

# Sequence and Structure Signatures of Cancer Mutation Hotspots in Protein Kinases

Anshuman Dixit<sup>1,2</sup>, Lin Yi<sup>1</sup>, Ragul Gowthaman<sup>1</sup>, Ali Torkamani<sup>3</sup>, Nicholas J. Schork<sup>3</sup>, Gennady M. Verkhivker<sup>1,2,4\*</sup>

**1** Graduate Program for Bioinformatics, Center for Bioinformatics, The University of Kansas, Lawrence, Kansas, United States of America, **2** Department of Pharmaceutical Chemistry, School of Pharmacy, The University of Kansas, Lawrence, Kansas, United States of America, **3** Scripps Genomic Medicine, Department of Molecular and Experimental Medicine, Scripps Health and The Scripps Research Institute, La Jolla, California, United States of America, **4** Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America

## Abstract

Protein kinases are the most common protein domains implicated in cancer, where somatically acquired mutations are known to be functionally linked to a variety of cancers. Resequencing studies of protein kinase coding regions have emphasized the importance of sequence and structure determinants of cancer-causing kinase mutations in understanding of the mutation-dependent activation process. We have developed an integrated bioinformatics resource, which consolidated and mapped all currently available information on genetic modifications in protein kinase genes with sequence, structure and functional data. The integration of diverse data types provided a convenient framework for kinome-wide study of sequence-based and structure-based signatures of cancer mutations. The database-driven analysis has revealed a differential enrichment of SNPs categories in functional regions of the kinase domain, demonstrating that a significant number of cancer mutations could fall at structurally equivalent positions (mutational hotspots) within the catalytic core. We have also found that structurally conserved mutational hotspots can be shared by multiple kinase genes and are often enriched by cancer driver mutations with high oncogenic activity. Structural modeling and energetic analysis of the mutational hotspots have suggested a common molecular mechanism of kinase activation by cancer mutations, and have allowed to reconcile the experimental data. According to a proposed mechanism, structural effect of kinase mutations with a high oncogenic potential may manifest in a significant destabilization of the autoinhibited kinase form, which is likely to drive tumorigenesis at some level. Structure-based functional annotation and prediction of cancer mutation effects in protein kinases can facilitate an understanding of the mutation-dependent activation process and inform experimental studies exploring molecular pathology of tumorigenesis.

**Citation:** Dixit A, Yi L, Gowthaman R, Torkamani A, Schork NJ, et al. (2009) Sequence and Structure Signatures of Cancer Mutation Hotspots in Protein Kinases. *PLoS ONE* 4(10): e7485. doi:10.1371/journal.pone.0007485

**Editor:** Kumar Selvarajoo, Keio University, Japan

**Received:** July 16, 2009; **Accepted:** September 25, 2009; **Published:** October 16, 2009

**Copyright:** © 2009 Dixit et al. 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 was supported by The University of Kansas start-up funding. 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: verk@ku.edu

## Introduction

A central goal of cancer research involves the discovery and functional characterization of the mutated genes that drive tumorigenesis [1]. The Human Genome Project has provided researchers with unprecedented insights into the structure and organization of genes. Large-scale resequencing and polymorphism characterization studies have subsequently focused on the identification and cataloguing of naturally occurring gene and sequence variation [2–5]. The Cancer Genome Atlas and related DNA sequencing initiatives have specifically investigated the genetic determinants of cancer [6]. These studies have determined that only a fraction of genetic alterations contributing to tumorigenesis may be inherited, while somatically acquired mutations can contribute decisively during the progression of a normal cell to a cancer cell. Protein kinases play a critical role in cell signaling and have emerged as the most common protein domains that are implicated in cancer [7–11]. Although the kinase catalytic domain is highly conserved, protein kinase crystal structures have revealed considerable structural differences

between closely related active and highly specific inactive forms of kinases [12–17]. Evolutionary conservation and conformational plasticity of the kinase catalytic domain allow for a dynamic equilibrium between active and inactive kinase forms, which can facilitate regulation of the catalytic activity [15–17]. There are more than 500 protein kinases encoded in the human genome and many members of this family are prominent therapeutic targets for combating diseases caused by abnormalities in signal transduction pathways, especially various forms of cancer [18–22].

The complete sequencing of the human genome and high-throughput generation of genomic data have opened up avenues for a systematic approach to understanding the complex biology of cancer and clinical targeting of activated oncogenes. Large-scale tumor sequencing studies have identified a rich source of naturally occurring mutations in the protein kinase genes with many being simple single nucleotide polymorphisms (SNPs) [23–32]. A subset of these SNPs could occur in the coding regions (cSNPs) and lead to the same polypeptide sequence (synonymous SNPs, sSNPs) or result in a change in the encoded amino acid sequence (nonsynonymous coding SNP, nsSNPs). Resequencing studies of

the kinase coding regions in tumors have classified tumor-associated somatic mutations revealing that only a small number of kinase mutations may contribute to tumor formation (known as cancer driver mutations) while the majority could be neutral mutational byproducts of somatic cell replication (known as passenger mutations) [23–28]. While protein kinases have a prominent role in tumorigenesis, commonly mutated protein kinases in cancer appeared to be the exception to the rule and most of kinase driver mutations are expected to be distributed across many protein kinase genes [27]. Cancer mutations in protein kinases could often exemplify the phenomenon of oncogene addiction whereby, despite the accrual of numerous genetic alterations over the maturation of a tumor, cancer cells could remain reliant upon particular oncogenic pathways and may become addicted to the continued activity of specific activated oncogenes [33,34]. The dominant oncogenes that confer the oncogene addiction effect include ABL, EGFR, VEGFR, BRAF, FLT3, RET, and MET kinase genes [34].

The recent discovery of lung cancer mutations in the EGFR kinase domain [35–37] and their differential sensitivity to EGFR inhibitors have suggested that genetic alterations may be associated with structural changes, rendering tumors sensitive to selective inhibitors. Structural determinations of the EGFR [38–41] and ABL cancer mutants [42,43] have suggested that molecular mechanisms of kinase activation by cancer mutations and activity signatures of cancer drugs may be associated with the dynamics of functional transitions between inactive and active kinase forms. Biophysical modeling of protein kinase structure and dynamics has revealed important mechanistic features of kinase activation at atomic resolution. Molecular dynamics (MD) simulations of large-scale conformational transitions have been performed for many therapeutically important protein kinases, including HCK kinase [44], adenylate kinase [45], Src kinase [46–51], cyclin-dependent kinase 5 (CDK5) [52], ABL kinase [53], KIT kinase [54] EGFR, RET and MET kinase domains [55–57]. These studies have suggested that cancer mutations can have a subtle, yet profoundly important functional affect not only on local conformational changes at the mutational site, but also on allosteric regulation and cooperative interactions in signal transduction networks [58,59]. According to the proposed mechanism of kinase activation, structural effect of cancer mutations could manifest in shifting the dynamic equilibrium between inactive and active kinase forms towards a constitutively active kinase, thereby causing deleterious consequences for kinase regulation.

Cancer biology studies of protein kinase genes have integrated genetic, structural and functional approaches to characterize underlying molecular signatures of cancer mutations. High-throughput DNA sequence analysis and functional assessment of candidate cancer mutations in the tyrosine kinase genes have identified point mutations in the conserved hot spots from the activation loop in leukemia-associated tyrosine kinases [60–63]. A high-throughput platform has been used to interrogate the entire FLT3 coding sequence in AML patients and experimentally test the functional consequences of each candidate tumorigenic allele [63]. These studies have indicated that rare driver variants could often occur at frequencies indistinguishable from passenger mutations. As a result, functional analysis of candidate mutations identified in genome-wide screens can be ultimately required to determine which mutations contribute to cell transformation. Computational approaches, when combined with structural and functional studies, have also facilitated the identification and prediction of candidate cancer genes and individual alleles contributing to tumorigenesis [64–67].

Bioinformatics tools were recently developed to distinguish between driver and passenger nsSNPs [68,69]. Though quite

powerful, generalized prediction methods may fail to achieve the sensitivity and specificity attainable by prediction models tailored to individual protein families. We have developed kinase-targeted machine learning models that focused on nsSNPs in protein kinases by leveraging known sequence-based and structure-based protein kinase features to identify patterns in residues and sequence motifs harboring functionally relevant variations [70–72]. The developed support-vector machine (SVM) method has been shown to differentiate between disease-associated nsSNPs and neutral nsSNPs with ~80% accuracy [70]. These findings have suggested that the predictive power of machine learning models in assessing functionally important mutations can be significantly enhanced by selecting informative attributes characteristic of a specific protein family. Furthermore, we have found that kinase regions harboring a large number of cancer mutations in multiple protein kinases could contain a high proportion of the predicted driver mutations, while kinase subdomains devoid of cancer mutations were more likely to contain passenger mutations [71,72]. These results have suggested that biological characteristics and functional consequences separating cancer driver mutations from passenger mutations in protein kinases may differ from those separating disease-associated from neutral nsSNPs across the entire genome.

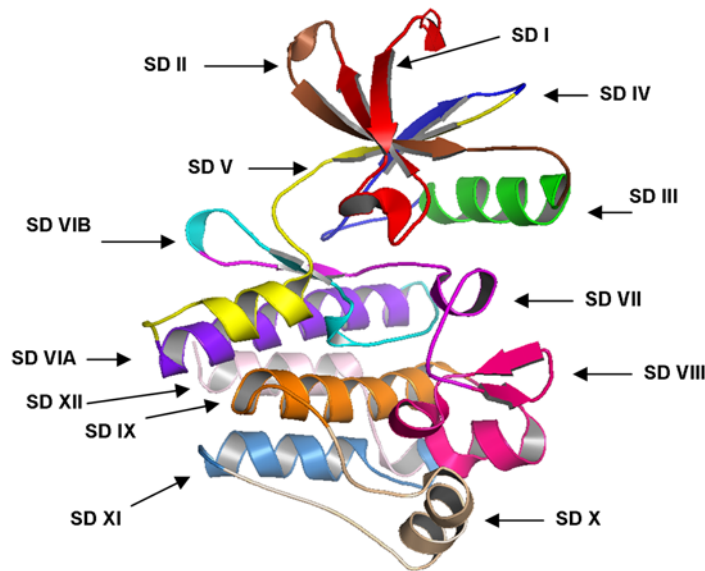
The growing body of genetic, molecular and functional information about protein kinase genes, combined with their prominent role as therapeutic targets for cancer intervention have produced an unprecedented explosion of diverse data. A large amount of information about genetic modifications in protein kinase families has been accumulated in different sources, including PupaSNP [73], dbSNP database [74], Online Mendelian Inheritance in Man (OMIM) from National Center for Biotechnology Information (NCBI) [75,76], KinMutBase [77,78], BTKbase [79], Human gene mutation database (HGMD) [80,81], Catalogue of Somatic Mutations in Cancer database (COSMIC) [82], Protein Kinase Resource (PKR) [83], and Mutations of Kinases in Cancer (MoKCa) [84]. While current databases and information portals have accumulated a large amount of information on kinase SNPs, there is a growing need for integration and comprehensive mapping of diverse data categories on protein kinase genes within a central resource.

In this work, we introduce Composite Kinase Mutation Database (CKMD), a single repository and integrated bioinformatics resource that consolidated and unequivocally mapped all currently available information on genetic variations in protein kinase genes with sequence, structural and functional data. CKMD and web-based resource are freely available at <http://verklab.bioinformatics.ku.edu/database/>. The functionality and capabilities of CKMD portal can allow for robust functional annotation of protein kinase genes and enable kinome-wide prediction and structure-functional analysis of cancer mutations. The database-driven analysis of sequence and structure-based signatures of kinase SNPs has clarified salient aspects of sequence conservation patterns and structural profiles of cancer-causing mutations, including the emergence of structurally conserved tumorigenic hotspots across multiple protein kinases. Furthermore, structural modeling and energetic analysis of kinase cancer mutations, which constitute the largest mutational hotspot, have provided useful insights into a common mechanism of kinase activation.

## Results

### Sequence-Structure Classification and Mapping of Kinase SNPs

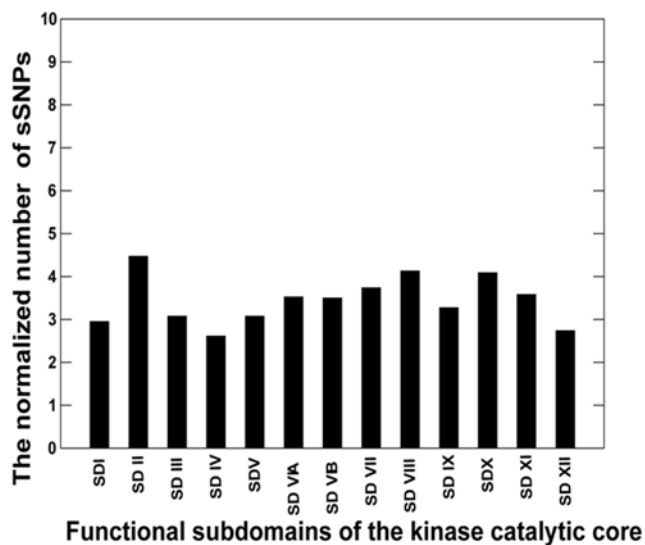
The integration and mapping of diverse data types in CKMD provided a convenient framework for kinome-wide analysis of



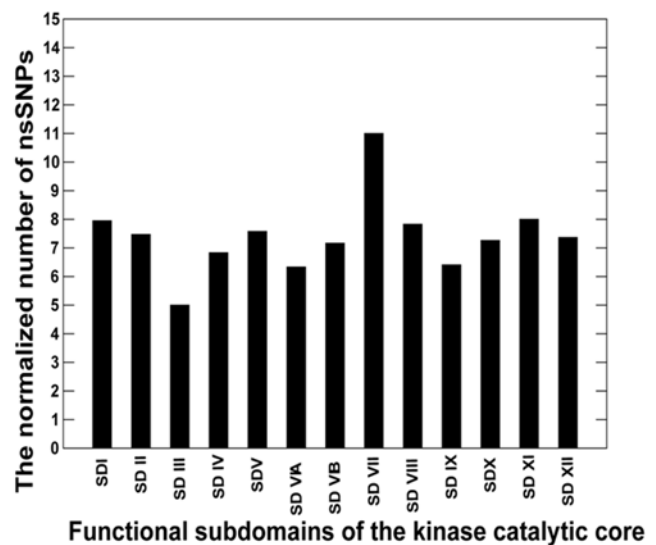
**Figure 1. Functional Subdomains of the Kinase Catalytic Core.** The kinase catalytic domain was subdivided into 12 subdomains (SD) using the ABL kinase crystal structure (pdb entry 1IEP) as the reference for defining the residue ranges as follows : SD I:242–261(P-loop region); SD2 :262–278; SD3:279–291( $\alpha$ C-helix); SD4:292–309; SD5:310–335 (hinge region); SD6A:336–356; SD6B357–374 (catalytic loop); SD7:375–393 (activation loop) ; SD8:394–416 (P+I loop); SD9:417–438; SD10:439–461; SD11:462–480; SD12:481–498. The alignment of functional subdomains for protein kinase genes was done using structure-informed multiple sequence alignment.  
doi:10.1371/journal.pone.0007485.g001

sequence-based and structure-based signatures of cancer mutations. Genetic variations in protein kinase genes are widely spread across both phylogenetic and structural space, and only a subset of all SNPs could be directly mapped to the kinase catalytic domain. We began by analyzing the distribution of various SNPs categories that could be mapped onto the 12 functional subdomains (SDs) of the kinase catalytic core [7] (**Figure 1**). Structural mapping of sSNPs resulted in a uniform coverage of kinase subdomains,

showing only a weak preference towards SD II which has no obvious functional role in kinase regulation (**Figure 2A**). In contrast, the distribution of nsSNPs highlighted the preferential bias towards specific functional regions. Indeed, functionally important P-loop (SD I), hinge region (SD V), catalytic loop (SD VIB), and especially activation loop (SD VII) along with the downstream P+I loop region (SD VIII) tend to be more densely populated (**Figure 2B**). The P+I segment links the subdomains in



A



B

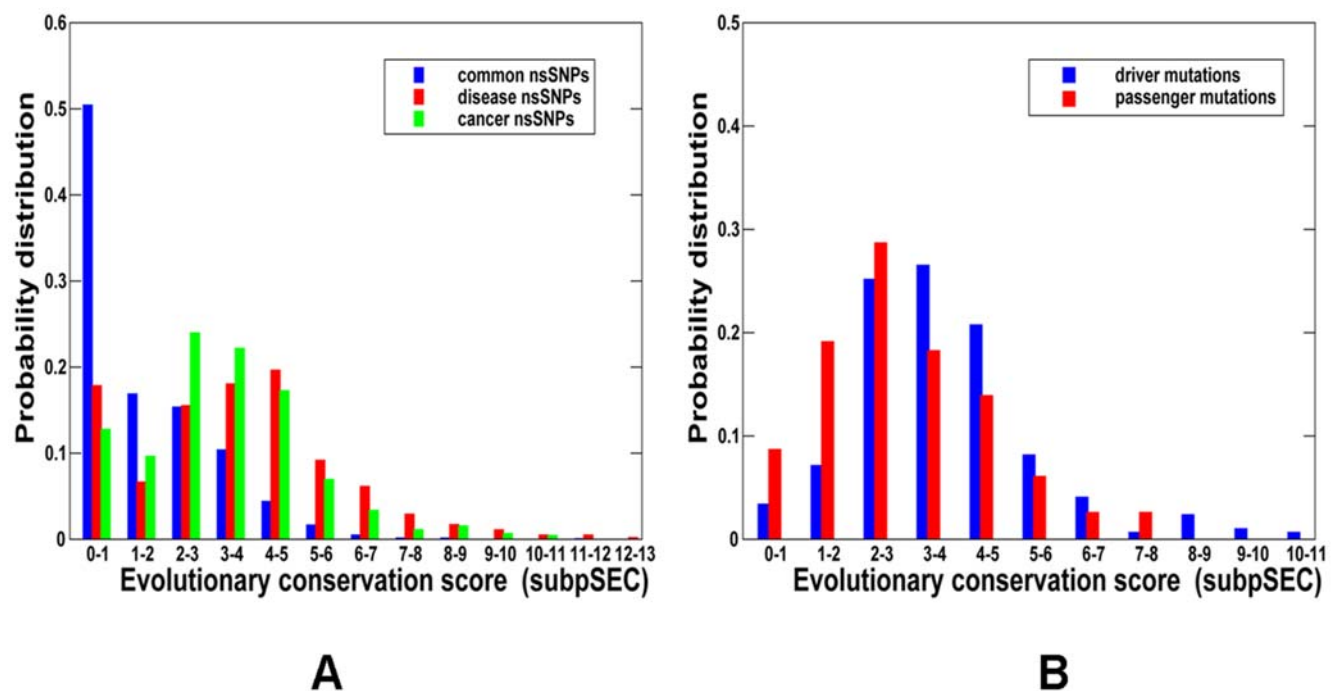
**Figure 2. The Distribution of SNPs Types across Functional Subdomains of the Kinase Catalytic Core.** The distribution of kinase sSNPs is shown in panel (A) and the distribution of nsSNPs is presented in panel (B).  
doi:10.1371/journal.pone.0007485.g002

the C-terminal lobe with the ATP and substrate binding regions in the N-terminal lobe. Moreover, the P+1 loop is directly connected to the F-helix, which serves as a central scaffold in the assembly of active kinase form [85–87].

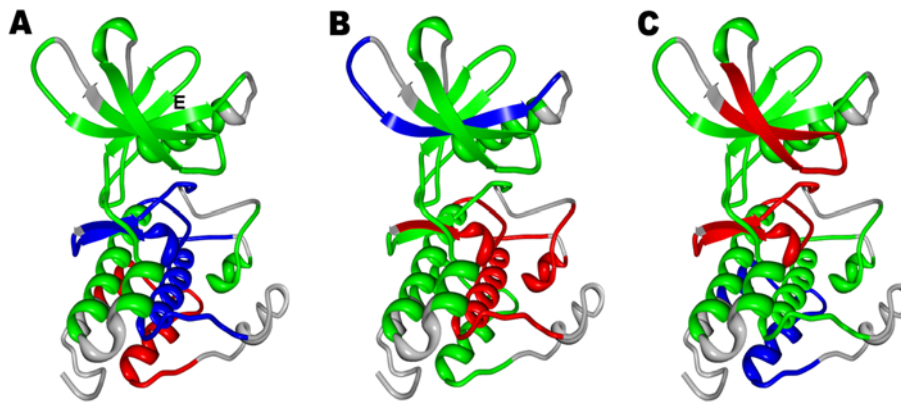
The kinase catalytic domain harbors a significant number of nsSNPs falling into three major categories: common and likely neutral nsSNPs, inherited disease-causing nsSNPs, and cancer-causing (somatic) nsSNPs. We analyzed evolutionary conservation patterns among these three different categories of kinase nsSNPs (Figure 3). A measure of conservation was derived from the absolute value of the substitution position-specific evolutionary conservation score, termed “subPSEC,” which was obtained by aligning a given protein against a library of Hidden Markov Models (HMM) representing distinct protein families [88,89]. The score was defined as  $-\ln(P_{a_{ij}}/P_{b_{ij}})$ , where  $P_{a_{ij}}$  is the probability of observing amino acid  $a$  at position  $i$  in HMM  $j$ . According to the PANTHER website [89], a score of -3 would correspond to an estimated 50% probability that the SNP may be a disease causing variant. The SNPs conservation profiles for kinase genes could be described as the absolute value of subPSEC score, where the higher the score, the greater the degree of evolutionary conservation. The distribution of common nsSNPs was biased towards a lower level of conservation, as would be expected for neutral variants with little or no functional significance. Cancer-associated nsSNPs appeared to fall into positions with a higher level of conservation than common nsSNPs, yet could be as conserved as disease-causing nsSNPs (Figure 3A). This analysis indicated that either cancer-associated nsSNPs may not necessarily fall into evolutionary highly conserved positions, or the distribution may be skewed towards a lower conservation level by cancer variants of no functional consequence (passenger mutations). Using a recently developed SVM-based method capable of predicting functionally important cancer mutations [70,71], we

compared the evolutionary conservation distributions of cancer driver mutations and passenger mutations at different levels of conservation (Figure 3B). Although the predicted cancer driver mutations did fall at the positions exhibiting slightly higher conservation level, as compared to the passenger mutations, the difference was rather modest. Hence, it appeared that cancer mutations in protein kinases may not display strong sequence conservation signals and consequently, functional importance of kinase genetic variants may not be directly related with their evolutionary conservation.

We also analyzed molecular determinants of genetic variations in protein kinases utilizing CKMD resource for a comprehensive structural mapping of nsSNPs onto the kinase catalytic core. The database-driven analysis revealed a differential enrichment of SNPs categories in functional regions of the kinase domain (Figures 4, 5). Common nsSNPs tend to be randomly distributed within the catalytic core, only sparsely populating functional segments of the catalytic core, such as the catalytic or activation loops, whereas these nsSNPs more densely occupy evolutionary unconserved regions of the C-terminal tail (Figure 4A). The disease-causing nsSNPs primarily mapped to the regions involved in regulation and substrate binding, such as the APE-loop and the P+1 region, as well as the catalytic loop (Figure 4B). Cancer-associated nsSNPs tend to target regions directly involved in the catalytic activity that are mainly localized in the P-loop, activation loop and catalytic loop (Figures 4C). The distribution of kinase nsSNPs across functional kinase subdomains reinforced the notion that the kinase regions that are enriched (or devoid) of SNPs could be markedly different across the three mutation types, with a minimal overlap. Indeed, the distribution shows a clear preference for cancer-causing nsSNPs to accumulate mostly in the activation loop region (SDVII) as well as populating the P-loop (SD I) (Figure 5A). A significant number of disease-associated nsSNPs



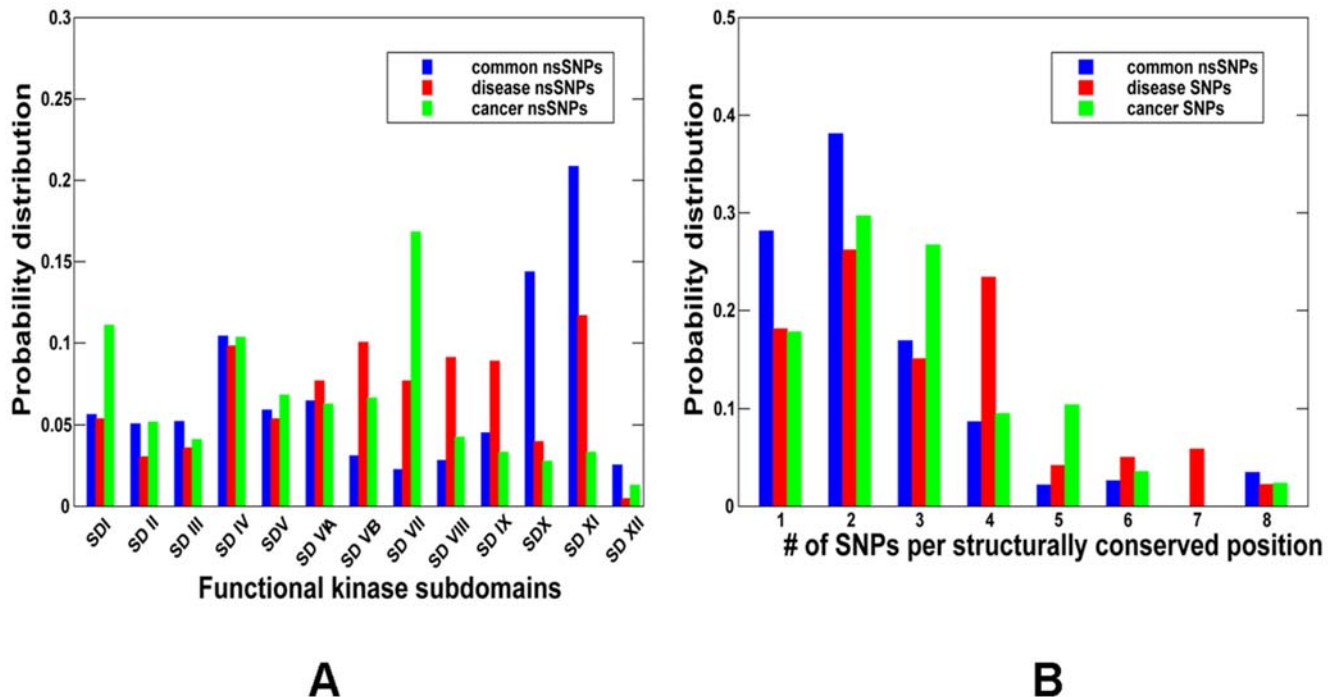
**Figure 3. The Distribution of nsSNPs Types across Evolutionary Conservation Levels.** (A) The probability distribution of common nsSNPs (shown in blue bars), disease-causing SNPs (shown in red bars) and cancer-causing nsSNPs (shown in green bars) as a function of evolutionary conservation level. (B) The probability distribution of cancer driver mutations (shown in blue bars) and passenger nsSNPs (shown in red bars) as a function of evolutionary conservation level. For both panels (A) and (B), a higher score corresponds to a higher level of conservation. doi:10.1371/journal.pone.0007485.g003



**Figure 4. Structural Mapping of nsSNPs onto the Kinase Catalytic Domain.** Structural mapping is shown for common nsSNPs (A), disease-causing nsSNPs (B), and cancer-causing nsSNPs (C). In all panels the green coloration represents regions with a SNP frequency equivalent to what would be expected by random chance, blue coloration represents regions that are statistically devoid of SNPs, and red coloration depicts regions that are statistically enriched in SNPs. Enrichment of SNPs in these regions was calculated as described in the Materials and Methods section. For clarity, the SNPs density was mapped onto a representative kinase crystal structure (EGFR, pdb entry 1M14) by projecting the multiple sequence kinase alignment onto the protein structure.  
doi:10.1371/journal.pone.0007485.g004

were not directly involved in the ATP binding, but rather buried in the catalytic core. Interestingly, the P+1 loop and the residues that anchor this pocket to the F-helix were some of the most enriched in disease-associated mutations, but not cancer-causing mutations. These results corroborate with previous findings indicating that disease-associated mutations could primarily affect the kinase regions involved in functional regulation, allosteric interactions and substrate binding [72].

Functional differences across different mutation types could be also reflected in the position-specific distribution of nsSNPs at the mutational hotspots determined by the number of structurally equivalent protein kinase positions (**Figure 5B**). The distribution of common nsSNPs, that have little or no functional affect and could be randomly distributed throughout the catalytic core, was dominated by weakly conserved positions mutated in a single, or two protein kinases. In contrast, the disease-causing nsSNPs tend



**Figure 5. The Distribution of nsSNPs Types across Functional Subdomains of the Catalytic Core.** (A) The distribution of common nsSNPs (shown in blue bars), disease-causing nsSNPs (shown in red bars), and cancer-causing nsSNPs (shown in green bars) in the functional subdomains of the kinase catalytic core. The expected probability of a SNP occurring in a kinase subdomain region was calculated for each SNP type as described in the Materials and Methods section. (B) The position-specific distribution of common nsSNPs (shown in blue bars), disease-causing nsSNPs (shown in red bars), and cancer-associated nsSNPs (shown in green bars) across different categories of structurally conserved mutational hotspots as determined by the number of SNPs per structurally identical position.  
doi:10.1371/journal.pone.0007485.g005

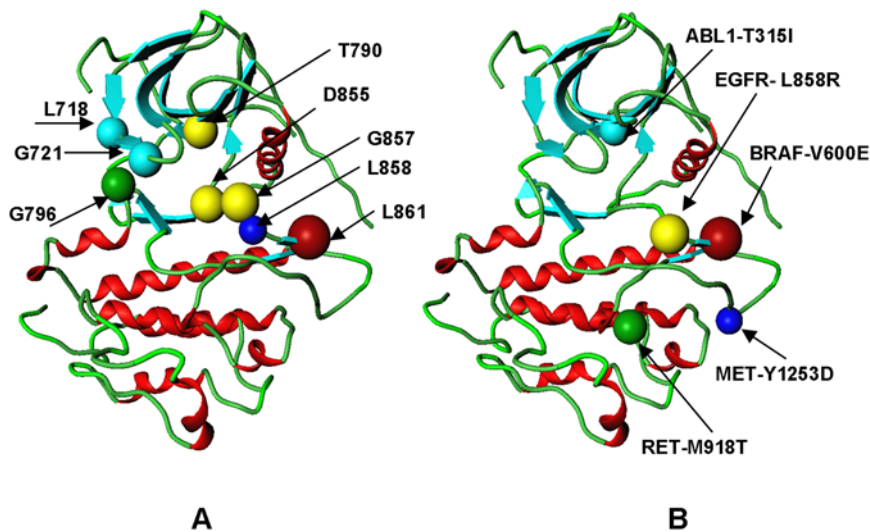
to be concentrated at structurally equivalent positions, with a significant excess of mutations occurring at positions mutated in four or more different protein kinases. The position-specific distribution of cancer nsSNPs was shifted towards a higher number of nsSNPs per position, probably due to the selection of tumorigenic mutational hotspots shared across multiple protein kinases (Figure 5B).

### Structural Bioinformatics Analysis of Kinase Mutational Hotspots

Kinome-wide analysis of sequence and structure-based signatures of cancer mutations, revealed that a significant number of cancer mutations could fall at structurally equivalent positions within the catalytic core. These structurally conserved mutations tend to cluster into specific mutational hotspots which may be shared by multiple kinase genes. Cancer mutation hotspots in protein kinases are largely localized within the P-loop, hinge region, and activation loop (Figure 6A, Table S1). Of special interest is a spectrum of EGFR, ABL, MET, FLT3 and KIT cancer mutations that correspond to the same structurally conserved position in the activation loop, which appeared to be mutated in at least 8 different kinases (Figure 6A, Table S1). This site corresponds to the known driver mutations BRAF-V600, FLT3-D835, KIT-D816, PDGFR $\alpha$ -D842, MET-D1228, EGFR-L861, ABL-L387, and ErbB2-L869. Despite a sequence-specific conservation pattern, many mutations at this structurally conserved position are commonly occurring activating mutations, including D1228H/N/V in MET [90,91], D835E/F/H/N/V/Y in FLT3 [92,93], D816E/F/H/N/I/V/Y in KIT [94,95] and V600D/E/G/K/L/M/R in BRAF [96]. In some cases, these

mutations could have important implications for targeted inhibitor therapies by leading to drug resistance effects in KIT [97], BRAF [98], EGFR [99], ABL [100], and MET [101]. Another functionally important mutational hotspot corresponds to the conserved gate-keeper kinase position and includes ABL-T315I, EGFR-T790M, KIT-T670E, and PDGFR $\alpha$ -T674I variants (Figure 6A, Table S1). Some of the structurally equivalent positions could be conserved across the kinome, as the aspartate and glycine residues from the DFG motif (corresponding to the reference positions EGFR-D855 and EGFR-G857), as well as a conserved glycine in the hinge region (which corresponds to the EGFR-G796 reference position). There are examples of cancer mutations displaying a subgroup level of conservation, including EGFR-L858 position, which bears a conserved leucine in EGFR and ABL kinases, or a conserved aspartate shared in FLT3, KIT, MET, PDGFR $\alpha$ .

While most of the cancer driver mutations are likely to be rather rare, it is striking that a significant number of functionally important cancer mutants fall at structurally conserved positions within the kinase catalytic core. Moreover, we have observed that structurally conserved hotspots of cancer driver mutations often bear mutations with a high oncogenic activity (Figure 6B). A quantitative characterization of “oncogenicity” could be described in a variety of ways, including cell transformation potential, substrate utilization, and catalytic efficiency. However such data are typically available only for a limited number of genes and mutations and are not suitable for genome-wide analysis. We used a convenient definition of an oncogenic potential that may be offered by using the frequency profiles of somatic mutations in the protein kinases genes obtained from the COSMIC repository [82]. This analysis revealed that a rather small number of somatic



**Figure 6. Structurally Conserved Mutational and Oncogenic Hotspots in the Kinase Catalytic Domain.** (A) Structural localization of the conserved mutational hotspots is illustrated using the crystal structure of the active EGFR kinase (pdb entry 2J6M). The large-size red ball corresponds to the structural position of L861, and denotes localization of the largest mutational hotspot shared in 8 different kinases. The medium-size yellow balls correspond to structural positions of T790, D855, and G857 residues (respective mutational hotspots shared by 6 different kinases). The smaller green ball corresponds to G796 position (5 structurally conserved kinase mutations); the cyan balls correspond to L718 and G721 positions (each position denote residues with 4 cancer mutations); and the smallest blue ball corresponds to L858 position (3 structurally conserved kinase mutations). Cancer mutation hotspots in protein kinases are largely localized within the P-loop, hinge region, and activation loop. See also Table S1 for a comprehensive annotation of structurally conserved mutational hotspots. (B) Structural localization of cancer driver mutations with the high oncogenic potential is illustrated using the crystal structure of the active EGFR kinase (pdb entry 2J6M). The dominant oncogenic mutations are BRAF-V600E, KIT-D816V, and PDGFR $\alpha$ -D842V which all correspond to the same structurally conserved mutational hotspot. Structural annotation of cancer driver mutations is arranged according to their oncogenic potential as determined by the frequency of observing respective somatic mutations in the protein kinases genes. The higher the oncogenic potential of the cancer drive, the larger the ball denoting structural position of the respective mutation.

doi:10.1371/journal.pone.0007485.g006

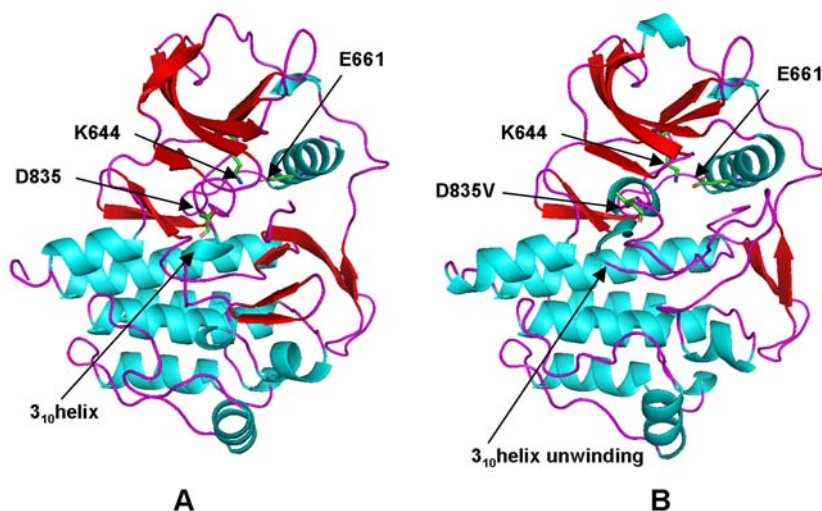
kinase mutations with the known oncogenic potential could emerge with a high frequency in the mutational samples (**Table S2**). Strikingly, these functionally important mutations fall into major structurally conserved positions in the kinase catalytic domain. Indeed, highly oncogenic mutations BRAF-V600E, KIT-D816V, and PDGFRA-D842V belong to the largest mutational hotspot (**Figure 6B**). The functional importance of oncogenic kinase mutations from mutational hotspots such as ABL-T315I, EGFR-L858R, and RET-M918T, is also widely recognized. For instance, structurally conserved RET-M918T and MET-M1250T cancer drivers are situated in the substrate binding C-lobe of the kinase core (**Figure 6B**) and are known to be associated with oncogenic activation by displaying the highest transforming potential among known RET [102–106] and MET mutations [107–110]. The presented analysis suggests that structurally conserved hotspots in the kinase catalytic domain may be statistically enriched by mutations with a high probability of being cancer drivers. We argue that the preferential structural localization of oncogenic mutations in the activation loop and the substrate binding C-lobe of the kinase domain may be determined by their strategic location critical for the kinase autoinhibition, regulation and allosteric interactions in signal transduction networks.

### Structural and Energetic Signatures of Kinase Mutational Hotspots

Structural modeling and energetic analysis of cancer mutation effects can provide further insights into molecular mechanisms of kinase activation. We employed homology modeling and MD simulations to analyze whether structurally conserved cancer drivers that target the same tumorigenic hotspot in the kinase catalytic domain may also share a common activation mechanism. Molecular modeling focused on a quantitative comparison of MET-D1228V, MET-D1228H [90,91], FLT3-D835V, FLT3-D835E [92,93], and KIT-D816V, KIT-D816H [94,95] mutants. Substitutions of D835 in FLT3 and D816 in KIT result in the constitutive activation of the receptor, this residue has been suggested to play an important regulatory role. The crystal structures of FLT3 [111], KIT [112] and MET kinases

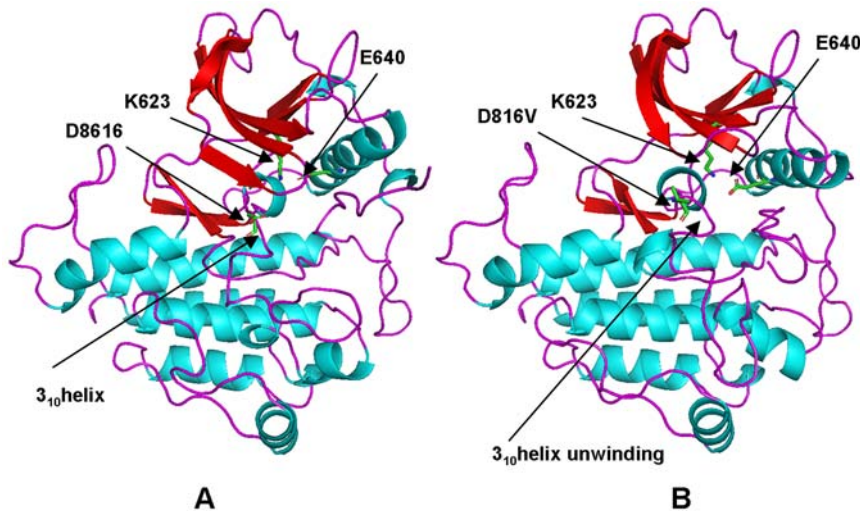
[113,114] have suggested that cancer mutations may destabilize the autoinhibited wild-type (WT) form. It is important to note that structural modeling studies were performed to evaluate the extent of local perturbations that could be induced by cancer mutations on the autoinhibited kinase structure. Given the absence of high resolution crystal structures of kinase cancer mutants and nature of large conformational changes caused by activating mutations, we focused on understanding local functional effects of cancer mutations rather than attempting to make computational predictions of the mutant structures.

Homology modeling and MD simulations of commonly occurring activating mutations in this mutational hotspot revealed a significant local reorganization of the autoinhibited kinase conformation. This is reflected in the local structural variations near the site of mutation (root mean square deviations, RMSD = 3 Å–4 Å) (**Table S3**). The majority of cancer mutations resulted in moderate global changes, but considerable local structural changes near the mutational site and in the activation loop. The results revealed that structurally conserved FLT3-D835V (**Figure 7**) and KIT-D816V mutations (**Figure 8**) enhanced the local protein mobility near the mutational site and destabilized the autoinhibited kinase conformation through a similar molecular mechanism. Interestingly, FLT3-D835 and KIT-D816 participate in stabilization of the  $3_{10}$ -helix (**Figures 7A, 8A**), which includes a stretch of residues (I836, M837, S838, D839, N841 in FLT3 and I817, K818, N819, D820 and S821 in KIT). During simulations the  $3_{10}$ -helix rapidly unfolded and remained in the unfolded state for both FLT3-D835V (**Figure 7B**) and KIT-D816 mutants (**Figure 8B**). Local perturbations induced by these mutations caused similar disruptions in the interaction networks responsible for stabilization of the inactive kinase form. In agreement with earlier studies [115–117], our results confirmed that deleterious effects of FLT3-D835V and KIT-D816V substitutions could primarily result from destabilization of the  $3_{10}$ -helix motif that is critical for the integrity of the inactive kinase form. Homology modeling and MD refinement of the EGFR-L861Q mutant, initiated from the inactive, Src-like EGFR crystal structure (**Figure 9A**), reproduced conformational changes in the activation loop leading to the active kinase form (**Figure 9B**). This may be attributed to a considerable incompat-



**Figure 7. Structural Modeling of the FLT3-D835V Mutant.** (A) The crystal structure of the autoinhibited wild-type FLT3 (pdb entry 1RJB). The position of D835 and key conserved residues K644 and E661 are highlighted. The location of the critical  $3_{10}$ -helix is indicated with an arrow. (B) Structural model of FLT3-D835V cancer mutant. Structural change in FLT3-D835V position and unwinding of the  $3_{10}$ -helix are highlighted with arrows.

doi:10.1371/journal.pone.0007485.g007



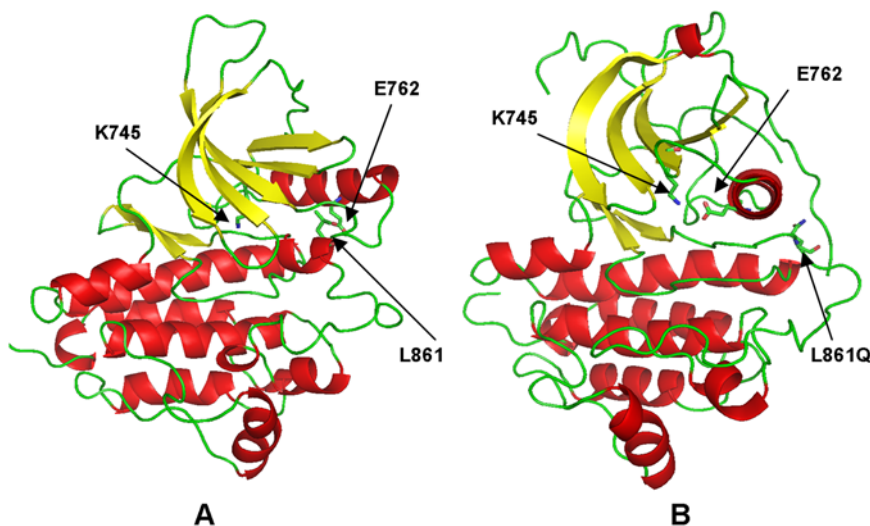
**Figure 8. Structural Modeling of the KIT-D816V Mutant.** (A) The crystal structure of the autoinhibited wild-type KIT (pdb entry 1T46). The position of D816 and key conserved residues K623 and E640 are highlighted. The location of the critical  $3_{10}$ -helix is indicated with an arrow. (B) Structural model of KIT-D816V cancer mutant. Structural change in KIT-D816V position and unwinding of the  $3_{10}$ -helix are highlighted with arrows. doi:10.1371/journal.pone.0007485.g008

ibility of the activating mutation with the Src-like structure of the WT EGFR. While the hydrophobic Leu-861 is packed in a hydrophobic core of the WT structure (**Figure 9A**), switching to a polar residue triggered a conformational transition of the activation loop folding outwards, towards an active-like kinase state (**Figure 9B**).

According to our recent findings [56,57], cancer mutations in ABL and EGFR kinases, that display high oncogenic activity, may also induce the greater differential effect on thermodynamic stability of the inactive and active kinase forms. These energetic factors may serve as thermodynamic catalysts of kinase activation by cancer mutations. In line with this hypothesis, structural signatures of the cancer mutational hotspot may manifest in deleterious protein stability changes in the inactive state of the enzyme, thereby promoting transitions to the constitutively active

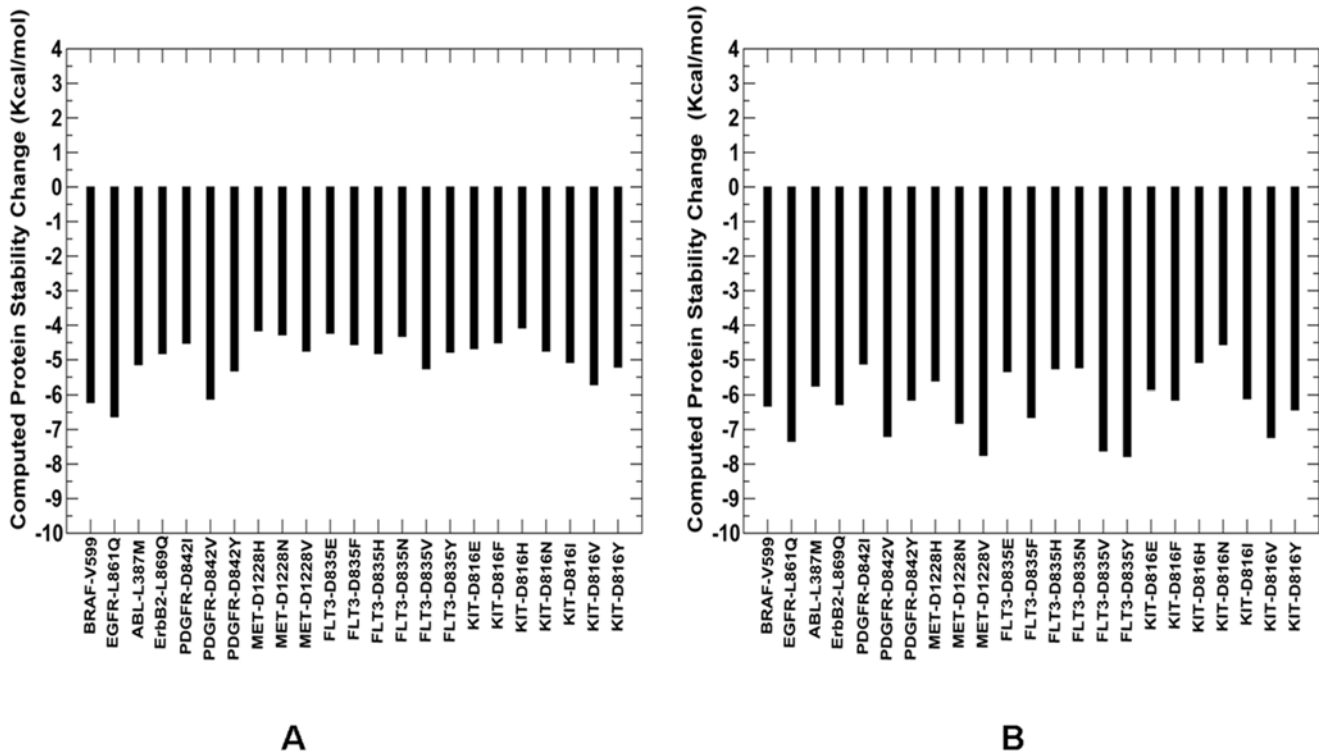
kinase form. In the present study, we verified and expanded the initial conjecture by analyzing structural mapping of mutational hotspots and performing computational evaluation of protein stability changes using CUPSAT and FOLDx methods (**Figures 10,11**). Both approaches revealed a consistent trend, whereby commonly occurring activating mutations with an appreciable oncogenic activity resulted in a considerable destabilization of the autoinhibited WT structure (**Figure 10**). For example, mutations D1228H, D1228N, and D1228V in MET from the mutational hotspot are known to have significant oncogenic transformation effect of NIH 3T3 cells [118,119]. Accordingly, these mutations were shown to have a significant destabilization effect on the protein structure (**Figure 10**).

In order to illustrate functional significance of structural effects and concomitant protein stability changes for kinase cancer

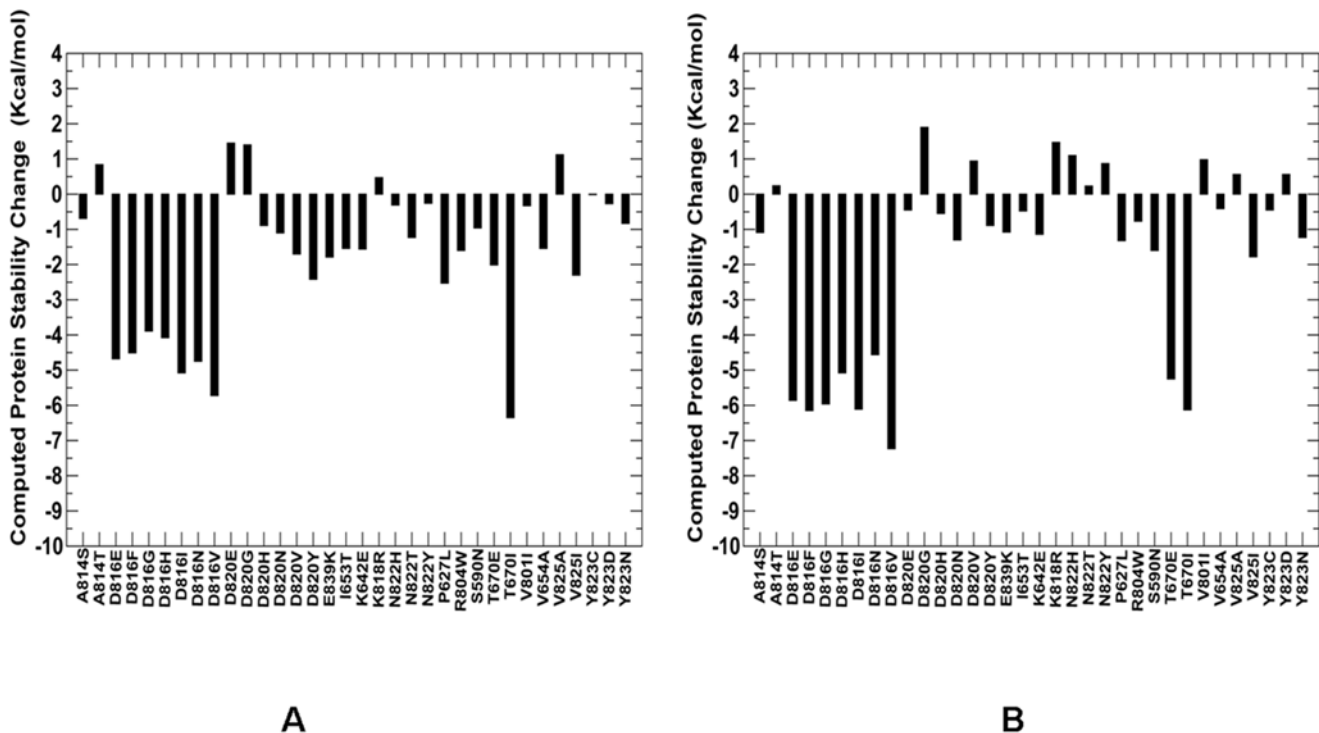


**Figure 9. Structural Modeling of the EGFR-L861Q Mutant.** (A) The inactive, Src-like structure of EGFR (pdb entry 2G7). The position of L861 is indicated with an arrow. The conserved salt bridge between K645 and E762 is broken in the inactive structure. (B) The model of the EGFR-L861Q mutant displays the active-like conformation of the activation loop. The new position of EGFR-L861Q residue and the restored salt bridge between K745 and E762 are indicated with arrows. doi:10.1371/journal.pone.0007485.g009





**Figure 10. Protein Stability Analysis of the Cancer Mutation Hotspot.** Protein stability differences calculated between the WT and mutants for structurally conserved mutations using CUPSAT (A) and FOLDx approaches (B). Negative values of protein stability changes correspond to destabilizing mutations.  
doi:10.1371/journal.pone.0007485.g010



**Figure 11. Protein Stability Analysis of KIT Mutations.** Protein stability differences between the WT and mutants for a panel of KIT mutations using CUPSAT (A) and FOLDx approaches (B). The panel included both disease-causing mutations and commonly occurring cancer mutations at D816 position. Negative values of protein stability changes correspond to destabilizing mutations.  
doi:10.1371/journal.pone.0007485.g011

mutations, we compared protein stability differences between oncogenic KIT mutations at the D816 position and a spectrum of disease-causing KIT variants (**Figure 11**). A considerable destabilization effect on the autoinhibited inactive kinase was observed for the activating KIT mutations. In contrast, disease-causing SNPs only marginally affected protein stability of the WT KIT structure. Despite simplified energy models employed in the CUPSAT and FOLDx approaches, we observed consistent trends, capturing highly oncogenic mutations as the mutations which elicit larger and more detrimental protein stability changes. These results are consistent with our earlier studies; supporting the hypothesis that functional role of cancer mutations may be associated with their impact on the protein kinase stability.

## Discussion

Development of the integrated bioinformatics resource CKMD has enabled structure-based functional annotation and prediction of cancer mutation effects in protein kinases. Structural mapping of kinase genetic variants onto aligned crystal structures and mutational models has allowed to characterize molecular effects of nsSNPs. We have found an enrichment of different categories of SNPs in the different structural regions of the kinase domain, suggesting structure-based determinants responsible for selection of tumorigenic mutational hotspots. The distributions of nsSNPs types has shown that (a) neutral kinase nsSNPs are randomly distributed within the catalytic core; (b) disease-causing nsSNPs map to regulatory and substrate binding regions; and (c) cancer-causing nsSNPs can target catalytic and nucleotide binding functions, preferentially clustering in the activation loop of the kinase domain. Based on these results, we could speculate about potential diversity of structural mechanisms that may be associated with the effects of genetic alterations. It is possible that disease-causing mutations may function by perturbing the local environment near the organizing F-helix, which is responsible for maintaining structural plasticity and correct positioning of the key catalytic and regulatory spine regions [85–87]. On the other hand, structural effects of cancer-causing mutations may manifest in perturbing flexible regions that are directly involved in conformational transitions between inactive and active kinase forms. The preferential localization of cancer-causing mutations in the P-loop and the activation loop may lower the energetic barrier for triggering the dynamic imbalance shifted towards the constitutively active kinase conformation. The earlier analysis of protein kinase motions indicated that conformational motions in functionally important protein regions which harbor cancer mutations, namely the P-loop and activation loop, are coupled and may be highly correlated [56,57].

Although kinase cancer mutations may not exhibit a strong sequence conservation signal, we have identified a number of structurally equivalent positions within the protein kinase catalytic core can be frequent targets of tumorigenic mutations. These structurally conserved mutations tend to cluster into specific mutational hotspots which may be shared by multiple kinase genes. Sequence and structure-based methods were used to characterize molecular determinants of mutational hotspots in protein kinases. We have determined that structurally conserved hotspots in the kinase catalytic domain can be often enriched by cancer driver mutations with a high oncogenic potential. Structural modeling and energetic analysis of the mutational hotspots have also suggested a common molecular mechanism of kinase activation by cancer mutations, which may be determined by a combined effect of the partial destabilization of the inactive state and a concomitant stabilization of the active-like form of the

enzyme. Furthermore, the results have indicated that cancer mutations with the higher oncogenic potential can have a greater differential effect on thermodynamic stability of the inactive and active kinase forms. Structure-based computational prediction and analysis of cancer mutation effects may thus be helpful for integrative cancer biology studies exploring the molecular pathology of tumorigenesis.

Ongoing development of database-oriented research tools within the CKMD environment will allow for automated structural and network-based bioinformatics analyses of rapidly growing knowledge-base of resequencing data on protein kinase genes. Further integration of genetic, functional, and structural insights about the molecular basis of tumorigenesis into robust bioinformatics infrastructure can ultimately help to discover molecular signatures of cancer mutations.

## Materials and Methods

### The Database Content and Organization

CKMD was developed as a bioinformatics resource for structure-functional analysis of genetic variations in protein kinases. We employed MySQL as a relational database management system for storing and managing the information content. Perl, a widely used scripting language was used to parse the data into various table forms. PHP5 Hypertext preprocessor was used in the design of the database interface, while Apache was used as the web server. Data stored in CKMD were mainly gathered from NCBI [74–76], COSMIC [82], SwissProt [120–122], and Protein Data Bank (PDB) [123]. We have also integrated non-redundant information about genetic variations in protein kinases from more specialized resources PupaSNP [73], KinMutBase [77,78], BTKbase [79], HGMD [80,81], PKR [83], and MoKCa [84].

Main entries in CKMD were indexed as genes and each gene entry contained many sub-entries of related information associated with that gene. We opted the gene id (GeneID) from Entrez Gene database as the unique identifier to index all entries in CKMD. This was partly due to the fact that the COSMIC database also referenced to GeneID in its entries. SwissProt, however, did not reference to GeneID and thus we developed a relation that matched SwissProt accession numbers with GeneIDs. This relation was crucial to coherently incorporate SwissProt data into CKMD along with the data from other sources. The raw data gathered from NCBI, SwissProt, and COSMIC were text files. All MySQL tables in CKMD referenced to either GeneID or SwissProt accession number. For each SNP entry, information about its position, nucleotide change and corresponding amino acid change was uniquely mapped on the protein kinase sequence and structure. The main information sources and a general architectural framework of CKMD are summarized in the design diagrams (**Figure S1**).

CKMD provides a simple and intuitive user interface that allows users to browse, search, download, and analyze genetic, sequence, structure and functional data on protein kinase data within a single integrated source. There are five main options available in CKMD: Composite, Browse, Search, Download, and Statistics. The “Composite” option offers a convenient and transparent way to view all information stored in CKMD for kinases genes. The “Browse” option allows to browse through entries in CKMD in three major categories: Gene, Mutation, and Structure. The “Search” option permits to query CKMD for a particular entry using many different searching criteria. The “Download” option allows to download and view all available protein kinase crystal structures and a large number of mutational

models. Finally, the “Statistics” option offers various sequence and structure-based statistical analyses of SNPs distributions across kinase genes. The important CKMD functionality is that the database stores and provides a convenient access to protein kinase crystal structures and mutational models with the mapped nsSNPs. A total of 989 crystal structures corresponding to 126 kinase genes were collected from PDB and consolidated in CKMD. 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 [124]. We have developed Java applet using Jmol, an open-source Java viewer for chemical structures in 3D (<http://www.jmol.org/>), to provide graphical representation of protein kinase structures. This interface could allow users to load and view multiple and aligned protein kinase structures along with convenient tools for manipulation of three-dimensional structures, localization and molecular analysis of SNPs.

Protein kinase sequences were obtained from Kinbase (<http://kinase.com/kinbase/>). Common SNPs were retrieved from PupaSNP [73] and dbSNP [74] using the Ensembl data mining tool, Biomart ([http://www.ensembl.org/Homo\\_sapiens/martview](http://www.ensembl.org/Homo_sapiens/martview)). The disease causing SNPs were retrieved from OMIM [75,76], KinMutBase [77,78], and HGMD resources [80,81]. Currently, there are 518 kinase gene entries in CKMD, both referenced in NCBI [74–76] and SwissProt database [120–122], and 7955 unique SNP entries corresponding to these kinase genes that are referenced in NCBI. These unique SNP entries include 3722 synonymous, 3985 missense, 75 nonsense and 173 frameshift mutations. We have also gathered 780 OMIM variant entries from NCBI and 3542 SwissProt variant entries. Cancer mutations were retrieved from OMIM [75,76] and COSMIC resources [82]. The complete lists of mRNA and protein products for each unique SNP entry were also included and cross-linked to NCBI database. All nsSNPs were assigned to positions in Kinbase protein sequence using flanking sequences in the Ensembl and Entrez Gene sequences because of higher confidence in Kinbase sequences versus other publicly available sequences. Corresponding positions in DNA sequences were determined using a combination of flanking sequences given in dbSNP data and Genewise (<http://www.ebi.ac.uk/Wise2/>).

### Motif-based and Structure-based Multiple Sequence Alignments

Motif-based alignments of kinase sequences to the catalytic core were first generated by implementation of the Gibbs motif sampling method [125,126]. This method identifies characteristic motifs for each individual subdomain of the kinase catalytic core, which are then used to generate high-confidence motif-based Markov chain Monte Carlo multiple alignments based on these motifs [127,128]. These subdomains define the core structural components of the protein kinase catalytic core. Intervening regions between these subdomains were not aligned. The nsSNPs were then mapped to the kinase catalytic domain in accordance with this alignment. Cancer driver predictions were performed by using the SVM approach as described in our earlier work [70,71]. Sequence analysis was done with the aid of the subPSEC conservation measure [88,89].

To further verify structural distribution of nsSNPs in functional kinase regions, we also performed structure-informed multiple alignment of kinase sequences using PROMALS3D approach [129]. In this approach, 30 different kinase crystal structures (Table S4) (the maximum allowed limit of structural information used by PROMALS3D) and kinase catalytic domain sequences for

445 different genes were used for the multiple sequence alignment. The obtained alignment was then matched against the alignment of the kinase sequences with the available crystal structure to ascertain the quality of the sequence alignment. The predicted and observed residue ranges for the catalytic loop, hinge region,  $\alpha$ C-helix, activation loop and P-loop are in excellent agreement with the observed residue ranges for these functional kinase regions (Table S5).

### Kinase SNP Distribution and Enrichment Analysis

Functionally important subdomains of the kinase catalytic core, as in the nomenclature defined by Hanks and Hunter [7], were examined to determine the distribution of nsSNPs and identify structurally conserved hotspots of functionally important mutations. The number of SNPs in each of the subdomains was calculated from the structure-informed multiple sequence alignment described in the previous section. The expected probability  $E(p)$  of a SNP occurring in a kinase subdomain region was calculated separately for each SNP type as previously documented [71,72]. In brief, the average length of each region was calculated as the weighted average of the region length in each kinase considered, where weights correspond to the total number of SNPs occurring within each kinase. This weighting helps avoid biases that might arise as a result of some kinases simply harboring more SNPs than others. The probability of a SNP occurring within a particular region purely by chance was computed as its weighted average length over the sum of every region’s weighted average length. The probability (p-value) of the observed total number ( $x$ ) of SNPs occurring within each region, where  $n$  is the total number of SNPs considered, was calculated using the general binomial distribution as follows:

If  $x/n < E(p)$ :

$$p\text{-value}(x) = \left( \sum_{x=0}^x \binom{n}{x} \bullet E(p)^x (1 - E(p))^{n-x} \right) \bullet 2$$

If  $x/n > E(p)$ :

$$p\text{-value}(x) = \left( \sum_{x=x}^n \binom{n}{x} \bullet E(p)^x (1 - E(p))^{n-x} \right) \bullet 2$$

### Structural Modeling and MD Refinement of Kinase Cancer Mutants

We have also consolidated in CKMD all publicly available crystal structures of WT and mutant protein kinases from PDB. A total of 989 kinase crystal structures corresponding to 126 genes were deposited in CKMD. Although a number of kinase crystal structures including mutants have been solved, there is still very little structural information about most cancer kinase mutants. To facilitate structure-functional analysis of cancer mutation effects in protein kinases we have generated and stored in CKMD structural models of a large number of protein kinase mutants (Figure S2). Only a subset of all SNPs can be directly mapped onto the kinase crystal structures. As a result, there are some protein kinases with the known WT crystal structure and known SNPs, yet no mutational models could be generated, because either all known mutations reside outside of the resolved crystal structure of the kinase catalytic domain or only synonymous mutations were available.

Structural modeling of nsSNPs was carried out using MODELLER [130,131] with a subsequent refinement of side-chains by the SCRWL3 program [132]. Initial models were built in MODELLER using a flexible sphere of 5 Å around mutated residue and the inactive crystal structures of the WT EGFR, FLT3, and KIT kinases as the templates. 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. In the optimization stage, we initially used 5000 steps of conjugate gradient (CG) minimization to remove unfavorable contacts and ensure sufficient relaxation of the local environment near mutational site. The predicted mutational models were chosen out of the 100 models as scored by the MODELLER default scoring function. These final models were then refined in 2ns MD simulations using NAMD 2.6 [133] with the CHARMM27 force field [134,135] and the explicit TIP3P water model as implemented in NAMD 2.6 [136]. Equilibration was done in stages by gradually increasing the system temperature in steps of 20K starting from 10K until 310K. At each stage, 10,000 equilibration steps was employed, while applying a harmonic restraining force of  $10 \text{ Kcalmol}^{-1} \text{ \AA}^{-2}$  to all backbone C $\alpha$  atoms. Subsequently, the system was equilibrated for 150,000 steps at 310K (NVT) and then for additional 150,000 steps at 310K using Langevin piston (NPT) to maintain the pressure. Finally the restrains were removed and the system was equilibrated for 500,000 steps to prepare the system for simulation. An NPT simulation was run on the equilibrated structure keeping the temperature at 310K and pressure at 1 bar using Langevin piston coupling algorithm. Nonbonded van der Waals interactions were treated by using a switching function at 10Å and reaching zero at 12 Å distance.

### Protein Stability Calculations

To quantify the destabilization effect of cancer mutations on the inactive, autoinhibited kinase form, we computed the protein stability change upon these mutations using CUPSAT (Cologne University Protein Stability Analysis Tool) approach for the prediction and analysis of protein stability changes upon point mutations [137,138] and Foldx method [139,140]. In the CUPSAT approach, coarse-grained atom potentials and torsion angle potentials are used to predict protein stability upon point mutations. Foldx analysis of protein stability is based on the empirical force field which was developed for the rapid evaluation of the effect of mutations on the stability, folding and dynamics of

proteins and nucleic acids. The free energy of folding is evaluated in this approach from the difference in Gibbs free energy between the crystal structure of the protein and a hypothetical unfolded reference state of which no structural details are known.

### Supporting Information

**Figure S1** CKMD Architecture and Information Sources.

Found at: doi:10.1371/journal.pone.0007485.s001 (0.20 MB TIF)

**Figure S2** The Gene-based Distribution of Structurally Mapped Kinase Cancer Mutations. For clarity of presentation, only top 70 kinase genes that have cancer-causing nsSNPs mapped onto three-dimensional protein structure are presented.

Found at: doi:10.1371/journal.pone.0007485.s002 (0.75 MB TIF)

**Table S1** Structurally conserved cancer mutation hotspots in the protein kinase genes.

Found at: doi:10.1371/journal.pone.0007485.s003 (0.03 MB XLS)

**Table S2** Distribution of kinase cancer mutations in the mutational samples

Found at: doi:10.1371/journal.pone.0007485.s004 (0.18 MB DOC)

**Table S3** Analysis of the crystal structures and mutational models

Found at: doi:10.1371/journal.pone.0007485.s005 (0.09 MB DOC)

**Table S4** The list of protein kinase crystal structures used for the multiple sequence alignment by PROMALS3D.

Found at: doi:10.1371/journal.pone.0007485.s006 (0.06 MB DOC)

**Table S5** Structure-based multiple alignment of kinase sequences.

Found at: doi:10.1371/journal.pone.0007485.s007 (0.04 MB XLS)

### Author Contributions

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

### References

- Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100: 57–70.
- International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437: 1299–1320.
- International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851–861.
- Eichler EE, Nickerson DA, Altshuler D, Bowcock AM, Brooks LD, et al. (2007) Completing the map of human genetic variation. *Nature* 447: 161–165.
- Lander ES, Weinberg RA (2000) Genomics: journey to the center of biology. *Science* 287: 1777–1782.
- Collins FS, Barker AD (2007) Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am* 296: 50–57.
- Hanks SK, Hunter T (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 9: 576–596.
- Hunter T, Plowman GD (1997) Review: the protein kinases of budding yeast: six score and more. *Trends Biochem Sci* 22: 18–22.
- Hunter T (2000) Signaling – 2000 and beyond. *Cell* 100: 113–127.
- Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. *Science* 298: 1912–1934.
- Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 10: 514–520.
- Noble ME, Endicott JA, Johnson LN (2004) Protein kinase inhibitors: insights into drug design from structure. *Science* 303: 1800–1805.
- Cherry M, DH (2004) Recent kinase and kinase inhibitor X-ray structures: mechanisms of inhibition and selectivity insights. *Curr Med Chem* 11: 663–673.
- Cheatham GM (2004) Novel protein kinases and molecular mechanisms of autoinhibition. *Curr Opin Struct Biol* 14: 700–705.
- Scheeff ED, Bourne PE (2005) Structural evolution of the protein kinase-like superfamily. *PLoS Comput Biol* 2005 1: e49.
- Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. *Cell* 109: 275–282.
- Nolen B, Taylor SS, Ghosh G (2004) Regulation of protein kinases. Controlling activity through activation segment conformation. *Molecular Cell* 15: 661–675.
- Sridhar R, Hanson-Painton O, Cooper DR (2000) Protein kinases as therapeutic targets. *Pharm Res* 17: 1345–1353.
- Madhusudan S, Ganesan TS (2004) Tyrosine kinase inhibitors in cancer therapy. *Clin Biochem* 37: 618–635.
- Sawyers CL (2003) Opportunities and challenges in the development of kinase inhibitor therapy for cancer. *Genes and Dev* 17: 2998–3010.
- Sawyer TK (2004) Novel oncogenic protein kinase inhibitors for cancer therapy. *Curr Med Chem Anticancer Agents* 4: 449–455.
- Knight ZA, Shokat KM (2005) Features of selective kinase inhibitors. *Chem Biol* 12: 621–637.
- Davies H, Hunter C, Smith R, Stephens P, Greenman C, Bignell G, et al. Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res.* 2005 65(17): 7591–5.

24. Stephens P, Edkins S, Davies H, Greenman C, Cox C, et al. (2005) A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat Genet* 37: 950–952.
25. Sjöblom T, Jones S, Wood LD, Parsons WD, Lin J (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314: 268–274.
26. Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, et al. (2007) High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 39: 347–351.
27. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, et al. (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318: 1108–1113.
28. Greenman C, Stephens P, Smith R, Dalgleish GL, Hunter C, et al. (2007) Patterns of somatic mutation in human cancer genomes. *Nature* 446: 153–158.
29. Jones S, Zhang X, Parsons DW, Lin JCH, Leary RJ, et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321: 1801–1806.
30. Parsons DW, Jones S, Zhang X, Lin JCH, Leary RJ, et al. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321: 1807–1812.
31. Kiemeny LA, Thorlacius S, Sulem P, Geller F, Aben KKH, et al. (2008) Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet* 40: 1307–1312.
32. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. (2009) Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 41: 324–328.
33. Weinstein BI (2002) Cancer. Addiction to oncogenes—the Achilles heel of cancer. *Science* 297: 63–64.
34. Sharma SV, Settleman J (2007) Oncogene addiction: setting the stage for molecularly targeted cancer therapy. *Genes Dev* 21: 3214–3231.
35. Lynch TJ, Bell DW, Sordella R, Gurubhagavata S, Okimoto RA, et al. (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 350: 2129–2139.
36. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, et al. (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304: 1497–1500.
37. Pao W, Miller V, Zakowski M, Doherty J, Politi K, et al. (2004) EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci USA* 101: 13306–13311.
38. 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.
39. 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.
40. 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.
41. Kumar A, Petri ET, Halmos B, Boggon TJ (2008) Structure and clinical relevance of the epidermal growth factor receptor in human cancer. *J Clin Oncol* 26: 1742–1751.
42. 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.
43. 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.
44. 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.
45. Arora K, Brooks CL 3rd (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.
46. Ozkirimli E, Post CB (2006) Src kinase activation: A switched electrostatic network. *Protein Sci* 15: 1051–1062.
47. Ozkirimli E, Yadav SS, Miller WT, Post CB (2008) An electrostatic network and long-range regulation of Src kinases. *Protein Sci* 17: 1871–1880.
48. Banavali NK, Roux B (2007) Anatomy of a structural pathway for activation of the catalytic domain of Src kinase Hck. *Proteins* 67: 1096–1112.
49. Yang S, Roux B (2008) Src kinase conformational activation: Thermodynamics, pathways, and mechanisms. *PLoS Comput Biol* 4: e1000047.
50. Banavali NK, Roux B (2009) Flexibility and charge asymmetry in the activation loop of Src tyrosine kinases. *Proteins* 74: 378–389.
51. 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.
52. Bertecchi 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.
53. Shan Y, Seeliger MA, Eastwood MP, Frank F, Xu H, Jensen MO, Dror RO, Kuriyan J, Shaw DE (2009) A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc Natl Acad Sci U S A* 106: 139–144.
54. Zou J, Wang YD, Ma FX, Xiang ML, Shi B, Wei YQ, Yang SY (2008) Detailed conformational dynamics of juxtamembrane region and activation loop in c-Kit kinase activation process. *Proteins* 72: 323–332.
55. Papakyriakou A, Vourloumis D, Tzortzou-Stathopoulou F, Karpusas M (2008) Conformational dynamics of the EGFR kinase domain reveals structural features involved in activation. *Proteins* 76: 375–386.
56. 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.
57. 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.
58. Pellicena P, Kuriyan J (2006) Protein-protein interactions in the allosteric regulation of protein kinases. *Curr Opin Struct Biol* 16: 702–709.
59. Masterson LR, Mascioni A, Traaseth NJ, Taylor SS, Veglia G (2008) Allosteric cooperativity in protein kinase A. *Proc Natl Acad Sci U S A* 105: 506–511.
60. Loriaux MM, Levine RL, Tyner JW, Fröhling S, Scholl C, et al. (2008) High-throughput sequence analysis of the tyrosine kinome in acute myeloid leukemia. *Blood* 111: 4788–4796.
61. Tyner JW, Loriaux MM, Erickson H, Eide CA, Deininger J, et al. (2008) High-throughput mutational screen of the tyrosine kinome in chronic myelomonocytic leukemia. *Leukemia* 23: 406–409.
62. Tomasson MH, Xiang Z, Walgren R, Zhao Y, Kasai Y, et al. (2008) Somatic mutations and germline sequence variants in the expressed tyrosine kinase genes of patients with de novo acute myeloid leukemia. *Blood* 111: 4797–4808.
63. Fröhling S, Scholl C, Levine RL, Loriaux M, Boggon TJ, et al. (2007) Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell* 12: 501–513.
64. Parmigiani G, Boca S, Lin J, Kinzler KW, Velculescu V, et al. (2009) Design and analysis issues in genome-wide somatic mutation studies of cancer. *Genomics* 93: 17–21.
65. Torkamani A, Verkhivker G, Schork NJ (2009) Cancer driver mutations in protein kinase genes. *Cancer Lett* 281: 117–127.
66. Krallinger M, Izarzugaza JMG, Rodriguez-Penagos C, Valencia A (2009) Extraction of human kinase mutations from literature, databases and genotyping studies *BMC Bioinformatics* 10(Suppl 8): S1.
67. Izarzugaza JM, Redfern OC, Orengo CA, Valencia A (2009) Cancer associated mutations are preferentially distributed in protein kinase functional sites. *Proteins*. In press.
68. Kaminker JS, Zhang Y, Waugh A, Haverty PM, Peters B, et al. (2007) Distinguishing cancer-associated missense mutations from common polymorphisms. *Cancer Res* 67: 465–473.
69. Kaminker JS, Zhang Y, Watanabe C, Zhang Z (2007) CanPredict: a computational tool for predicting cancer-associated missense mutations. *Nucleic Acids Res (Web Server issue)* 35: W595–598.
70. Torkamani A, Schork NJ (2007) Accurate prediction of deleterious protein kinase polymorphisms. *Bioinformatics* 23: 2918–2925.
71. Torkamani A, Schork NJ (2008) Prediction of cancer driver mutations in protein kinases. *Cancer Res* 68: 1675–1682.
72. 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.
73. 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.
74. 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.
75. 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.
76. Reibholz-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.
77. Stenberg KA, Riikonen PT, Vihinen M (2000) KinMutBase, a database of human disease-causing protein kinase mutations. *Nucleic Acids Res* 28: 369–371.
78. Ortutay C, Valiaho J, Stenberg K, Vihinen M (2005) KinMutBase: a registry of disease-causing mutations in protein kinase domains. *Hum Mutat* 25: 435–442.
79. Valiaho J, Smith CI, Vihinen M (2006) BTKbase: the mutation database for X-linked agammaglobulinemia. *Hum Mutat* 27: 1209–1217.
80. 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.
81. 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.
82. 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.

83. 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.
84. Richardson CJ, Gao Q, Mitsopoulos C, Zvelebil M, Pearl LH, et al. (2009) MoKCa database—mutations of kinases in cancer. *Nucleic Acids Res* 37(Database issue): D824–31.
85. Kannan N, Neuwald AF (2005) Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *J Mol Biol* 351: 956–972.
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. Thomas PD, Kejarawal A (2004) Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci U S A* 101: 15398–15403.
89. Thomas PD, Campbell MJ, Kejarawal A, Mi H, Karlak B, et al. (2003) PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res* 13: 2129–2141.
90. Chiara F, Michieli P, Pugliese L, Comoglio PM (2003) Mutations in the met oncogene unveil a “dual switch” mechanism controlling tyrosine kinase activity. *J Biol Chem* 278: 29352–29358.
91. Lorenzato A, Olivero M, Patane S, Rosso E, Oliaro A, et al. (2002) Novel somatic mutations of the Met oncogene in human carcinoma Metastases activating cell motility and invasion. *Cancer Res* 62: 7025–7030.
92. Yamamoto Y, Kiyoi H, Nakano Y, Suzuki R, Kodera Y, et al. (2001) Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. *Blood* 97: 2434–2439.
93. Abu-Duhier FM, Goodeve AC, Wilson GA, Care RS, Peake IR, et al. (2001) Identification of novel FLT-3 Asp835 mutations in adult acute myeloid leukaemia. *Br J Haematol* 113: 983–988.
94. Ferrao PT, Gonda TJ, Ashman LK (2003) Constitutively active mutant D816VKit induces megakaryocyte and mast cell differentiation of early haemopoietic cells from murine foetal liver. *Leuk Res* 27: 547–555.
95. Tan A, Westerman D, McArthur GA, Lynch K, Waring P, et al. (2006) Sensitive detection of KIT D816V in patients with mastocytosis. *Clin Chem* 52: 2250–2257.
96. Sensi M, Nicolini G, Petti C, Bersani I, Lozupone F, et al. (2006) Mutually exclusive NRASQ61R and BRAFV600E mutations at the single-cell level in the same human melanoma. *Oncogene* 25: 3357–3364.
97. Furitsu T, Tsujimura T, Tono T, Ikeda H, Kitayama H, et al. (1993) Identification of mutations in the coding sequence of the proto-oncogene c-kit in a human mast cell leukemia cell line causing ligand-independent activation of c-kit product. *J Clin Invest* 92: 1736–1744.
98. 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.
99. Fu YN, Yeh CL, Cheng HH, Yang CH, Tsai SF, et al. (2008) EGFR mutants found in non-small cell lung cancer show different levels of sensitivity to suppression of Src: implications in targeting therapy. *Oncogene* 27: 957–965.
100. Corbin AS, La Rosée P, Stoffregen EP, Druker BJ, Deininger MW (2003) Several Bcr-Abl kinase domain mutants associated with imatinib mesylate resistance remain sensitive to imatinib. *Blood* 101: 4611–4614.
101. Maritano D, Accornero P, Bonifazi N, Ponzetto C (2000) Two mutations affecting conserved residues in the Met receptor operate via different mechanisms. *Oncogene* 19: 1354–1361.
102. Gujral TS, Singh VK, Jia Z, Mulligan LM (2006) Molecular mechanisms of RET receptor-mediated oncogenesis in multiple endocrine neoplasia 2B. *Cancer Res* 66: 10741–10749.
103. Gujral TS, Mulligan LM (2006) Molecular implications of RET mutations for pheochromocytoma risk in multiple endocrine neoplasia 2. *Ann N Y Acad Sci* 1073: 234–240.
104. Lai AZ, Gujral TS, Mulligan LM (2007) RET signaling in endocrine tumors: delving deeper into molecular mechanisms. *Endocr Pathol* 18: 57–67.
105. Cranston AN, Carmiti C, Oakhill K, Andzelm ER, Stone EA, et al. (2006) RET is constitutively activated by novel tandem mutations that alter the active site resulting in multiple endocrine neoplasia type 2B. *Cancer Res* 66: 10179–10187.
106. Knowles PP, Rust JM, Kjaer S, Scott RP, Hanrahan S, et al. (2006) Structure and chemical inhibition of the RET tyrosine kinase domain. *J Biol Chem* 281: 33577–33587.
107. Berthou S, Aebersold DM, Schmidt LS, Stroka D, Heigl C, et al. (2004) The Met kinase inhibitor SU11274 exhibits a selective inhibition pattern toward different receptor mutant variants. *Oncogene* 23: 5387–5393.
108. Morotti A, Mila S, Accornero P, Tagliabue E, Ponzetto C (2002) K252a inhibits the oncogenic properties of Met, the HGF receptor. *Oncogene* 21: 4885–4893.
109. Nakaigawa N, Weirich G, Schmidt L, Zbar B (2000) Tumorigenesis mediated by MET mutant M1268T is inhibited by dominant-negative Src. *Oncogene* 19: 2996–3002.
110. Miller M, Ginalski K, Lesyng B, Nakaigawa N, Schmidt L, et al. (2001) Structural basis of oncogenic activation caused by point mutations in the kinase domain of the MET proto-oncogene: modeling studies. *Proteins* 44: 32–43.
111. Griffith J, Black J, Faerman C, Swenson L, Wynn M, et al. (2004) The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. *Mol Cell* 13: 169–178.
112. 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.
113. Schiering N, Knapp S, Marconi M, Flocco MM, Cui J, et al. (2003) Crystal structure of the tyrosine kinase domain of the hepatocyte growth factor receptor c-Met and its complex with the microbial alkaloid K-252a. *Proc Natl Acad Sci U S A* 100: 12654–12659.
114. Wang W, Marimuthu A, Tsai J, Kumar A, Krupka HI, et al. (2006) Structural characterization of autoinhibited c-Met kinase produced by coexpression in bacteria with phosphatase. *Proc Natl Acad Sci U S A* 103: 3563–3568.
115. Foster R, Griffith R, Ferrao P, Ashman L (2004) Molecular basis of the constitutive activity and STI571 resistance of Asp816Val mutant KIT receptor tyrosine kinase. *J Mol Graph Model* 23: 139–152.
116. Torrent M, Rickert K, Pan BS, Sepp-Lorenzino L (2004) Analysis of the activating mutations within the activation loop of leukemia targets Flt-3 and c-Kit based on protein homology modeling. *J Mol Graph Model* 2004 23: 153–165.
117. Vendôme J, Letard S, Martin F, Svinarchuk F, Dubreuil P, et al. (2005) Molecular modeling of wild-type and D816V c-Kit inhibition based on ATP-competitive binding of ellipticine derivatives to tyrosine kinases. *J Med Chem* 48: 6194–201.
118. Jeffers M, Schmidt L, Nakaigawa N, Webb CP, Weirich G, et al. (1997) Activating mutations for the met tyrosine kinase receptor in human cancer. *Proc Natl Acad Sci USA* 94: 11445–11450.
119. Bardelli A, Longati P, Gramaglia D, Basilico C, Tamagnone L, et al. (1998) Uncoupling signal transducers from oncogenic MET mutants abrogates cell transformation and inhibits invasive growth. *Proc Natl Acad Sci USA* 95: 14379–14383.
120. 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.
121. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A (2007) UniProtKB/Swiss-Prot. *Methods Mol Biol* 406: 89–112.
122. The UniProt Consortium The universal protein resource (UniProt) (2008) *Nucleic Acids Res* 36: D190–D195.
123. 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(Database issue): D302–D305.
124. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33: 2302–2309.
125. Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, et al. (1993) Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* 262: 208–214.
126. Neuwald AF, Liu JS, Lawrence CE (1995) Gibbs motif sampling: detection of bacterial outer membrane protein repeats. *Protein Sci* 4: 1618–1632.
127. Neuwald AF, Liu JS (2004) Gapped alignment of protein sequence motifs through Monte Carlo optimization of a hidden Markov model. *BMC Bioinformatics* 5: 157.
128. Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G (2007) Structural and functional diversity of the microbial kinome. *PLoS Biol* e17.
129. Pei J, Tang M, Grishin NV (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res* 36(Web Server issue): W30–34.
130. 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.
131. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Science* 9: 1753–1773.
132. Canutescu AA, Shelenkov AA, Dunbrack RL Jr (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* 12: 2001–2014.
133. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802.
134. 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.
135. MacKerell AD Jr, Banavali N, Foloppe N (2001) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 56: 257–265.
136. 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.
137. Parthiban V, Gromiha MM, Schomburg D (2006) CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res* 34: W239–242.
138. Parthiban V, Gromiha MM, Abhinandan M, Schomburg D (2007) Computational modeling of protein mutant stability: analysis and optimization of statistical potentials and structural features reveal insights into prediction model development. *BMC Struct Biol* 7: 54.
139. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005) The FoldX web server: an online force field. *Nucleic Acids Res* 33: W382–388.
140. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 320: 369–87.