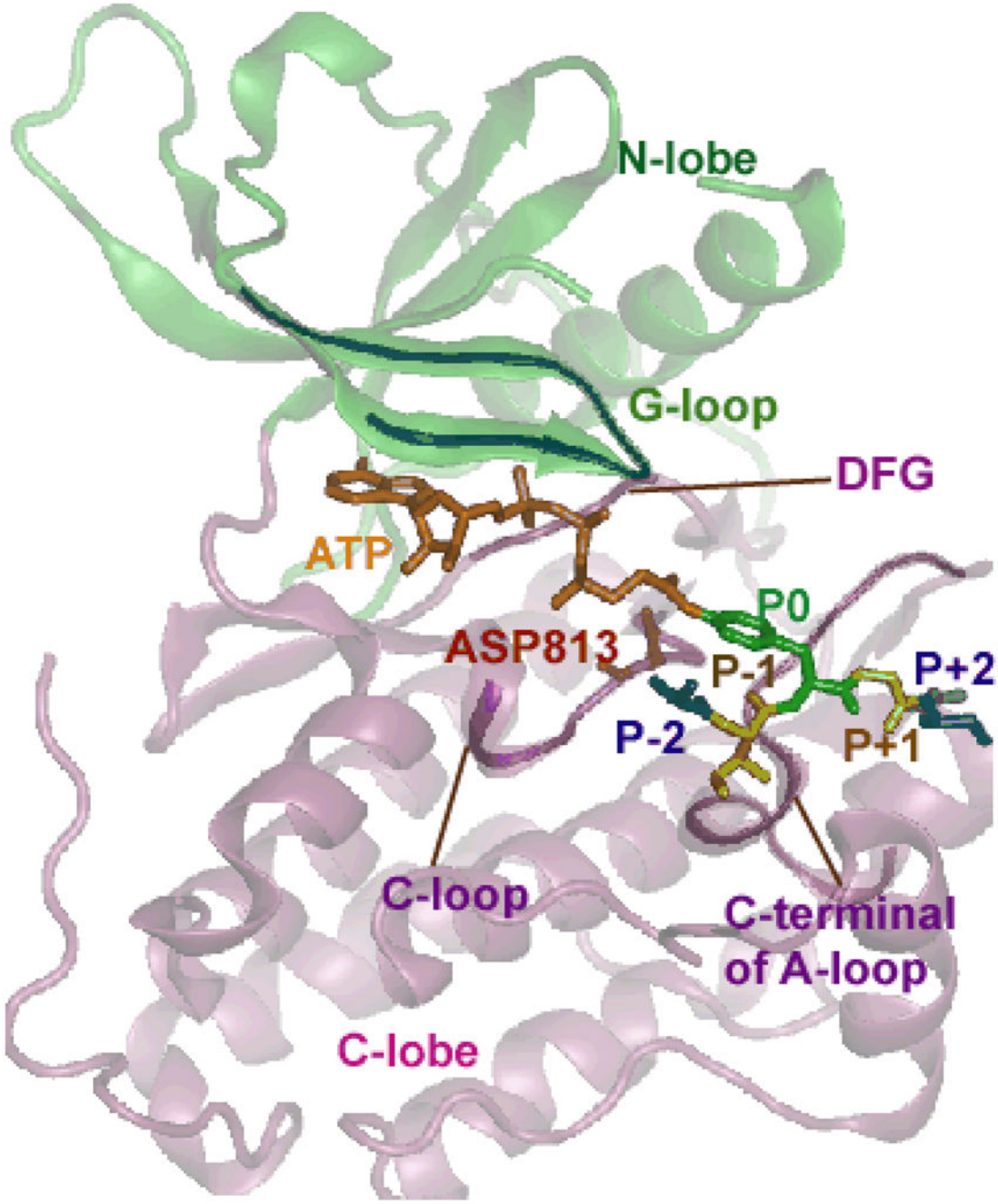

Figure 1.

Structure of EGFR TKD co-crystallized with a bi-substrate analog. The kinase is depicted in ribbon representation, green color for N-lobe and pink color for C-lobe. The bisubstrate is represented as licorice with the color code as follows: ATP is shown in orange; tyrosine residue in the peptide is colored green and numbered as P0; the amino acids situated immediately proximal to the P0 site are numbered as P−1, P−2, P−3, etc., (going towards the N-terminus from the P0 amino acid), and P+1, P+2, P+3, etc., (going towards the C-terminus from the P0 amino acid), which are colored individually. The conserved aspartic

acid residue (ASP813 in EGFR) is highlighted in red (stick representation). Other conserved motifs, namely the DFG motif, the C-loop, and the A-loop are also highlighted as opaque ribbons.

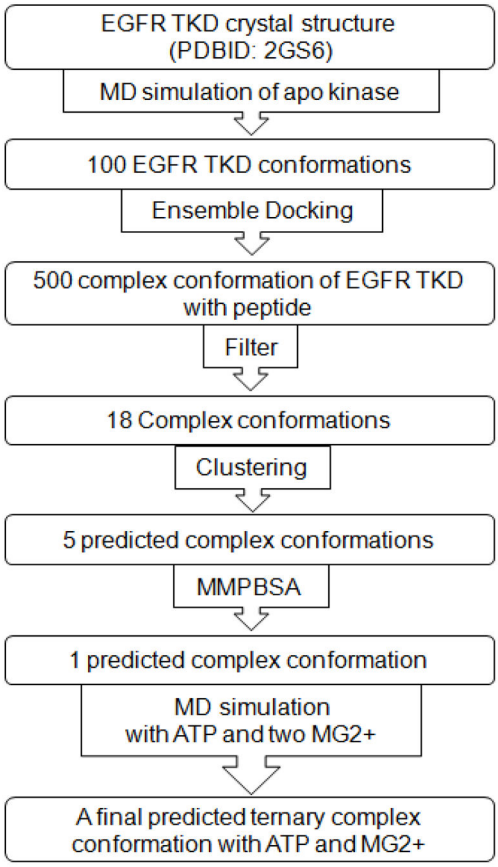

Figure 2.
Docking protocol to predict peptide bound conformation in EGFR TKD.

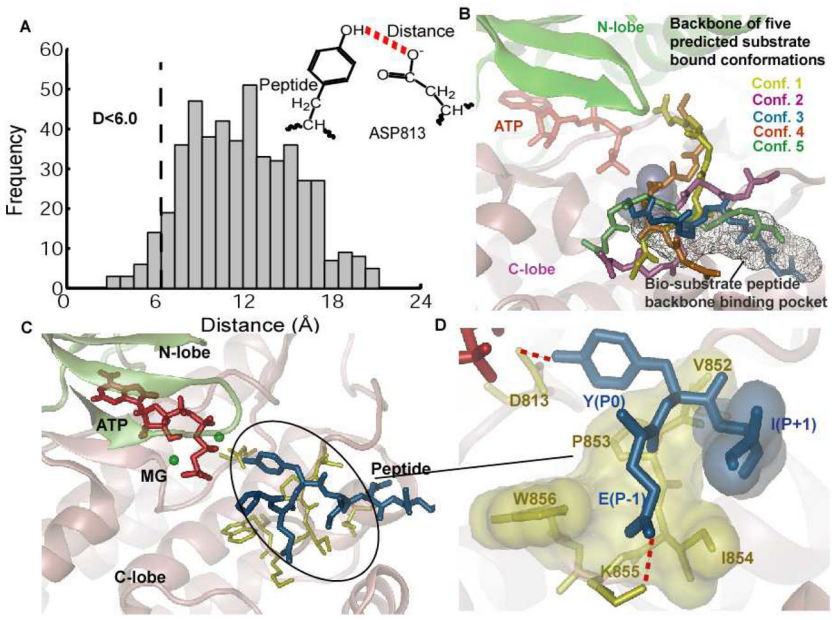

Figure 3.

Results for docking the peptide Y1068 to the EGFR TKD. A) Histogram of the distance between the oxygen in hydroxyl group of the tyrosine and the oxygen atom in the carboxylate group of the D813 for all the 500 predicted conformations by ensemble docking. Conformations with the above distance less than 6 Å were clustered into 5 clusters based on their backbone RMSD are shown in panel B. B) Five predicted peptide clusters of Y1068 bound conformations generated by ensemble docking that satisfied the distance criteria in panel A. The bi-substrate peptide binding pocket is also shown in the background, for reference. All 5 conformations were re-ranked using the MMPBSA method (see Table 1). C) Bound conformation of the peptide Y1068 to EGFR TKD in the presence of ATP and two MG^{2+} ions after 10 ns of MD simulation. The ATP atoms are depicted in red, MG^{2+} in green, peptide in blue, and residues in EGFR TKD that interact with the peptide in yellow. D) Depiction of the main interactions between the peptide and the kinase: Y(P0) hydrogen-bonds with D813; E(P−1) forms a salt bridge interaction with residue K855; the hydrophobic residue I(P+1) interacts with a hydrophobic surface in EGFR TKD consisting of the residues V852-P853-I854-W856. All hydrophobic residues are depicted as transparent surfaces.

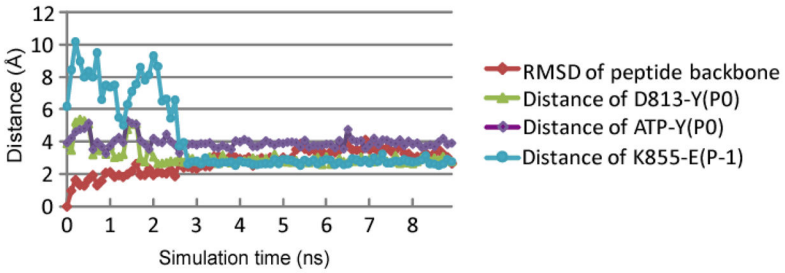

Figure 4.

Evolution of the RMSD and key distances in the 10 ns MD simulation of EGFR TKD bound to ATP and the Y1068 peptide substrate. The color schemes are: the RMSD calculated based the backbone atoms of the central 5 residues of the peptide relative to the initial conformation (red); the distance between D813:Oδ_2 and Y(P0):Oη (green); the distance between Y(P0):Oη and ATP:P_γ (purple); and the salt bridge distance between K855 and E(p −1) (turquoise).

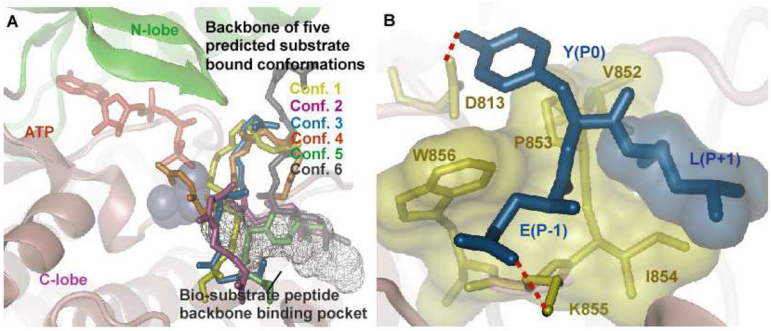

Figure 5.

Results for docking the peptide Y1173 to EGFR TKD. A) Six predicted peptide clusters of the Y1173 bound conformations generated by ensemble docking that satisfied the distance criteria in panel A. The bi-substrate peptide binding pocket is shown in the background, for reference. All 6 conformations were re-ranked using the MMPBSA method (see Table 2). B) A depiction of the main interactions between the peptide and kinase: Y(P0) hydrogen-bonds with D813; E(P−1) forms a salt bridge interaction with residue K855; hydrophobic residue L(P+1) interacts with a hydrophobic surface in EGFR TKD consisting of residues V852-P853-I854-W856. All hydrophobic residues are depicted as transparent surfaces.

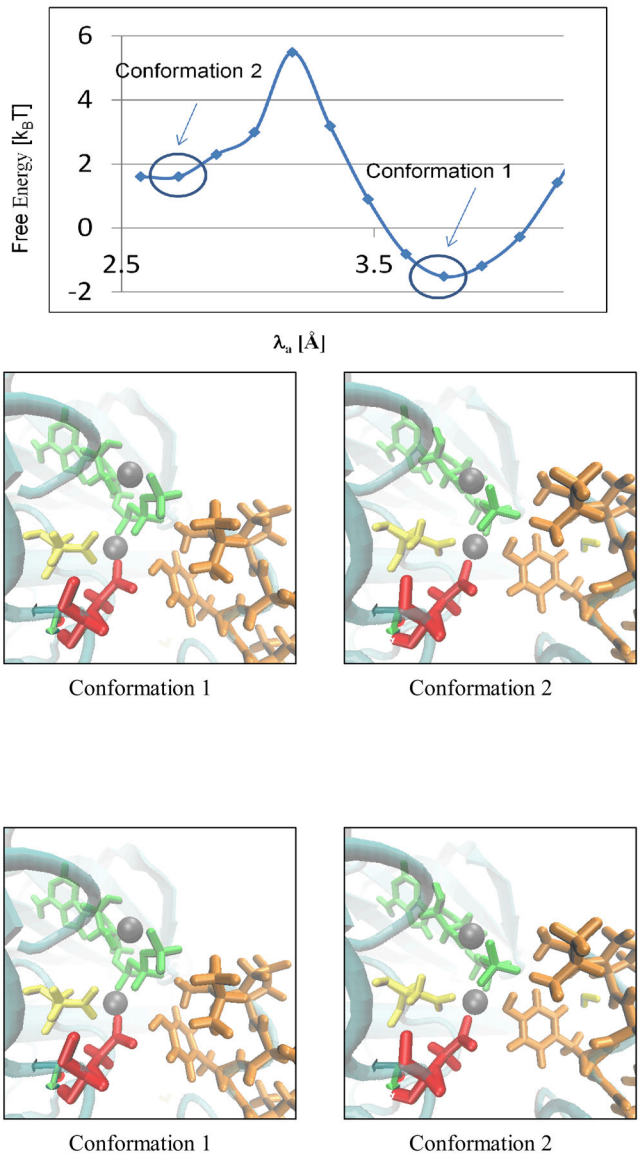

Figure 6.

Pre-Catalytic conformational landscape along the nucleophilic attack distance λ_a. Representative snapshots of conformations 1 and 2 are also depicted. In these snapshots the ATP is in green, the peptide is in orange, D813 is in red, D831 is in yellow, and the Mg^{2+} ions are in grey.

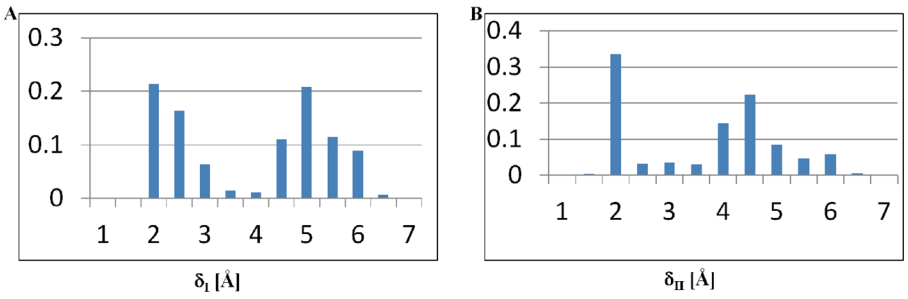

Figure 7.

Normalized histograms representing the distributions of values for the distances between the proton of the substrate tyrosine hydroxyl and (*A*) the $O2_{\delta}$ oxygen of D813 in EGFR (δ_I); and (*B*) the $O1_{\gamma}$ oxygen of ATP (δ_{II}) recorded in our simulations of the ground state (reactant) dynamics.

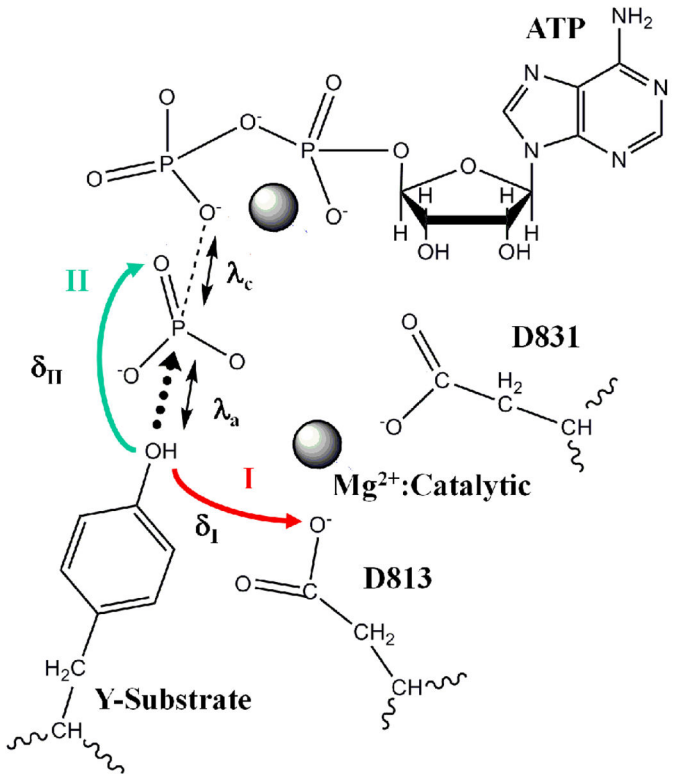

Figure 8.

Schematic of the proposed pathways for phosphoryl-transfer in EGFR TKD. Mg^{2+} ions and the catalytic aspartates (D831) are marked. Two pathways for proton migration (concomitant with phosphoryl-transfer) are shown. Pathway I (red) involves proton abstraction from the substrate tyrosine –OH group by D813 of EGFR, and pathway II (turquoise) involves proton migration to the $O1_{\gamma}$ oxygen of ATP. δ_I and δ_{II} denote distances between the proton of the substrate tyrosine hydroxyl and either the $O_{\delta 2}$ oxygen of D813 (δ_I) or the $O1_{\gamma}$ oxygen of ATP (δ_{II}). The nucleophilic attack distance, λ_a (distance between the tyrosine oxygen and the ATP P_{γ}) and the bond cleavage distance, λ_c (distance between the ATP P_{γ} and ATP $O2/3_{\beta}$) are also marked.

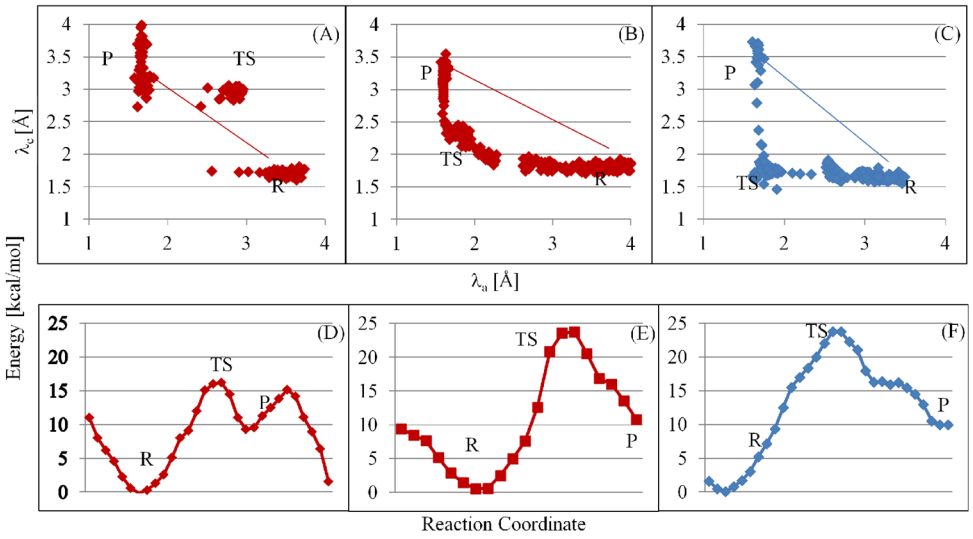

Figure 9.

QM/MM results for proposed pathways in figure 8 for the EGFR TKD/ATP/Y1068 peptide ternary complex. (A)&(D): Pathway I, dissociative mechanism; (B)&(E): Pathway I, associative mechanism; and (C)&(F): Pathway II, associative mechanism. R = reactant; P = product; and TS = transition state characterized by a trigonal bipyramidal geometry about P_γ. (A–C) Distributions of values for the nucleophilic attack distance, λ_a (distance between the tyrosine oxygen and the ATP P_γ) and the bond cleavage distance, λ_c (distance between the ATP P_γ and ATP O2/3_β) along the reaction pathways. *(D–F)* Energy changes along the reaction pathways in QM/MM simulations for the different mechanisms. In the energy landscape in panel D, the first peak corresponds to the formation of the trigonal-bipyramidal transition state where the proton from the tyrosine hydroxyl has been dissociated, while the second peak corresponds to the proton migrating to the ADP molecule.

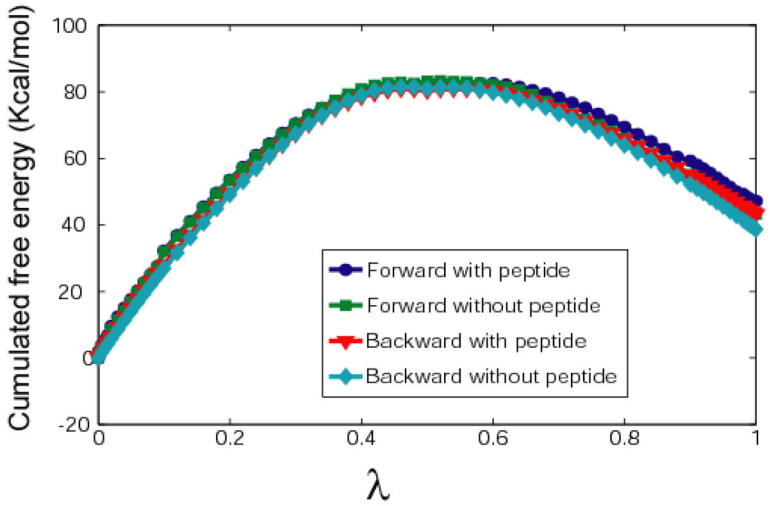

Figure 10.

Cumulative free energy changes upon L834R mutation (Kcal/mol) calculated for bound and unbound Y1068 peptide substrates in both forward and backward directions in the free energy perturbation simulations.

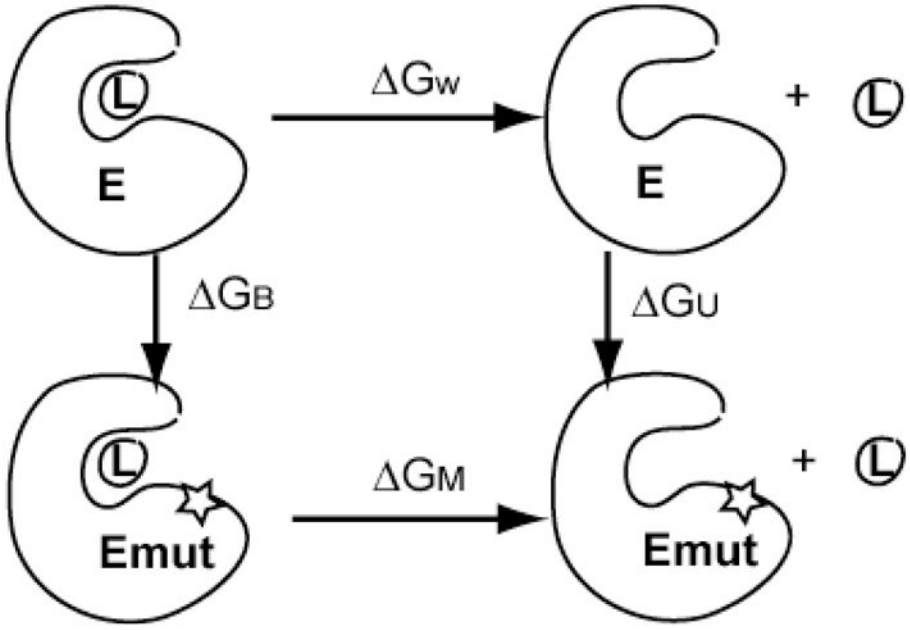

Figure 11.
Thermodynamic cycle for the calculation of binding affinity change of ligand/substrate due to a mutation employed in the FEP simulations.

Table 1

Rescoring of the top five predicted Y1068 peptide conformations predicted by ensemble docking using the MMPBSA method based on 1 ns MD simulation. The RMSD is calculated between the backbone atoms of the central five residues of the modeled peptide to the backbone atoms of the five-residue peptide portion of the bi-substrate in the structure 2GS6.

Conf.	$\Delta\bar{U}_{MM}$ (Kcal/mol)	$\Delta\bar{W}_{PB}$ (Kcal/mol)	$\Delta\bar{W}_{SA}$ (Kcal/mol)	$\Delta\bar{G}_{trans}$ ($-T\Delta\bar{S}_{trans}$) (Kcal/mol)	$\Delta\bar{G}_{rot}$ ($-T\Delta\bar{S}_{rot}$) Kcal/mol)	$\Delta\bar{G}_{total}$ (Kcal/mol)	RMSD to bi-substrate (Å)
1	−157.4	153.3	−7.7	5.2	4.2	−2.4	5.45
2	−131.4	121.9	−7.5	5.5	4.3	−7.2	5.09
3	−134.4	109.4	−6.5	4.9	4.3	−22.2	2.72
4	−206.3	197.4	−7.9	5.4	4.2	−7.3	5.97
5	−107.5	103.7	−7.3	5.2	4.3	−1.6	3.92

Table 2

Rescoring of the top six predicted Y1173 peptide conformations predicted by ensemble docking using the MMPBSA method based on 1 ns MD simulation. The RMSD is calculated between the backbone atoms of the central five residues of the modeled peptide to the backbone atoms of the five-residue peptide portion of the bi-substrate in the structure 2GS6.

Conf.	$\Delta\bar{U}_{MM}$ (Kcal/mol)	$\Delta\bar{W}_{PB}$ (Kcal/mol)	$\Delta\bar{W}_{SA}$ (Kcal/mol)	$\Delta\bar{G}_{trans}$ ($-T\Delta\bar{S}_{trans}$) (Kcal/mol)	$\Delta\bar{G}_{rot}$ ($-T\Delta\bar{S}_{rot}$) Kcal/mol)	$\Delta\bar{G}_{total}$ (Kcal/mol)	RMSD to bi-substrate (Å)
1	−436.5	394.6	−8.0	5.9	4.5	−39.4	9.6
2	−238.0	217.1	−6.9	4.7	4.3	−18.8	10.0
3	−243.8	226.2	−8.1	5.4	4.3	−16.1	7.9
4	−114.2	111.8	−7.4	5.1	3.9	−0.7	10.6
5	−297.0	241.3	−7.1	5.3	4.7	−52.8	3.6
6	−250.1	228.5	−8.0	5.4	4.3	−19.9	9.7

Table 3

Cumulative free energy change upon L834R mutation (Kcal/mol) calculated for bound and unbound Y1068 peptide substrate in both forward and backward directions.

	ΔF (Kcal/mol)		ΔΔF (Kcal/mol)
	With peptide	**No peptide**	
Forward	47.188	43.247	3.941
Backward	43.3	38.716	4.584