

# The Energy Landscape Analysis of Cancer Mutations in Protein Kinases

Anshuman Dixit<sup>1</sup>, Gennady M. Verkhivker<sup>2,3\*</sup>

**1** Department of Pharmaceutical Chemistry, School of Pharmacy, The University of Kansas, Lawrence, Kansas, United States of America, **2** School of Computational Sciences and Crean School of Health and Life Sciences, Schmid College of Science and Technology, Chapman University, Orange, California, United States of America, **3** Department of Pharmacology, University of California, San Diego, La Jolla, California, United States of America

## Abstract

The growing interest in quantifying the molecular basis of protein kinase activation and allosteric regulation by cancer mutations has fueled computational studies of allosteric signaling in protein kinases. In the present study, we combined computer simulations and the energy landscape analysis of protein kinases to characterize the interplay between oncogenic mutations and locally frustrated sites as important catalysts of allosteric kinase activation. While structurally rigid kinase core constitutes a minimally frustrated hub of the catalytic domain, locally frustrated residue clusters, whose interaction networks are not energetically optimized, are prone to dynamic modulation and could enable allosteric conformational transitions. The results of this study have shown that the energy landscape effect of oncogenic mutations may be allosteric eliciting global changes in the spatial distribution of highly frustrated residues. We have found that mutation-induced allosteric signaling may involve a dynamic coupling between structurally rigid (minimally frustrated) and plastic (locally frustrated) clusters of residues. The presented study has demonstrated that activation cancer mutations may affect the thermodynamic equilibrium between kinase states by allosterically altering the distribution of locally frustrated sites and increasing the local frustration in the inactive form, while eliminating locally frustrated sites and restoring structural rigidity of the active form. The energy landscape analysis of protein kinases and the proposed role of locally frustrated sites in activation mechanisms may have useful implications for bioinformatics-based screening and detection of functional sites critical for allosteric regulation in complex biomolecular systems.

**Citation:** Dixit A, Verkhivker GM (2011) The Energy Landscape Analysis of Cancer Mutations in Protein Kinases. PLoS ONE 6(10): e26071. doi:10.1371/journal.pone.0026071

**Editor:** Jie Zheng, University of Akron, United States of America

**Received:** June 21, 2011; **Accepted:** September 19, 2011; **Published:** October 6, 2011

**Copyright:** © 2011 Dixit, Verkhivker. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work is partly supported by funding from The University of Kansas. No additional external funding was received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: gverkhiv@ucsd.edu

## Introduction

Rapid and efficient communication of long-range conformational changes in proteins plays a vital role in allosteric regulation of biological systems [1,2]. Recent seminal reviews of protein allostery have emphasized a central role of cooperativity and the notion that catalysis and allostery may emerge via common communication routes [3,4]. Modeling of allosteric transitions in biological molecules has been significantly advanced by the development of elastic network models and normal mode analysis approaches [5-22]. Elastic network models of protein dynamics and signal propagation theory have allowed for a quantitative analysis of long-range allosteric protein interactions [13-16]. Sequence-based evolutionary analysis [23,24] and structure-based approaches [19,20,25-27] have demonstrated that allosteric pathways in proteins may be formed through interactions of evolutionary conserved and sparsely connected clusters of residues that are energetically coupled to mediate long-range communication. A comprehensive analysis of allosteric mechanisms has led to a unified view of allosteric regulation that implies the existence of preexisting conformational states and multiple communication pathways on the conformational landscape [28-32]. Energy landscape theories and simplified energy models have provided a robust theoretical framework to elucidate fundamental aspects of

protein structure, dynamics and allosteric regulation [33-43]. According to the modern energy landscape theory, random sequences have rugged landscapes with many local minima due to severe conflicting interactions (a phenomenon termed “frustration”) and, as a result, the prevalence of structurally alternative yet energetically similar conformations. The energy landscape models have also suggested that protein-like sequences may have evolved to partially eliminate frustrated interactions between amino acids and have smoothed (“funnel-like”) landscapes to ensure fast folding to their thermodynamically stable native structures. This has become known as the ‘principle of minimal frustration’ [44,45]. The funneled-like nature of the energy landscapes for natural proteins implies that the conformations that are structurally similar to the native state are also low in energy, and the native state interactions are minimally frustrated [33-45]. A generalized view of allosteric regulation based on the energy landscape theory (often termed as a “conformational selection model”) suggests that a protein may function in a dynamic equilibrium of structurally different conformational states, whereby the effect of binding or mutation can be propagated over long distances by cooperatively shifting the equilibrium towards a functionally relevant conformation [46-52]. The “old” view (induced fit mechanism) and the “new” view (conformational selection mechanism) of protein allostery appeared not to be mutually exclusive but rather

complementary in rationalizing allosteric mechanisms at the molecular level [53-56]. Physics-based simulation approaches have provided a compelling evidence of coupling between collective motions and local structural changes as an important underlying principle of allosteric communication in biomolecules [53-60]. Thermodynamics-based approaches have further linked global and local structural perturbations with free energy changes of allosteric coupling in mechanisms conformational switching [61-64]. Moreover, the energy landscape models have suggested that long-range cooperativity of protein-protein interactions during allosteric transitions may favor a combination of the population-shift and induced-fit mechanisms, whereas short-range, allosteric binding of proteins with inhibitors could often proceed via the population-shift mechanism [65-72].

Ferreiro and Wolynes [69] have recently advanced the energy landscape theory by combining biophysical modeling and structural bioinformatics analyses of local protein interactions that are fundamental for folding, binding and allosteric regulation. According to this model, minimally frustrated landscapes of protein networks may have evolved to acquire the ability for regulation via cooperative allosteric changes. The proposed method has quantified the degree of spatial local frustration in proteins using a local version of the global gap criterion formulation of the minimal frustration principle [33-45]. This model introduced a local frustration metric termed “configurational frustration index” as a measure of local stabilization for an individual native pair with respect to a set of structural decoys generated by perturbing both the identities and location of the interacting amino acids [69,70]. According to this criterion, if the interaction energy of a native pair of residues is sufficiently stabilizing as compared to the set of structural decoys, this residue pair is designated as “minimally frustrated”, otherwise the interactions may be classified as either “neutral” or “locally frustrated”. It is worth noting that the principle of minimal frustration does not require a complete elimination of locally stable alternative structures. A certain degree of local frustration is always present in an otherwise largely unfrustrated protein structure and may have arisen from evolutionary requirements to adapt protein dynamics for specific functions [43].

The analysis of locally frustrated protein regions using a non-redundant set of 314 monomeric protein domains and a curated set of nonredundant dimeric complexes has shown that the locally frustrated sites correspond to the regions involved in binding with other macromolecules and ligands and could often collocate with the functional groups prone to large structural changes [69,70]. Wolynes and coworkers have recently surveyed a curated database of allosteric proteins with known inactive and active crystal structures and have demonstrated that allosteric protein domains are connected by a web of minimally frustrated interactions, while highly frustrated residues could be preferentially clustered near the protein surface [71,72]. According to this study, minimally frustrated regions in allosteric proteins domains constitute nearly 40% of the total contacts, with about 10% of the total interactions considered to be “highly frustrated”, and the remainder of interactions attributed to the “neutral” category.

Protein kinases are signaling switches with a conserved catalytic domain that phosphorylate protein substrates and play a critical role in cell signaling pathways [73-82]. Protein kinase genes constitute ~2% of all genes in human genome and this protein family consists of more than 500 diverse members. The crystal structures of human protein kinases include 167 unique human protein kinase domains and 170 kinases, considering closely related orthologues (<http://www.sgc.ox.ac.uk/research/kinases/>). Structural studies of protein kinase catalytic domain structures and

regulatory protein complexes have revealed distinct scenarios by which kinases can control a dynamic equilibrium between structurally similar active and highly specific inactive kinase states - a structural hallmark of the kinase domain critical for its normal function [83-89]. Allosteric regulation may be achieved via different mechanisms including inhibitor-induced stabilization of the specific inactive conformation in ABL [90-95], BRAF [96], KIT [97], PDGFR, P38 [98], PI3K kinases [99] and binding to the allosteric myristoyl-binding pocket in ABL [100-103]. Protein kinase activation can be also regulated via formation of structurally diverse regulatory complexes most notably exemplified for ABL [104,105] and EGFR kinases [106-109], yet a unifying structural mechanism associated with asymmetric tyrosine kinase arrangements in regulatory complexes could underlie the activation mechanism of the entire EGF protein family [110-113]. A steady progress in understanding of protein kinase mechanisms has fueled a considerable effort to discover and design selective ATP-competitive and allosteric inhibitors targeting specific forms of cancer, kinase cancer mutants and associated targeted pathways [114-116].

Abnormal activation of regulation in protein kinases is a dominant source of tumor-associated somatic mutations. Structural and mutagenesis investigations of ABL [93-95] and EGFR kinases [117-119] have revealed structural divergence of the kinases in response to activating mutations. Kinome-wide bioinformatics studies have contributed to the identification of conserved sequence motifs harboring disease-associated and cancer mutations, suggesting that a significant number of oncogenic cancer mutations could form structurally conserved mutational hotspots within the kinase catalytic core [120-123]. Computer simulation studies have investigated molecular mechanisms of protein kinase activation in c-Src [124-127], adenylate kinase [128], ABL [129], CDK5 [130], KIT [131], RET, MET [132] and EGFR kinase [133-135]. Multi-scale simulation studies of conformational transitions in the normal and oncogenic forms of ABL and EGFR kinases have indicated that the impact of the oncogenic mutants may spread beyond the immediate site of mutation leading to global allosteric changes [134]. Most recently, computational modeling of allosteric regulation has revealed organizing principles of mutation-induced activation in ABL and EGFR kinases, which may be determined by a dynamic coupling between structurally rigid  $\alpha$ F-helix and conformationally adaptive  $\alpha$ I-helix and  $\alpha$ C-helices [136]. These structural elements form a dynamic network of efficiently interacting functional regions that may universally control the long-range interdomain communication and allosteric activation in protein kinases. The energy landscape studies have previously suggested that localized frustration may be connected with allosteric conformational changes in proteins [69-72].

Collectively, computational studies have suggested that molecular mechanisms of allosteric regulation in protein kinases can be described using models of mutation-induced modulation of the conformational landscape and conformational selection principles of the thermodynamically relevant states. In this work, kinome-based structural bioinformatics analysis and biophysical modeling of protein kinase structures were employed to characterize and quantify the interplay between oncogenic kinase mutations and locally frustrated sites as potential catalysts and mediators of kinase activation. The results of this study suggest that the energy landscape effect of oncogenic mutations may be allosteric in nature, eliciting global changes in the spatial distribution of highly frustrated residues. We show that cancer mutations could act by simultaneously perturbing the network of minimally frustrated interactions in the inactive kinase state, while reducing local

frustration and restoring allosteric interactions in the active kinase form. Hence, locally frustrated sites in the catalytic core may serve an important functional role by enabling mutation-induced conformational transitions towards the constitutively active kinase conformation.

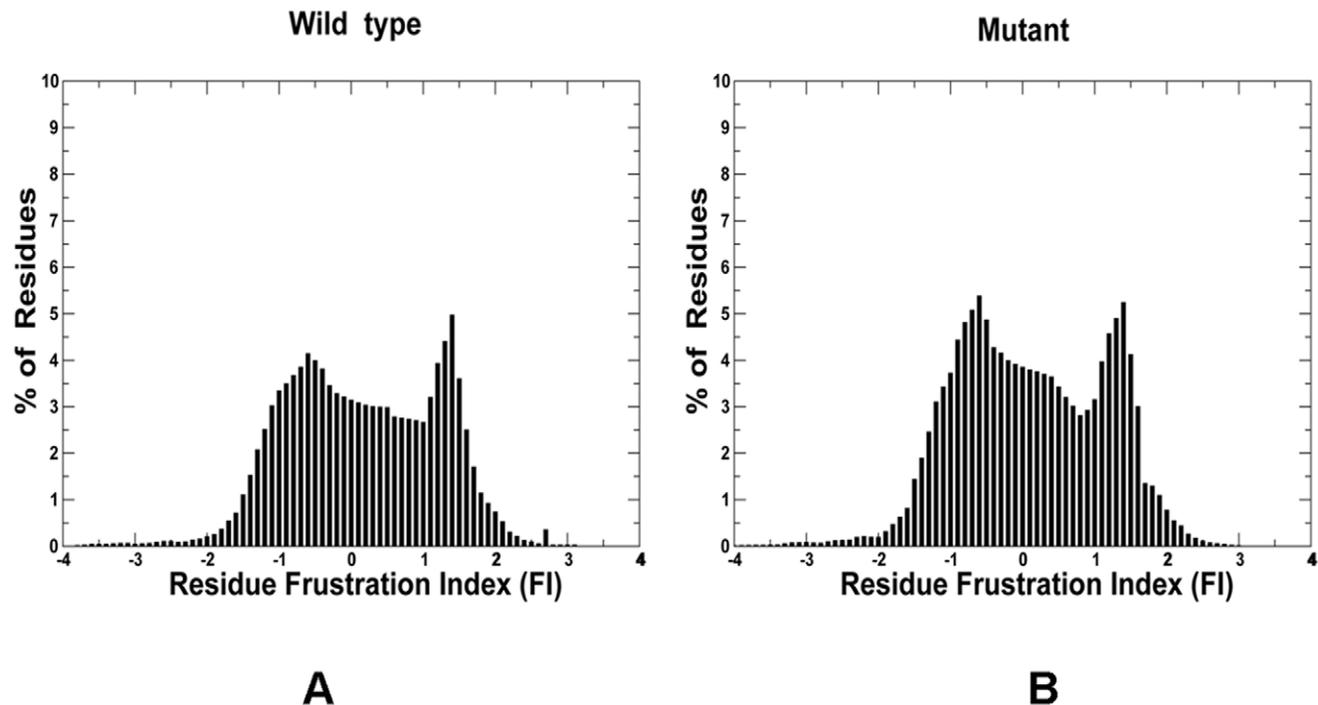
## Results

### The Energy Landscapes and Local Frustration in Protein Kinases

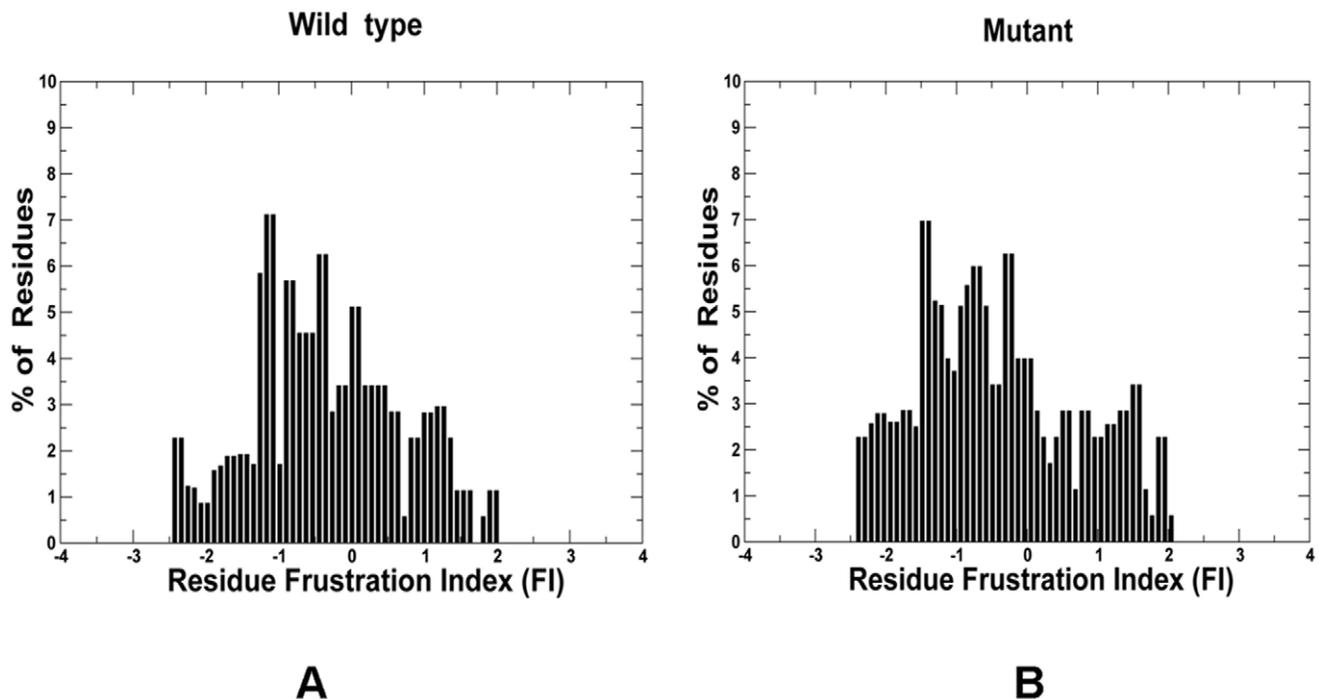
In the present study, we combined molecular dynamics (MD) simulations of protein kinases with the energy landscape analysis to characterize the role of local frustration as an important factor associated with allosteric kinase activation. From the energy landscape perspective, the mechanistic features of the activation transitions should be determined by the structural topology of the kinase domain fold and therefore could capture salient aspects of the activating mechanism. To investigate the role of local frustration in conformational transitions between structurally different functional states, we surveyed the local frustration profiles in protein kinase structures and characterized the network of minimally frustrated interactions responsible for structural stability of the kinase catalytic core. We also located and characterized clusters of locally frustrated sites where the minimal frustration principle could be violated. The change in the configurational frustration index upon mutation can provide a quantitative measure of a tendency to bring about a conformational change in the protein. The effect of kinase cancer mutations on local frustration profiles allowed us to quantify how mutation-induced redistribution of locally frustrated residues can promote allosteric transitions between structurally distinct functional states. The configuration frustration index could measure the relative stability of a particular native contact relative to the set of all possible contacts in that location, thus allowing to classify the individual

native contacts in the protein structure according to their frustration level. A kinome-wide examination of the configurational frustration index computed for a large number of protein kinase crystal structures (**Table S1** in **File S1**) revealed that the typical values may range between  $-4$  to  $+4$ . The overall residue-based distribution of a frustration index for the wild type (WT) kinases was biased towards minimally frustrated residues with the positive values of the frustration index (**Figure 1A**). The distribution also displayed a smaller shallow peak corresponding to locally frustrated residues with the frustration index in the range of  $-0.6$  to  $-0.7$  units. The impact of mutations resulted in a subtle yet noticeable change in the distribution local of frustration in the catalytic core, revealing a second equally important peak around  $-1.0$  value. Hence, the overall distribution was almost evenly divided between the minimally frustrated and frustrated residues, leaving fewer residues at the neutral status (**Figure 1B**).

We focused then on the local frustration analysis performed for a subset of protein kinase genes (ABL, EGFR, BTK, KIT, BRAF, MET, and RET) that account for the vast majority of highly oncogenic mutations in the catalytic domain (**Table S2** in **File S1**). These protein kinase genes were chosen for a more detailed analysis because of the wealth of structural and functional information that provided complementary experimental data for validation of our models. More importantly, however, a diverse repertoire of activating and drug resistant mutations in these kinases genes represent critical cancer culprits that could frequently contribute to a state of oncogene dependency in a variety of cancers. The distribution of local frustration in these kinase genes, as measured by the configurational frustration index, revealed a distinct pattern where the peaks were noticeably shifted towards more frustrated residues for both the WT and mutant kinases (**Figure 2**). The percentage of minimally frustrated interactions in the catalytic core accounted for more than 40% of the total contacts, with about 15-20% of the interactions be



**Figure 1. The histogram of a residue-based frustration index values in the catalytic domain from a kinome-based analysis.** (A) Wild-type kinases and (B) Mutant kinases. doi:10.1371/journal.pone.0026071.g001

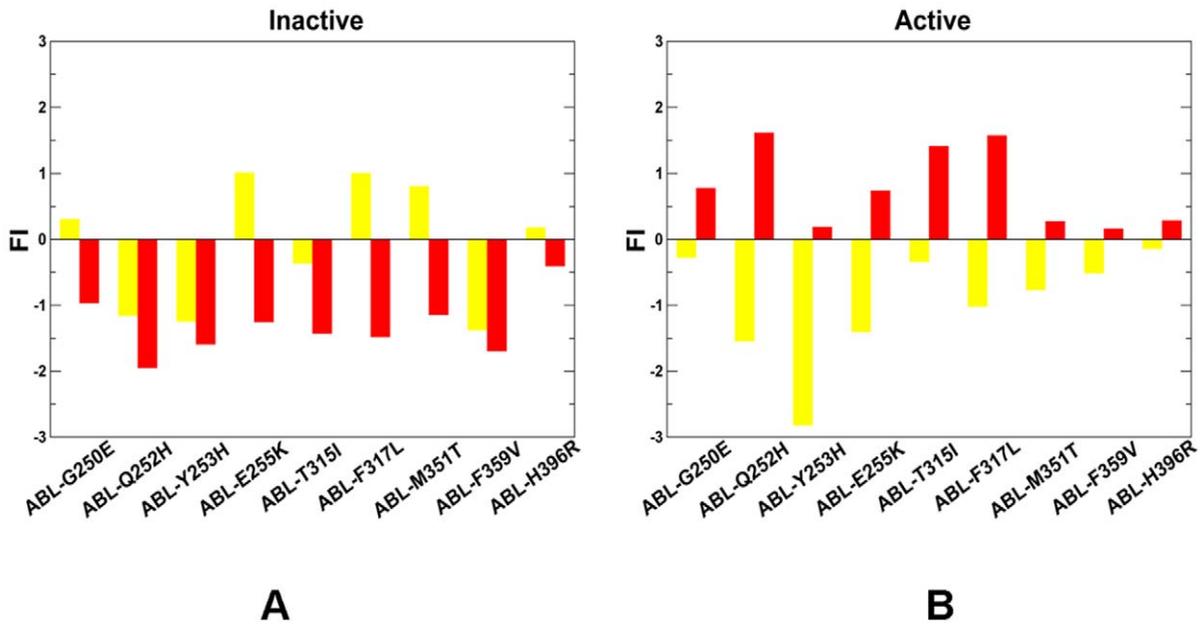


**Figure 2. The histogram of a residue-based frustration index values in the kinase catalytic domain from a statistical analysis of dominant kinase oncogenes.** (A) Wild-type kinases and (B) Mutant kinases. A set of employed kinase oncogenes included ABL, EGFR, BTK, KIT, MET, BRAF, and RET kinases. The analysis included mutants of these kinase genes with high oncogenic potential according to the frequency profiles in the mutational samples (>5) obtained from the COSMIC repository [153]. doi:10.1371/journal.pone.0026071.g002

considered as frustrated and the remainder neutral. This analysis generally agreed with the reported distribution of frustrated regions and partition of minimally and highly frustrated residues in proteins [69,70]. However, the average fraction of locally frustrated residues was higher in protein kinases than the one reported for small monomeric proteins. Hence, our data suggested that conformational landscapes of kinase oncogenes may be characterized by an increased level of local frustration and protein mobility.

Based on this analysis, we proposed that the spatial distribution of local frustration in protein kinases may be regulated and readily changed by oncogenic mutations. According to our conjecture, activating kinase mutations could amplify the local frustration in the inactive state, while eliminating (or partly removing) locally frustrated sites in the active state. As result, mutation-induced redistribution of local frustration in protein kinase structures may contribute to the molecular mechanisms that control kinase activity by altering the dynamic equilibrium between functional kinase forms. To verify this hypothesis, we analyzed changes in the local frustration profiles for a representative set of highly oncogenic ABL (Figure 3) and EGFR kinase mutants (Figure 4) in both inactive and active states. We observed that highly oncogenic mutations may indeed cause an increase in the local frustration of mutated residues in the inactive autoinhibitory state of ABL (PDB ID 1IEP) [90] and EGFR (PDB ID 1XKK) [117] (Figures 3, 4). Hence, kinase mutations with a high oncogenic potential may destabilize the autoinhibited kinase form. Importantly, oncogenic mutations could partly alleviate local frustration in the active kinase state (Figures 3, 4). A more extensive minimally frustrated network of interactions rigidifies the active form of the catalytic domain for ABL (PDB ID 1M52) [91,92] and EGFR (PDB ID 2J6M) [118]. Although the crystal structures employed in our

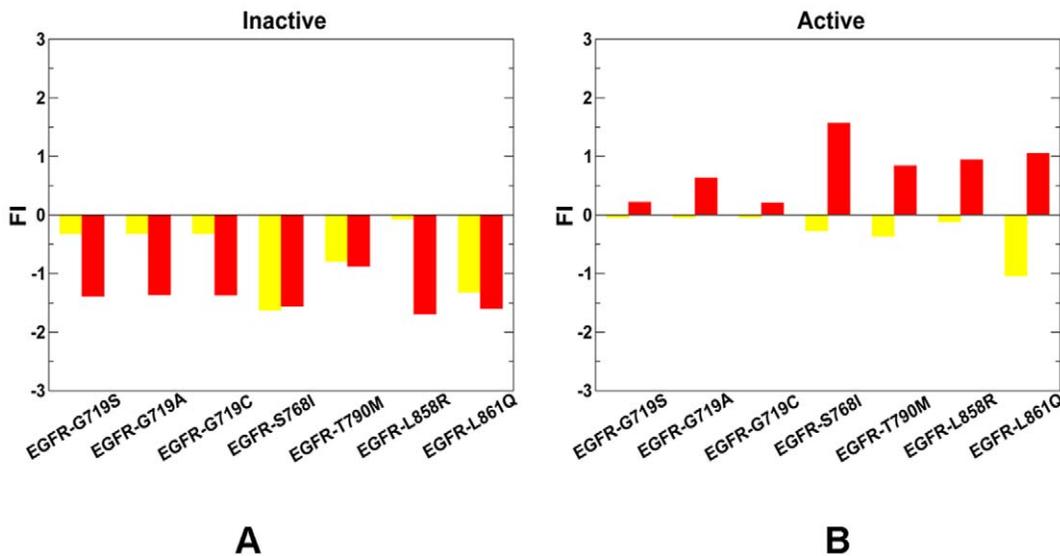
study (Table S1 in File S1) were mostly solved in the unbound form, there were some structures in the studied set (particularly ABL and EGFR kinases) that were originally crystallized in complexes with ATP or small molecule inhibitors. Since ATP binding could potentially increase structural rigidity of the catalytic domain, the local frustration analysis that included structures with a removed ATP may produce artificial changes in the frustration index for binding site residues. We evaluated the overall statistical distribution of the configurationally frustration index using crystal structures with bound molecules. The effect was found to be rather negligible and the resulting distribution was virtually indistinguishable from the one shown in Figure 1. Indeed, a significant fraction of the protein kinase residues involved in binding site interactions belong to the structurally rigid hinge region which is a minimally frustrated element of the catalytic core and as such robust to minor perturbations of interactions. However, we found some interesting small variations in the local frustration profiles of ABL (Figure S1) and EGFR kinases (Figure S2), which were observed for oncogenic mutations in the glycine-rich P-loop (ABL-G250E, ABL-Q252H, ABL-E255K, EGFR-G719S, EGFR-G719A, EGFR-G719C). It is known that the P-loop in ABL kinase may be stabilized in the Imatinib-bound inactive structure, which may explain the increased local frustration upon P-loop mutations in the inactive state (Figures S1,S2). Interestingly, these point mutations are known to impair the binding of Imatinib (Gleevec) to ABL by shifting the thermodynamic equilibrium towards the active form incompatible with the inhibitor binding [114-116]. Our data suggested that mutation-induced local frustration in the inhibitor-bound inactive kinase state may partly contribute to initiating a population shift between functional forms. We also analyzed the distribution and structural partition of minimally



**Figure 3. The effect of oncogenic mutations on local frustration in the ABL kinase.** The residue-based frustration index values are shown for a set of oncogenic ABL kinase mutants in the inactive (A) and active forms (B). The frustration index values are shown in filled yellow bars for the wild-type kinase form and in red filled bars for the mutant forms. The analysis was performed on the unbound form of the crystal structures of ABL in the inactive form (PDB ID 1IEP) [90] and active form (PDB ID 1M52) [91,92]. doi:10.1371/journal.pone.0026071.g003

frustrated and locally frustrated regions in the ABL kinase (Figure S3). A dense network of minimally frustrated residues was found in the structurally rigid core of the catalytic domain (connected by green lines). This minimally frustrated web was formed by structurally conserved  $\alpha$ F-helix and  $\alpha$ E-helix. In contrast, the clusters of locally frustrated residues (connected by red lines) assembled on the protein periphery, including the

$\alpha$ C-helix, activation loop, the P+1 loop in the C-terminal lobe. As the autoinhibiting interactions released in the active form, protein kinases could become more flexible with a considerable degree of residual local frustration. This was reflected in the increased presence of locally frustrated residues connected by red lines in the  $\alpha$ C-helix, and the C-terminal lobe of the active ABL (Figure S3B).



**Figure 4. The effect of oncogenic mutations on local frustration in the EGFR kinase.** The residue-based frustration index values are shown for a set of oncogenic EGFR kinase mutants in the inactive (A) and active forms (B). The frustration index values are shown in filled yellow bars for the wild-type kinase form and in red filled bars for the mutant forms. The analysis was performed on the unbound form of the crystal structures of EGFR in the inactive form (PDB ID 1XKK) [117] and active form (PDB ID 2J6M) [118]. doi:10.1371/journal.pone.0026071.g004

## Local Frustration and Protein Flexibility

We also investigated a relationship between local frustration and protein flexibility of kinase structures. In our previous studies, we have characterized the conformational landscapes of ABL, EGFR, RET and MET kinases as well as various cancer mutants using MD simulations of the Apo kinase and complexes with ATP and small molecule inhibitors [132,134,136]. Here, we compared the results of local frustration analysis with the kinase flexibility profiles, which were inferred from MD simulations and evaluated using the root mean square fluctuations (RMSF) of the catalytic domain residues. In particular, MD studies of ABL and EGFR kinases in the normal and oncogenic states displayed a high local flexibility in the lower portion of the activation loop [134,136]. Similarly, the bundle of  $\alpha$ -helices in the C-terminal, which represented the densest cluster of minimally frustrated residues (Figures S4,S5), also demonstrated the smallest variation in the RMSF values - a characteristic of structurally rigid protein core [134]. The local frustration profiles also matched up nicely with the B-factors of the protein kinase residues. An example of such comparative analysis was detailed for the EGFR-WT in the active form (Figure S5). A robust correlation was found between the residue-based local frustration index and the B-factor values (Figure S5 D). We also observed that the highly frustrated EGFR residues corresponded to the conformationally mobile regions with the higher B-factor values. To further illustrate these findings, we performed structural mapping of the average B-factors onto a set of inactive (Figure S5 A) and active structures (Figure S5 B) of ABL, EGFR, BTK, KIT, BRAF, MET, and RET kinases. Additionally, locally frustrated residues were also mapped onto the catalytic core. The locally frustrated sites corresponded to protein regions with the increased thermal mobility and overlapped with the protein residues of higher B-factors.

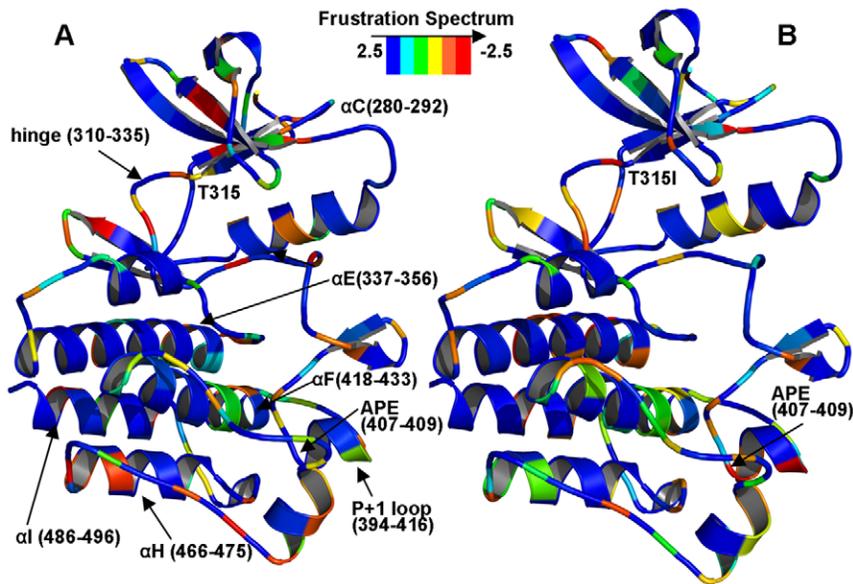
The analysis of protein kinase flexibility has also demonstrated that conformational changes in functionally important kinase regions may be allosterically coupled and highly correlated. More specifically, we found evidence of highly correlated protein motions and allosteric coupling of the  $\alpha$ C-helix and activation loop with all other kinase regions (Table S3 in File S1). Interestingly, the  $\alpha$ C-helix and the activation loop represented two most highly coupled protein kinase regions. Other highly correlated segments of the catalytic domain included (a) the hinge region and catalytic loop, and (b) the P-loop and activation loop. These findings consistent with our recent analysis of collective motions in ABL and EGFR regulatory complexes that manifested in “breathing” rigid body movements of the catalytic core coupled with the fluctuations of the P-loop, activation loop,  $\alpha$ C-helix and the  $\alpha$ G-helix of the C-terminal [136]. Numerous structural biology studies have also indicated a central involvement of the  $\alpha$ C-helix and activation loop in allosteric coupling that control regulation of protein kinase activity [110-113].

## Allosteric Effect of Oncogenic Mutations on Local Frustration

We investigated if spatial distribution of local frustration may present initiation points for global conformational changes and whether the effect of oncogenic mutations on the local frustration would be local or allosteric. If the effect of oncogenic mutations was local, it would cause only local perturbations and result in the negative values of the frustration index for residues in the immediate proximity of the mutational site. However, if the effect of oncogenic mutations was global, the spatial distribution of highly frustrated residues may be allosterically affected and result in noticeable changes at the remote from the mutational site regions. A comparison of locally frustrated residues mapping in the

ABL-WT and ABL-T315I mutant revealed subtle yet relevant changes, where most of the effected residues were remote from the mutational site (Figure 5). We observed that the gate-keeper mutation in the inactive kinase form may allosterically perturb structural rigidity of the catalytic core and increase local frustration of the  $\alpha$ F-helix,  $\alpha$ E-helix, and  $\alpha$ C-helix regions. Our findings corroborated with a hydrogen exchange mass spectrometry (HX MS) study of ABL kinase [100], indicating that the effect of the ABL-T315I mutation could result not only in local conformational disturbances near the  $\alpha$ C-helix, but also allosterically change protein flexibility in the distant from mutation protein regions. The changes in the local frustration induced by ABL-T315I mutation in the inactive kinase could be illustrated in examples presented in Figures S6, S7. While frustration plots of ABL-WT and ABL-T315I were generally similar, there were some changes in the red line clusters connecting Asp-381 of the DFG motif to Glu-286 which makes an important hydrogen bond with Lys-271 (Figure S6). Another change could be noted in the anti-parallel  $\beta$ -sheet from the lower part of the activation loop (Figure S7). In this region a few of the residues become highly frustrated upon mutation as evident by red lines connecting residues Tyr-393, Ala-395 and Pro-402.

Importantly, partial unfolding of the anti-parallel  $\beta$ -sheet at the lower end of the activation loop was previously determined as a prerequisite for stabilization of the intermediate Src-like structure and a common mechanistic feature of the ABL and EGFR activation pathways [134]. In the Src-like conformation,  $\alpha$ C-helix was rotated and moved out of the active site ( $\alpha$ C-helix-Glu-out position), the DFG motif flipped in the intermediate DFG-in position, the anti-parallel  $\beta$ -sheet from the lower part of the activation loop unfolded and the P-loop moved in towards the active site. These structural changes were accompanied by a concerted breakage of the K271-E286 ion pair and formation of the E286-R386 salt bridge. A conserved salt bridge between the K271 and E286 is a structural hallmark of the inactive and active forms of ABL, while it is absent in the Src-like inactive structure. The formation and breakage of this critical interaction coupled with the conformational changes in the DFG motif are critical structural features underlying mechanisms of kinase activation [134]. A mutation-induced development of local frustration in the DFG motif and the  $\beta$ -sheet of the activation loop could present the “initiation cracking points” [65-72] that would likely to perturb the inactive kinase form and facilitate conformational transitions between alternative kinase states. These results agree with the energy landscape analysis of adenylate kinase [66], in which the high stress region in the activation loop may “crack” or locally unfold releasing the strain and thus catalyzing a global conformational transformation. According to the “cracking” model [137-139], allosteric conformational changes can be triggered by the increased local frustration causing thermodynamic destabilization of a protein region via local unfolding. We found that mutation-induced perturbation of minimally frustrated interactions and amplification of protein flexibility in the inactive kinase state could be compounded by a partial reduction of local frustration and structural consolidation of the active kinase. Collectively these effects may present a feasible mechanism of kinase activation by cancer mutations via exploiting redistribution of local frustration to facilitate conformational transitions and enhance the thermodynamic stability of the constitutively active kinase. The proposed model is consistent with the energy landscape ideas according to which low local stability should accompany high local frustration and locally frustrated regions may act as local cracking points or specific hinges during allosteric changes [65-72,137-139].

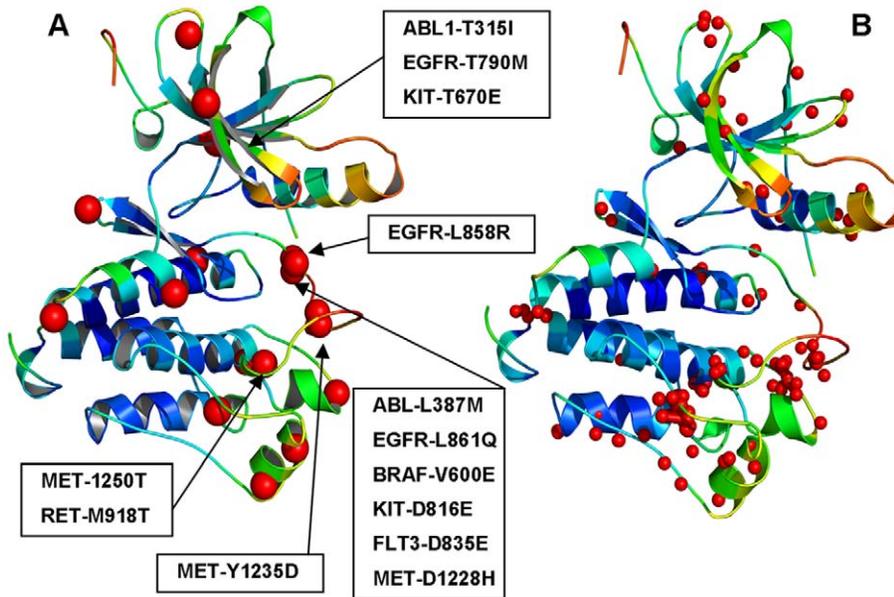


**Figure 5. The effect of oncogenic mutation on spatial distribution of local frustration in the ABL kinase.** The spatial distribution of local frustration in the inactive forms of ABL-WT (A) and ABL-T315I (B). The color sliding scheme of local frustration ranges from minimally frustrated (shown in blue) to highly frustrated (shown red). The key functional regions of the protein kinase along with the respective range of protein residues are referred to by arrows. Structural mapping of local frustration on the ABL kinase catalytic core is shown for the inactive ABL-WT structure (PDB ID 1IEP) [90]. The effect of T315I on the inactive ABL structure was evaluated via structural modeling detailed in the Materials and Methods section. The PyMol program was used for visualization of protein kinase structures and the local frustration mapping (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, and LLC). doi:10.1371/journal.pone.0026071.g005

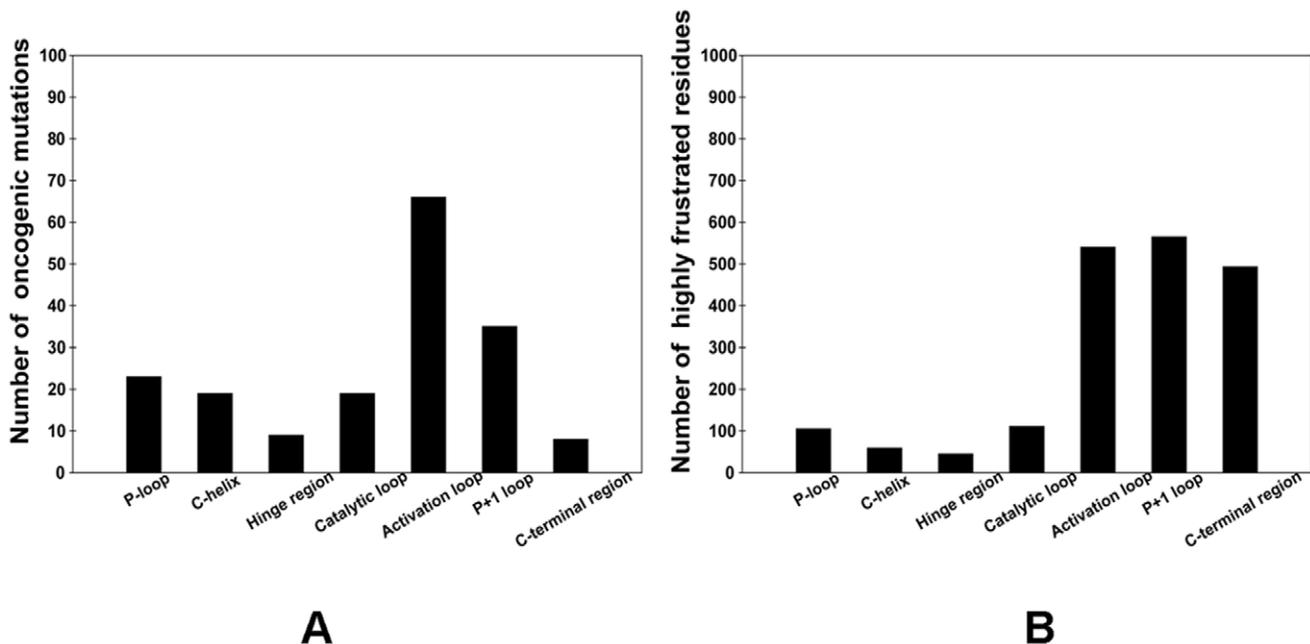
### Oncogenic Mutations as Allosteric Switches of Local Frustration

We proposed that locally frustrated kinase sites may catalyze large scale cooperative transitions by activating specific pathways of allosteric transformation, which may be modulated by cancer mutations. According to this conjecture, structural localization of kinase cancer mutations would be collocated with the locally frustrated sites. Structural bioinformatics analysis of protein kinases has previously revealed that highly oncogenic kinase mutations could fall at structurally conserved positions within the kinase catalytic core [123]. Moreover, these structurally conserved mutational hotspots could be shared by multiple kinase genes. To test our hypotheses, we performed a comparative analysis of the spatial distribution of highly oncogenic kinase mutations and highly frustrated residues mapped onto the kinase catalytic domain (Figure 6). We found that local frustration was not randomly scattered on the protein surface or uniformly distributed in the protein kinase structure. Interestingly, locally frustrated clusters could overlap with the kinase segments involved in allosteric interactions and collocate with the regions directly involved in conformational changes associated with the kinase function (Figure 6). In particular, our analysis revealed that the vast majority of locally frustrated sites resided in the C-terminal lobe, most notably populating the substrate binding region of the catalytic core framed by the  $\alpha$ F,  $\alpha$ G,  $\alpha$ H, and  $\alpha$ I helices, including the activation loop segment, and the P+1 loop (Figure 6). We have previously demonstrated that coupling between structurally rigid  $\alpha$ F-helix (minimally frustrated site) and conformationally adaptive  $\alpha$ I-helix,  $\alpha$ C-helix and the P+1 loop (more frustrated sites) may control allosteric activation in protein kinases [136]. Our present results indicated that highly frustrated residues could be localized near hinges (Figure 6) coordinating collective motions of kinase regions during allosteric conformational

changes. We also analyzed the distribution of known oncogenic mutations across catalytic core subdomains using a set of kinase oncogenes ABL, EGFR, BTK, KIT, BRAF, MET, and RET (Figure 7A). It appeared that this distribution was characterized by a bias towards specific functional regions, and functionally important activation loop along with the downstream P+1 loop region tend to be more densely populated than other subdomains. Other segments such as P-loop and catalytic loop could also harbor oncogenic mutations, but were less frequently populated by functionally important mutations. Parallel with this analysis, we carried out structural mapping of highly frustrated residues onto the kinase catalytic core and quantified the distribution of the local frustration index as a function of the kinase subdomain (Figure 7B). Importantly, the vast majority of highly frustrated kinase residues were mapped onto the C-terminal lobe, including the activation loop and regulatory P+1 loop. The locally frustrated residue clusters that populated the activation loop and the C-terminal lobe were collocated with the disease associated mutations and residues involved in allosteric conformational changes. Indeed, a relatively high concentration of highly frustrated residues in a single functional region is especially pronounced for the P+1 motif, which includes residues in the activation segment, and contains the conserved APE motif. The P+1 segment links the subdomains of the C-terminal lobe with the ATP and substrate binding regions in the N-terminal lobe. This segment is critical for protein substrate recognition and allosteric regulatory interactions [75-81], serving as a hydrophobic glue holding the sub-domains of the C-lobe together. The APE motif is involved in allosteric regulation, as it is anchored to the  $\alpha$ F,  $\alpha$ G and  $\alpha$ I-helices, providing direct communication between the activation segment and C-terminal. In addition, one of the highest concentrations of disease associated mutations localized in the vicinity of the P+1 pocket [120-123]. Functional role of these



**Figure 6. The spatial distribution of highly oncogenic mutations and highly frustrated residues in the kinase catalytic core.** The kinase mutations with known high oncogenic potential were mapped onto kinase catalytic domain (A). Structurally conserved hotspots of kinase cancer mutations are annotated by large red spheres and their location is indicated by arrows in (A). The locally frustrated sites ( $FI < -2.0$ ) mapped onto kinase catalytic domain and depicted as small red spheres (B). The crystal structure of EGFR-WT in the active form (PDB ID 2J6M) [118] was used as a template for structural mapping. The Pymol program was used for visualization of protein kinase structures and the local frustration mapping (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, and LLC).  
doi:10.1371/journal.pone.0026071.g006



**Figure 7. The distribution of oncogenic mutations and locally frustrated sites at each structural region in the catalytic core.** The distribution of cancer mutations (A) and the distribution of local frustration (B) across catalytic core subdomains. The analysis is performed based on known oncogenic mutations in the ABL, EGFR, BTK, KIT, BRAF, MET, and RET kinase genes. The residue ranges of the kinase subdomains (SD) were determined based on the ABL kinase crystal structure (PDB ID 1IEP) [90] as the reference and in accordance to our previous study [123]: SD1:242-261 (P-loop region); SD2:262-278; SD3:279-291 ( $\alpha$ C-helix); SD4:292-309; SD5:310-335 (hinge region); SD6A:336-356; SD6B:357-374 (catalytic loop); SD7:375-393 (activation loop); SD8:394-416 (P+1 loop); SD9:417-438; SD10:439-461; SD11:462-480; SD12:481-498. The C-terminal region encompasses SD8-SD12 subdomains.  
doi:10.1371/journal.pone.0026071.g007

residues as catalysts of kinase activation may be determined by their strategic location critical for regulation.

## Discussion

Activating kinase mutations result in a ligand-independent constitutive activation of the kinase activity. Among most prominent examples are activating mutations in the EGFR gene, where a single-point mutation L858R accounts for about 41% of all EGFR activating mutations [140,141]. Strikingly, recent functional studies have revealed the impaired nuclear EGFR accumulation in cells expressing EGFR-L858R may be due to the lack of allosteric activation rather than a direct consequence of constitutive kinase activity [142]. Hence, the primary functional effect of the activating EGFR-L858R mutation, which was shown to thermodynamically stabilize EGFR [134,143], is to allow for receptor activation that does not require the allosteric conformational change. The results of our current study corroborate with these central experimental findings. We observed a high density of locally frustrated residues in the regions involved in allosteric conformational transitions, particularly in the lower portion of the activation loop and the P+1 loop. While these locally frustrated sites could overlap with the mutational hotspots of disease-causing mutations, allosteric changes cannot occur if critical residues are mutated. This seeming contradiction may be partly explained based on the proposed functional role of locally frustrated sites as initiation points of allosteric transitions. Indeed, locally frustrated sites may trigger global structural changes via local rearrangements in the vicinity of pivotal hinge points and rigid body motions involving coupling of minimally frustrated and locally frustrated regions. The observed mutation-induced reduction of locally frustrated sites and thermodynamic stabilization of the active kinase form may thus help to suppress allosteric mechanism of activation.

This study suggested that the interplay between a minimally frustrated structural core and locally frustrated regions may collectively enable robust allosteric activation of protein kinases. Indeed, whereas a broad web of minimally frustrated residues in the kinase catalytic core could reflect robustness of the protein kinase fold to evolutionary pressure and mutations, the presence of locally frustrated residue clusters may not only be evolutionary tolerable but also potentially advantageous for tailoring protein kinase dynamics to maintain a dynamic equilibrium between alternative kinase states required for normal function. Diverse mechanisms of allosteric communication can span extreme cases, from a sequential model, where binding of a molecule at one site causes a sequential propagation of conformational changes across the protein to a fully cooperative model, where structural changes are tightly coupled and conformational switching is first-order phase transition. Our data seemed to support a “block-based” model of allosteric communication, according to which clusters of optimally interacting residues can recruit blocks of more flexible residues into communication pathways [32]. Although minimally frustrated residue clusters with optimized local interactions constitute the structurally rigid core of the kinase catalytic domain, locally frustrated residue clusters, whose interaction networks may not be energetically optimized, could define “soft spots”, that are weakly coupled to the kinase core and prone to dynamic modulation by mutations or binding.

In the present study, we combined computer simulations and the energy landscape analysis of protein kinases to characterize the interplay between oncogenic mutations, local frustration and protein flexibility as important catalysts of allosteric kinase activation and regulation. The results of this study suggested that mutation-induced allosteric signaling may involve a dynamic

coupling between structurally rigid (minimally frustrated) and plastic (locally frustrated) clusters of residues. We found that the energy landscape effect of oncogenic mutations may be allosteric in nature, eliciting global changes in the spatial distribution of highly frustrated residues. Furthermore, the protein kinase regions undergoing large structural changes during allosteric transitions could be enriched in clusters of highly frustrated residues. The present study indicated that activating cancer mutations could act as catalysts of kinase activation by simultaneously perturbing the network of minimally frustrated interactions in the inactive kinase state, while reducing local frustration and allosterically restoring structural stability in the active kinase form. Allosterically induced switch in the state of locally frustrated residues upon mutation can shift the thermodynamic equilibrium and “lock” the oncogenic kinase in a constitutively active form. This may present a feasible mechanism by which oncogenic mutations may function as catalysts of kinase activation by detrimentally affecting the thermodynamic equilibrium between kinase states. The energy landscape analysis of protein kinases and the proposed role of locally frustrated sites in activation mechanisms may have useful implications for bioinformatics-based screening and detection of functional sites critical for allosteric regulation in complex biomolecular systems. The results may be also potentially interesting for protein design, where rationale engineering of locally frustrated regions may provide means for probing activation mechanisms in a desired regime.

## Materials and Methods

### Protein Kinase Mutants

Protein kinase sequences were obtained from Kinbase (<http://kinase.com/kinbase/>). Sequence analysis of protein kinase mutations was done using data collected from different sources, including PupaSNP [144], dbSNP database [145], Online Mendelian Inheritance in Man (OMIM) from National Center for Biotechnology Information (NCBI) [146,147], KinMutBase [148,149], BTKbase [150], Human gene mutation database (HGMD) [151,152], Catalogue of Somatic Mutations in Cancer database (COSMIC) [153], Mutations of Kinases in Cancer (MoKCa) [154] SwissProt [155-157] Protein Kinase Resource (PKR) [158], and PDB [159]. The assembled set of somatic kinase mutations was categorized based on a quantitative metric of oncogenic potential corresponding to the frequency profiles of somatic mutations in the protein kinases genes obtained from the COSMIC repository [153]. Since only a subset of cancer mutations can be directly mapped onto the crystal structure of the catalytic domain, there are some protein kinase genes with the known WT crystal structures, yet no mutational models could be reliably produced, because either all known mutations reside outside of the resolved crystal structure of the kinase catalytic domain or only synonymous mutations were available. A collection of somatic kinase mutations that corresponded to the catalytic domain included ABL (36 mutations), EGFR (85 mutations), BTK (100 mutations), KIT (54 mutations), BRAF (62 mutations), MET (46 mutations), and RET (39 mutations) (Figure S8). To facilitate structure-functional analysis, we generated structural models of various protein kinase mutants using the respective WT crystal structure as a template (see Supporting Information in File S1). A total of 57 kinase genes that covered a wide range of kinase subfamilies were used in the present study (Table S1 in File S1).

### Analysis of Local Frustration in Protein Kinases

The protein kinase crystal structures as well as structural models of kinase mutants with the known WT crystal structure were used

in the calculation of the residue-based configurational frustration index. We focused on the local frustration analysis conducted for ABL, EGFR, BTK, KIT, BRAF, MET, and RET kinase genes based on simulations of these kinases in both the inactive and active forms (see Supporting Information in **File S1**). These kinase genes also account for the vast majority of highly oncogenic mutations in the catalytic domain. We computed residue-based configurational frustration index via a web server (<http://www.frustratometer.tk>). The local frustration analysis adapted a recently proposed method of quantifying the degree of frustration manifested by spatially local protein interactions [69]. The local frustration index for the contact between the amino acids  $i, j$  was defined as a Z-score of the energy of the native pair compared to the N decoys. According to the Ferreiro-Wolynes model, a residue-based frustration index can measure the energetic stability of a particular native contact as compared to a set of all possible contacts sampled by automatic generation of  $\sim 1000$  distributed decoys and recomputing the energy change. The frustration index can be calculated by mutating the identities and the distances between the interacting amino acids. In the mutational frustration index, the decoy set randomizes only the identities of the interacting amino acids  $i, j$  while keeping all other interaction parameters at their native value. We employed a more general configurational frustration index, where the decoy set involved randomizing not only the residue identities but also the distance between the interacting amino acids  $i, j$ . The index value that corresponded to a positive Z-score value would indicate that the majority of other amino acid pairs in that position were unfavorable. A contact was defined as minimally frustrated if its native energy was at the lower end of the distribution of decoy energies, and a frustration index as measured by a Z-score would be of 0.78 or higher magnitude. Conversely, a contact was defined as highly frustrated if its native energy was at the higher end of the distribution with a local frustration index lower than -1. If the native energy was in between these limits, the contact was defined as neutral.

### Structural Modeling of Protein Kinases

The protein kinase crystal structures corresponding to 57 kinase genes were collected from PDB [159] and were employed in the structural bioinformatics analysis and biophysical modeling (**Table S1** in **File S1**). To facilitate structure-functional analysis of genetic variations in kinase genes, all crystal structures and mutational models were structurally aligned using a java-based multiple alignment tool STRAP (<http://www.charite.de/bioinf/strap>) and TM-align algorithm [160]. Structural modeling of kinase mutants was carried out using MODELLER [161,162] with a subsequent refinement of side-chains by the SCRWL3 program [163]. Initial models were built in MODELLER using a flexible sphere of 5 Å around mutated residue. A protocol involving a conjugate gradient (CG) minimization, followed by simulated annealing refinement was repeated 20 times to generate 100 initial models for each studied mutant. The mutational models were chosen out of the 100 models as scored by the MODELLER default scoring function, followed by structural refinement using MD simulations protocol detailed in [123]. MD refinement simulations were done using NAMD 2.6 [164] with the CHARMM27 force field [165,166] and the explicit TIP3P water model as implemented in NAMD 2.6 [167]. The VMD program was used for the preparation and analysis of simulations [168,169]. Protein kinase flexibility was also analyzed by combining the results of MD simulations with the principal component analysis of conformational ensembles [170,171].

### Supporting Information

**Figure S1 The Effect of Oncogenic Mutations on Local Frustration in the Inhibitor-bound ABL Kinase Structures.** The residue-based frustration index values are shown for a set of oncogenic ABL kinase mutants in the inactive (A) and active forms (B). The frustration index values are shown in filled yellow bars for the WT kinase form and in red filled bars for the mutant forms. The analysis was performed using inhibitor-bound crystal structures of ABL in the inactive form (PDB ID 1IEP) [90] and active form (PDB ID 1M52) [91,92]. (TIF)

**Figure S2 The Effect of Oncogenic Mutations on Local Frustration in the Inhibitor-bound EGFR Kinase Structures.** The residue-based frustration index values are shown for a set of oncogenic EGFR kinase mutants in the inactive (A) and active forms (B). The frustration index values are shown in filled yellow bars for the WT kinase form and in red filled bars for the mutant forms. The analysis was performed using inhibitor-bound crystal structures of EGFR in the inactive form (PDB ID 1XKK) [117] and active form (PDB ID 2J6M) [118]. (TIF)

**Figure S3 The Energy Landscape of the ABL Kinase Catalytic Domain.** The spatial distribution and partition of minimally frustrated and locally frustrated regions in the inactive ABL kinase (A) and active ABL state (B). The protein backbone is displayed as blue ribbons, the direct residue interactions are shown with solid lines. Minimally frustrated interactions are shown in green, highly frustrated contacts in red, neutral contacts are not drawn. This analysis illustrated common similarities and differences in the local frustration of inactive and active kinase forms. Structural mapping of local frustration on the ABL kinase catalytic core is shown for the inactive ABL-WT structure (PDB ID 1IEP) [90] and active ABL-WT structure (PDB ID 1M52) [91,92]. The VMD program was used for protein kinase structure visualization [168,169]. (TIF)

**Figure S4 A Residue-based Comparative Analysis of Local Frustration and Protein Flexibility.** The values of the B-factors (A), the configurational frustration index FI (B), and the RMSF (C) for the protein kinase residues. The crystal structure of EGFR-WT in the active form (PDB ID 2J6M) [118] was used in this example of a comparative analysis. (TIF)

**Figure S5 Structural Mapping of B-factors and Locally Frustrated Sites onto the Kinase Catalytic Core.** Structural mapping of the average B-factors and locally frustrated residues onto a set of inactive (A) and active kinase structures (B). The set of inactive kinase conformations included ABL (PDB 1IEP), KIT (PDB ID 1T45), MET (PDB ID 2G15) and BRAF (PDB ID 1UWH). The set of active kinase conformations included EGFR (PDB ID 2J6M), BTK (PDB ID 1K2P), and RET (PDB ID 2IVS). The protein residues were colored accordingly to their B-factor (temperature factor) from dark blue for low B-factor to red for high B-factor. The locally frustrated residues are shown as red spheres. The PyMol program was used for visualization of protein kinase structures and the local frustration mapping (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, and LLC.). (TIF)

**Figure S6 Mutation-induced Redistribution of the Local Frustration in the ABL Kinase DFG Motif.** The mutation-

induced changes in the local frustration between ABL-WT (A) and ABL-T315I (B). (TIF)

**Figure S7 Mutation-induced Redistribution of the Local Frustration in the ABL Kinase Activation Loop.** The mutation-induced changes in the local frustration between ABL-WT (A) and ABL-T315I (B). (TIF)

**Figure S8 Gene-based Pie Diagram of Kinase Cancer Mutations.** The arc length of each sector is proportional to the number of cancer mutations for a given kinase gene. For clarity of presentation, only top 70 kinase genes with the cancer-causing mutations that can be mapped onto three-dimensional structure of the catalytic core are presented. (TIF)

## References

- Koshland DE, Némethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5: 365–385.
- Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: A plausible model. *J Mol Biol* 12: 88–118.
- Cui Q, Karplus M (2008) Allostery and cooperativity revisited. *Protein Sci* 17: 1295–1307.
- Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* 4: 474–482.
- Xu C, Tobi D, Bahar I (2003) Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T<->R2 transition. *J Mol Biol* 333: 153–168.
- Zheng W, Brooks B (2005) Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. *J Mol Biol* 346: 745–759.
- Dima RI, Thirumalai D (2006) Determination of network of residues that regulate allostery in protein families using sequence analysis. *Protein Sci.* 15: 258–268.
- Zheng W, Brooks BR, Thirumalai D (2006) Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc Natl Acad Sci U S A* 103: 7664–7669.
- Hyeon C, Lorimer GH, Thirumalai D (2006) Dynamics of allosteric transitions in GroEL. *Proc Natl Acad Sci U S A* 103: 18939–18944.
- Stan G, Lorimer GH, Thirumalai D, Brooks BR (2007) Coupling between allosteric transitions in GroEL and assisted folding of a substrate protein. *Proc Natl Acad Sci U S A* 104: 8803–8808.
- Zheng W, Brooks BR, Thirumalai D (2007) Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. *Biophys J* 93: 2289–2299.
- Chen J, Dima RI, Thirumalai D (2007) Allosteric communication in dihydrofolate reductase: signaling network and pathways for closed to occluded transition and back. *J Mol Biol* 374: 250–266.
- Chennubhotla C, Bahar I (2006) Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. *Mol Syst Biol* 2: 36.
- Chennubhotla C, Bahar I (2007) Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput Biol* 3: 1716–1726.
- Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* 17: 633–640.
- Chennubhotla C, Yang Z, Bahar I (2008) Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol Biosyst* 4: 287–292.
- Inis B, Schulten K, Tajkhorshid E, Bahar I (2008) Mechanism of signal propagation upon retinal isomerization: insights from molecular dynamics simulations of rhodopsin restrained by normal modes. *Biophys J* 95: 789–803.
- Zheng W, Thirumalai D (2009) Coupling between normal modes drives protein conformational dynamics: illustrations using allosteric transitions in myosin II. *Biophys J* 96: 2128–2137.
- Zheng W, Tekpinar M (2009) Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model. *BMC Struct Biol* 9: 45.
- Zheng W, Brooks BR, Thirumalai D (2009) Allosteric transitions in biological nanomachines are described by robust normal modes of elastic networks. *Curr Protein Pept Sci* 10: 128–132.
- Yang L, Song G, Jernigan RL (2009) Protein elastic network models and the ranges of cooperativity. *Proc Natl Acad Sci U S A* 106: 12347–12352.
- Yang Z, Majek P, Bahar I (2009) Allosteric transitions of supramolecular systems explored by network models: application to chaperonin GroEL. *PLoS Comput Biol* 5: e1000360.
- Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286: 295–299.
- Suel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 10: 59–69.
- Del-Sol AH, Fujihashi H, Amoros D, Nussinov R (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol Syst Biol* 0019.
- Del Sol, AH, Arauzo Bravo MJ, Amoros D, Nussinov R (2007) Modular architecture of protein structures and allosteric communications: potential implications for signaling proteins and regulatory linkages. *Genome Biol* 8: R92.
- Zandany NM, Ovadia MI, Orr I, Yifrach O (2008) Direct analysis of cooperativity in multisubunit allosteric proteins. *Proc Natl Acad Sci U S A* 105: 11697–702.
- Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins* 57: 433–443.
- Tsai CJ, Sol AD, Nussinov R (2008) Allostery: absence of a change in shape does not imply that allostery is not at play. *J Mol Biol.* 378: 1–11.
- Tsai CJ, Sol AD, Nussinov R (2009) Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol Biosyst* 5: 207–216.
- Sol AD, Sol, Tsai CJ, Ma B, Nussinov R (2009) The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure* 17: 1042–1050.
- Kidd BA, Baker D, Thomas WE (2009) Computation of conformational coupling in allosteric proteins. *PLoS Comput Biol* 5: e1000484.
- Onuchic JN, Schulten LZ, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem.* 48: 545–600.
- Socci ND, Onuchic JN, Wolynes PG (1998) Protein folding mechanisms and the multidimensional folding funnel. *Proteins* 32: 136–158.
- Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14: 70–75.
- Mirny L, Shakhnovich E (2001) Protein folding theory: from lattice to all-atom models. *Annu Rev Biophys Biomol Struct* 30: 361–396.
- Onuchic JN, Nymeyer H, Garcia AE, Chahine J, Socci ND (2000) The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv Protein Chem* 53: 87–152.
- Plotkin SS, Onuchic JN (2002) Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q Rev Biophys* 35: 111–167.
- Plotkin SS, Onuchic JN (2002) Understanding protein folding with energy landscape theory. Part II: Quantitative aspects. *Q Rev Biophys* 35: 205–286.
- Wolynes PG (2005) Energy landscapes and solved protein-folding problems. *Philos Transact A Math Phys Eng Sci* 363: 453–467.
- Wolynes PG (2005) Recent successes of the energy landscape theory of protein folding and function. *Q. Rev. Biophys.* 38: 405–410.
- Shakhnovich E (2006) Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. *Chem Rev* 106: 1559–1588.
- Zhuravlev PI, Papoian GA (2010) Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. *Q Rev Biophys.* 43: 295–332.
- Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* 84: 7524–7528.
- Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21: 167–195.
- Ma B, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. *Protein Eng* 12: 713–720.
- Tsai CJ, Kumar S, Ma B, Nussinov R (1999) Folding funnels, binding funnels and protein function. *Protein Sci* 8: 1181–1190.

**File S1 Supporting tables.**  
(DOC)

## Acknowledgments

We are very grateful to Prof. Peter G. Wolynes (Rice University) for his encouragement, support and stimulating discussions. We also thank Dr. Diego U. Ferreira (Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina) for his contribution, technical expertise, and fruitful discussions at the initial stages of this project.

## Author Contributions

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

48. Tsai CJ, Ma B, Nussinov R (1999) Folding and binding cascades: shifts in energy landscapes. *Proc Natl Acad Sci U S A* 96: 9970–9972.
49. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* 9: 10–19.
50. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* 97: 8868–8873.
51. Levy Y, Wolynes PG, Onuchic JN (2004) Protein topology determines binding mechanism. *Proc Natl Acad Sci U S A* 101: 5111–5116.
52. Verkhiyker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST, et al. (2002) Complexity and simplicity of ligand-macromolecule interactions: the energy landscape perspective. *Curr Opin Struct Biol* 12: 197–203.
53. Ma J, Karplus M (1998) The allosteric mechanism of the chaperonin GroEL: a dynamic analysis. *Proc Natl Acad Sci U S A* 95: 8502–8507.
54. Ma J, Sigler PB, Xu Z, Karplus M (2000) A dynamic model for the allosteric mechanism of GroEL. *J Mol Biol* 302: 303–313.
55. Kong Y, Ma J, Karplus M, Lipscomb WN (2006) The allosteric mechanism of yeast chorismate mutase: A dynamic analysis. *J Mol Biol* 356: 237–247.
56. Formanek MS, Ma L, Cui Q (2006) Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins* 63: 846–867.
57. Yu H, Ma L, Yang Y, Cui Q (2007) Mechanochemical coupling in the myosin motor domain. I. insights from equilibrium active-site simulations. *PLoS Comput Biol* 3: e21.
58. Yu H, Ma L, Yang Y, Cui Q (2007) Mechanochemical coupling in the myosin motor domain. II. analysis of critical residues. *PLoS Comput Biol* 3: e23.
59. Ma L, Cui Q (2007) Activation mechanism of a signaling protein at atomic resolution from advanced computations. *J Am Chem Soc* 129: 10261–10268.
60. Cecchini M, Houdusse A, Karplus M (2008) Allosteric communication in myosin V: From small conformational changes to large directed movements. *PLoS Comput Biol* 4: e1000129.
61. Pan H, Lee JC, Hiler VJ (2000) Binding sites in *Escherichia coli* dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc Natl Acad Sci U S A* 97: 12020–12025.
62. Liu T, Whitten ST, Hiler VJ (2007) Functional residues serve a dominant role in mediating the cooperativity of the protein ensemble. *Proc Natl Acad Sci U S A* 104: 4347–4352.
63. Sayar K, Uğur O, Liu T, Hiler VJ, Onaran O (2008) Exploring allosteric coupling in the alpha-subunit of heterotrimeric G proteins using evolutionary and ensemble-based approaches. *BMC Struct Biol* 8: 23.
64. Onaran HO, Costa T (2009) Allosteric coupling and conformational fluctuations in proteins. *Curr Protein Pept Sci* 10: 110–115.
65. Latzer J, Shen T, Wolynes PG (2008) Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles. *Biochemistry* 47: 2110–2122.
66. Miyashita O, Onuchic JN, Wolynes PG (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci U S A* 100: 12570–12575.
67. Okazaki KI, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci U S A* 103: 11844–11849.
68. Okazaki KI, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci U S A* 105: 11182–11187.
69. Ferreira DU, Hegler JA, Komives EA, Wolynes PG (2007) Localizing frustration in native proteins and protein assemblies. *Proc Natl Acad Sci U S A* 104: 19819–19824.
70. Sutto L, Lätzer J, Hegler JA, Ferreira DU, Wolynes PG (2007) Consequences of localized frustration for the folding mechanism of the IM7 protein. *Proc Natl Acad Sci U S A* 104: 19825–19830.
71. Li W, Wolynes PG, Takada S (2011) Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proc Natl Acad Sci U S A* 108: 3504–3509.
72. Ferreira DU, Hegler JA, Komives EA, Wolynes PG (2011) On the role of frustration in the energy landscapes of allosteric proteins. *Proc Natl Acad Sci U S A* 108: 3499–3503.
73. Hanks SK, Hunter T (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 9: 576–596.
74. Hunter T, Plowman GD (1997) Review: the protein kinases of budding yeast: six score and more. *Trends Biochem Sci* 22: 18–22.
75. Manning G, Whyte DB, Martínez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. *Science* 298: 1912–1934.
76. Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 10: 514–520.
77. Hunter T (2000) Signaling – 2000 and beyond. *Cell* 100: 113–127.
78. Pawson T, Scott JD (2005) Protein phosphorylation in signaling—50 years and counting. *Trends Biochem Sci* 30: 2862–2890.
79. Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. *Cell* 109: 275–282.
80. Nolen B, Taylor SS, Ghosh G (2004) Regulation of protein kinases. Controlling activity through activation segment conformation. *Molecular Cell* 15: 661–675.
81. Taylor SS, Kornev AP (2011) Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem Sci* 36: 65–77.
82. Shi Z, Resing KA, Ahn NG (2006) Networks for the allosteric control of protein kinases. *Curr Opin Struct Biol* 16: 686–692.
83. Pellicena P, Kuriyan J (2006) Protein-protein interactions in the allosteric regulation of protein kinases. *Curr Opin Struct Biol* 16: 702–709.
84. Kannan N, Neuwald AF (2005) Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *J Mol Biol* 351: 956–972.
85. Kannan N, Haste N, Taylor SS, Neuwald AF (2007) The hallmark of AGC kinase functional divergence is its C-terminal tail, a cis-acting regulatory module. *Proc Natl Acad Sci U S A* 104: 1272–1277.
86. Kornev AP, Haste NM, Taylor SS, Eyck LF (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc Natl Acad Sci U S A* 103: 17783–17788.
87. Kornev AP, Taylor SS, Ten Eyck LF (2008) A helix scaffold for the assembly of active protein kinases. *Proc Natl Acad Sci U S A* 105: 14377–14382.
88. Zhang J, Yang PL, Gray NS (2009) Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* 9: 28–39.
89. Knight ZA, Lin H, Shokat KM (2010) Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer* 10: 130–137.
90. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, et al. (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* 289: 1938–1942.
91. Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, et al. (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res* 62: 4236–4243.
92. Tokarski JS, Newitt JA, Chang CY, Cheng JD, Wittekind M, et al. (2006) The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against Imatinib-resistant ABL mutants. *Cancer Res* 66: 5790–5797.
93. Modugno M, Casale E, Soncini C, Rosettani P, Colombo R, et al. (2007) Crystal structure of the T315I Abl mutant in complex with the aurora kinases inhibitor PHA-739358. *Cancer Res* 67: 7987–7990.
94. Zhou T, Parillon L, Li F, Wang Y, Keats J, et al. (2007) Crystal structure of the T315I mutant of Abl kinase. *Chem Biol Drug Des* 70: 171–181.
95. Azam M, Seeliger MA, Gray NS, Kuriyan J, Daley GQ (2008) Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat Struct Mol Biol* 15: 1109–1118.
96. Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, et al. (2004) Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* 116: 855–867.
97. Mol CD, Dougan DR, Schneider TR, Skene RJ, Kraus ML, et al. (2004) Structural basis for the autoinhibition and STI-571 inhibition of c-kit tyrosine kinase. *J Biol Chem* 279: 31655–31663.
98. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, et al. (2002) Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Biol* 9: 268–272.
99. Knight ZA, Gonzalez B, Feldman ME, Zunder ER, Goldenberg DD, et al. (2006) A pharmacological map of the PI3-K family defines a role for p110alpha in insulin signaling. *Cell* 125: 733–747.
100. Iacob RE, Pene-Dumitrescu T, Zhang J, Gray NS, Smithgall TE, et al. (2009) Conformational disturbance in Abl kinase upon mutation and deregulation. *Proc Natl Acad Sci U S A* 106: 1386–1391.
101. Adrian FJ, Ding Q, Sim T, Velentza A, Sloan C, et al. (2006) Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nat Chem Biol* 2: 95–102.
102. Zhang J, Adrian FJ, Jahnke W, Cowan-Jacob SW, Li AG, et al. (2010) Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* 463: 501–506.
103. Iacob RE, Zhang J, Gray NS, Engen JR (2011) Allosteric interactions between the myristate- and ATP-site of the Abl kinase. *PLoS One* 6: e15929.
104. Nagar B, Hantschel O, Young MA, Scheffzek K, Veach DR, et al. (2003) Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* 112: 859–871.
105. Nagar B, Hantschel O, Seeliger M, Davies JM, Weis WI, et al. (2006) Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase. *Mol Cell* 21: 787–798.
106. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* 125: 1137–1149.
107. Jura N, Endres NF, Engel K, Deindl S, Das R, et al. (2009) Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell* 137: 1293–1307.
108. Red Brewer M, Choi SH, Alvarado D, Moravcevic K, Pozzi A, et al. (2009) The juxtamembrane region of the EGF receptor functions as an activation domain. *Mol Cell* 34: 641–651.
109. Jura N, Shan Y, Cao X, Shaw DE, Kuriyan J (2009) Structural analysis of the catalytically inactive kinase domain of the human EGF receptor 3. *Proc Natl Acad Sci U S A* 106: 21608–21613.
110. Dawson JP, Bu Z, Lemmon MA (2007) Ligand-induced structural transitions in ErbB receptor extracellular domains. *Structure* 15: 942–954.
111. Lemmon MA, Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. *Cell* 141: 1117–11134.

112. Bae JH, Schlessinger J (2010) Asymmetric tyrosine kinase arrangements in activation or autophosphorylation of receptor tyrosine kinases. *Mol Cells* 29: 443–448.
113. Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T, et al. (2011) Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. *Mol Cell* 42: 9–22.
114. Knight ZA, Shokat KM (2005) Features of selective kinase inhibitors. *Chem Biol* 12: 621–637.
115. Liu Y, Gray NS (2006) Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* 2: 358–364.
116. Okram B, Nagle A, Adrian FJ, Lee C, Ren P, et al. (2006) A general strategy for creating inactive-conformation" abl inhibitors. *Chem Biol* 13: 779–786.
117. Wood ER, Truesdale AT, McDonald OB, Yuan D, Hassell A, et al. (2004) A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res* 64: 6652–6659.
118. Yun CH, Boggon TJ, Li Y, Woo MS, Greulich H, et al. (2007) Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* 11: 217–227.
119. Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, et al. (2008) The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A* 105: 2070–2075.
120. Torkamani A, Schork NJ (2007) Accurate prediction of deleterious protein kinase polymorphisms. *Bioinformatics* 23: 2918–2925.
121. Torkamani A, Schork NJ (2008) Prediction of cancer driver mutations in protein Kinases. *Cancer Res* 68: 1675–1682.
122. Torkamani A, Kannan N, Taylor SS, Schork NJ (2008) Congenital disease SNPs target lineage specific structural elements in protein kinases. *Proc Natl Acad Sci USA* 105: 9011–9016.
123. Dixit A, Yi L, Gowthaman R, Torkamani A, Schork NJ, Verkhivker G (2009) Sequence and structure signatures of cancer mutation hotspots in protein kinases. *PLoS One* 4: e7485.
124. Young MA, Gonfloni S, Superti-Furga G, Roux B, Kuriyan J (2001) Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* 105: 115–126.
125. Banavali NK, Roux B (2007) Anatomy of a structural pathway for activation of the catalytic domain of Src kinase Hck. *Proteins* 67: 1096–1112.
126. Yang S, Roux B (2008) Src kinase conformational activation: Thermodynamics, pathways, and mechanisms. *PLoS Comput Biol* 4: e1000047.
127. Yang S, Banavali NK, Roux B (2009) Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. *Proc Natl Acad Sci U S A* 106: 3776–3781.
128. Arora K, Brooks CL, 3<sup>rd</sup> (2007) Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci U S A* 104: 18496–18501.
129. Shan Y, Seeliger MA, Eastwood MP, Frank F, Xu H, et al. (2009) A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc Natl Acad Sci U S A* 106: 139–144.
130. Berteotti A, Cavalli A, Branduardi D, Gervasio FL, Recanatini M, et al. (2009) Protein conformational transitions: the closure mechanism of a kinase explored by atomistic simulations. *J Am Chem Soc* 131: 244–250.
131. Zou J, Wang YD, Ma FX, Xiang ML, Shi B, et al. (2008) Detailed conformational dynamics of juxtamembrane region and activation loop in c-Kit kinase activation process. *Proteins* 72: 323–332.
132. Dixit A, Torkamani A, Schork NJ, Verkhivker G (2009) Computational modeling of structurally conserved cancer mutations in the RET and MET kinases: the impact on protein structure, dynamics, and stability. *Biophys J* 96: 858–874.
133. Papakyriakou A, Vourloumis D, Tzortzidou-Stathopoulou F, Karpusas M (2008) Conformational dynamics of the EGFR kinase domain reveals structural features involved in activation. *Proteins* 76: 375–386.
134. Dixit A, Verkhivker G (2009) Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations. *PLoS Comput Biol* 5: e1000487.
135. Mustafa M, Mirza A, Kannan N (2011) Conformational regulation of the EGFR kinase core by the juxtamembrane and C-terminal tail: a molecular dynamics study. *Proteins* 79: 99–114.
136. Dixit A, Verkhivker G (2011) Computational modeling of allosteric communication reveals organizing principles of mutation-induced signaling in ABL and EGFR kinases. *PLoS Comput Biol*. In press.
137. Whitford PC, Miyashita O, Levy Y, Onuchic JN (2007) Conformational transitions of adenylate kinase: switching by cracking. *J Mol Biol* 366: 1661–1671.
138. Whitford PC, Onuchic JN, Wolynes PG (2008) Energy landscape along an enzymatic reaction trajectory: hinges or cracks? *HFSP J* 2: 61.
139. Zhang HJ, Sheng XR, Pan XM, Zhou JM (1997) Activation of adenylate kinase by denaturants is due to the increasing conformational flexibility at its active sites. *Biochem Biophys Res Commun* 238: 382–386.
140. Shigematsu H, Gazdar AF (2006) Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int J Cancer* 118: 257–262.
141. Gazdar AF (2009) Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* 28 (Suppl1): S24–S31.
142. Liccardi G, Hartley JA, Hochhauser D (2011) EGFR nuclear translocation modulates DNA repair following cisplatin and ionizing radiation treatment. *Cancer Res* 71: 1103–1114.
143. Yang S, Park K, Turkson J, Arteaga CL (2008) Ligand-independent phosphorylation of Y869 (Y845) links mutant EGFR signaling to stat-mediated gene expression. *Exp Cell Res* 314: 413–419.
144. Conde L, Vaquerizas JM, Santoyo J, Al-Shahrour F, Ruiz-Llorente S, et al. (2004) PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. *Nucleic Acids Res* 32: W242–W248.
145. Sherry ST, Ward M, Sirotkin K (1999) dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 9: 677–679.
146. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, et al. (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 36: D13–D21.
147. Rebholz-Schuhmann D, Marcel S, Albert S, Tolle R, Casari G, et al. (2004) Automatic extraction of mutations from Medline and cross-validation with OMIM. *Nucleic Acids Res* 32: 135–142.
148. Stenberg KA, Riikonen PT, Vihinen M (2000) KinMutBase, a database of human disease-causing protein kinase mutations. *Nucleic Acids Res* 28: 369–371.
149. Ortutay C, Väliäho J, Stenberg K, Vihinen M (2005) KinMutBase: a registry of disease-causing mutations in protein kinase domains. *Hum Mutat* 25: 435–442.
150. Väliäho J, Smith CI, Vihinen M (2006) BTKbase: the mutation database for X-linked agammaglobulinemia. *Hum Mutat* 27: 1209–1217.
151. Krawczak M, Ball EV, Fenton I, Stenson PD, Abeyasinghe S, et al. (2000) Human gene mutation database – a biomedical information and research resource. *Hum Mut* 15: 45–51.
152. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, et al. (2003) Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 21: 577–581.
153. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, et al. (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 91: 355–358.
154. Richardson CJ, Gao Q, Mitsopoulous C, Zvebil M Pearl LH, et al. (2009) MoKCa database—mutations of kinases in cancer. *Nucleic Acids Res* 37: D824–D831.
155. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, et al. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31: 365–370.
156. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A (2007) UniProtKB/Swiss-Prot. *Methods Mol Biol* 406: 89–112.
157. The UniProt Consortium (2008) The universal protein resource (UniProt). *Nucleic Acids Res* 36: D190–D195.
158. Niedner RH, Buzko OV, Haste NM, Taylor A, Gribskov M, et al. (2006) Protein kinase resource: an integrated environment for phosphorylation research. *Proteins* 63: 78–86.
159. Kouranov A, Xie L, de la Cruz J, Chen L, Westbrook J, et al. (2006) The RCSB PDB information portal for structural genomics. *Nucleic Acids Res* 34: D302–D305.
160. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33: 2302–2309.
161. Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo A, et al. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29: 291–325.
162. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9: 1753–1773.
163. Canutescu AA, Shelenkov AA, Dunbrack RL, Jr. (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* 12: 2001–2014.
164. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802.
165. MacKerell AD, Jr., Bashford D, Bellott M, Dunbrack RL, Jr., Evanseck JD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102: 3586–3616.
166. MacKerell AD, Jr., Banavali N, Foloppe N (2001) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 56: 257–265.
167. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML, et al. (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935.
168. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33–8, 27–28.
169. Eargle J, Wright D, Luthey-Schulten Z (2006) Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics* 22: 504–506.
170. Amadei A, Linssen AB, Berendsen HJ (1993) Essential dynamics of proteins. *Proteins* 17: 412–425.
171. Zhou X, Chou J, Wong ST (2006) Protein structure similarity from Principle Component Correlation analysis. *BMC Bioinformatics* 7: 40.