JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

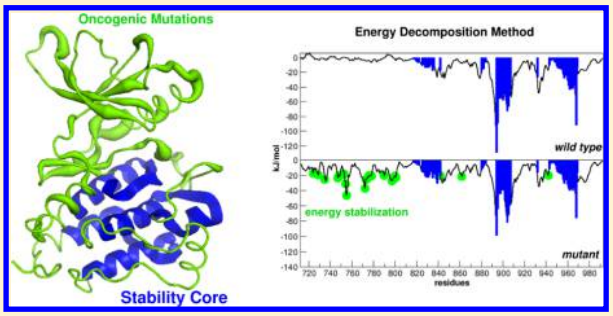
Structural Stability and Flexibility Direct the Selection of Activating Mutations in Epidermal Growth Factor Receptor Kinase

Antonella Paladino, Giulia Morra, and Giorgio Colombo*

Istituto di Chimica del Riconoscimento Molecolare, CNR Via Mario Bianco 9, 20131, Milano, Italy

Supporting Information

ABSTRACT: Herein we investigate the potential of novel methods of molecular dynamics analysis to provide information on the key factors that underlie the preferential localization and the effects of mutations modulating protein activities. Epidermal growth factor receptor (EGFR) kinases are selected as a test case. The combined analysis of protein energetics and internal dynamics indicates a clear polarization in the native protein, whereby a highly stable and ordered scaffold in one domain, namely the C-lobe, is combined to a flexible and loosely stabilized domain, the N-lobe. The subdivision in two portions with different properties directs the presence of point mutations mainly to the N-lobe. This allows modulating protein flexibility so that the protein can more efficiently sample the conformations necessary for substrate recognition, while leaving the stability of the protein unperturbed. In this context, comparative simulations of EGFR in the wild type sequence and in the presence of the activating oncogenic mutation G719S reveal flexibility changes in several key regions, involving in particular the part of the kinase devoted to the regulation of substrate recognition (regulatory core) and an increase in the number of stabilizing interactions in the N-lobe for the activated mutant. Our approaches represent a promising and simple strategy toward rationalizing the effects of mutations in modulating enzymatic activities.



INTRODUCTION

Protein kinases are signaling switches with a conserved catalytic domain that phosphorylate protein substrates and play a critical role in cell signaling pathways linked to cell maintenance, survival, and proliferation.[1] Protein kinase genes constitute 2% of all genes in the human genome, and this protein family consists of more than 500 different members.[2] The protein data bank reports 167 unique human protein kinase domains.

The kinase domain is also the most commonly found among known cancer genes. Since the discovery of the first oncogene, vSrc, in the 1970s and its identification as an enzyme with tyrosine phosphorylating activity, dysfunctional signaling by mutated or overexpressed kinases has been intimately linked to the development of cancer pathologies.[3−6]

Extensive efforts aimed at characterizing the reaction mechanisms have linked enzymatic activity to a "two-state" dynamic/conformational model, whereby proteins switch between an inhibited, structurally rigid ground state and a more dynamic and heterogeneous active state. Pathogenic mutations have different abilities to shift this balance toward the active state, arguably by lowering the energy costs linked to the demanding conformational transition to the active structure. The overactivation of the enzymatic activity confers cells a growth and survival advantage, so that it is not surprising that many kinase mutations have been linked to cancer phenotypes. All these properties have made kinases attractive drug targets for the development of antitumor therapies.[5−9]

In recent years, an increasing number of experimental and theoretical studies have focused on kinases. Different crystal structures in the presence or absence of ligands have revealed the major conformational states of these proteins.[5,6,10−12] Computational studies have addressed mechanistic aspects related to the recognition and binding of ligands, as well as to the possible effects of mutations on kinase conformational dynamics. In this context, multiscale and enhanced sampling approaches have been extensively used to describe conformational transitions, while the recent availability of unprecedented computer power and the optimization of existing molecular dynamics (MD) algorithms have started to access time scales on the order of micro- and milliseconds.[7,13] Nevertheless, the high computational costs still associated with these techniques make their routine application challenging.

The goal of this paper is to investigate the main conformational and physicochemical determinants of the modulation of kinase activity induced by mutations. To this end, we propose a novel computational scheme based on the analysis of internal protein energetics and flexibility/rigidity to describe the protein functional dynamics, starting from classical, accessible MD simulations.

In particular, we focus on EGFR as a paradigmatic example for the application of our approach. Members of the epidermal growth factor receptor (EGFR) family are transmembrane

Received: May 12, 2015
Published: June 29, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

tyrosine kinases that are activated by ligand-induced dimerization. The isolated wild-type (wt) EGFR domain is in an autoinhibited state, in which the activation loop is folded onto the kinase substrate-binding site. wt-EGFR's by themselves exist predominantly as monomers, indicative of a free-energy cost associated with their dimerization.[10,14] Mutations in EGFR are frequently found in many cancers, where they have been shown to impact the equilibrium between inactive and active states, favoring the latter.[7,9,15,16]

In this study, we have carried out and analyzed different MD simulations of EGFR (Figure 1) in the active and inactive forms, both in the wild type sequence and in the presence of a known activating mutation, namely G719S (see Table 1).

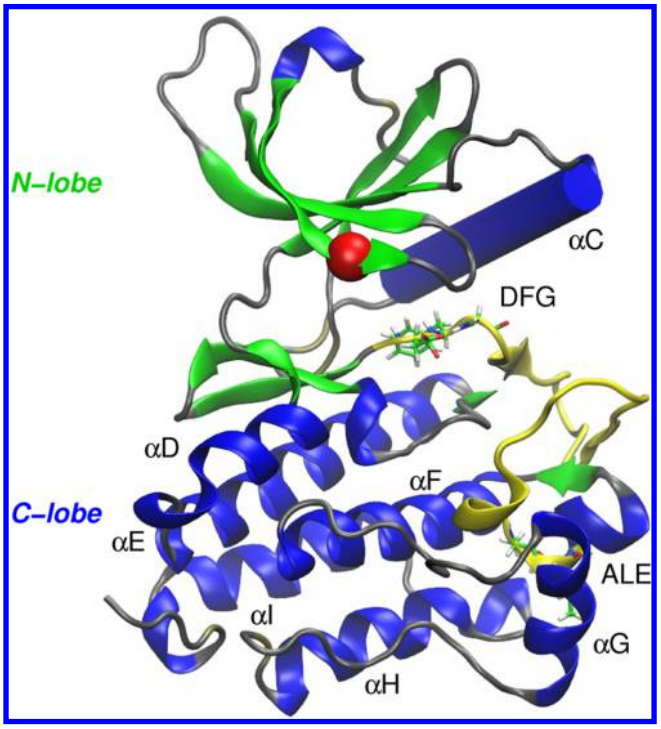


Figure 1. X-ray structure of active EGFR (pdb code 2ITP). Secondary structure elements are given in cartoon representation. The green N-lobe includes the five-stranded β sheet and the αC-helix. The C-lobe is formed by the blue bundle of α-helices. The activation (P+1) loop (yellow) extends between DFG and ALE motifs, shown in sticks. The G719S mutation is also shown as a red ball.

Table 1. List of Simulated EGFRs

name	group	family	PDB	state	αC-conformation
EGFR	Egfr	TK	2ITP	activated	in
EGFR[a]	Egfr	TK	2ITP[a]	activated	in
EGFR	Egfr	TK	2GS7	inactive	out
EGFR[a]	Egfr	TK	2GS7[a]	inactive	out

[a]Indicates point mutation G719S. Inactive and activated states refer to closed and open form of the kinase domain, respectively, due to the rearrangement of the αC-helix and activation loop, as described in the main text.

We first asked whether it is possible to correlate the frequency of occurrence and position of oncogenic mutations to specific energetic signatures, obtainable from the study of the WT protein. On this basis, we hypothesize a mechanistic model for kinase activation and test it on the G719S EGFR mutant. The knowledge gained here may be useful in the design of inhibitors or antibodies that selectively target specific regions in oncogenic mutants[17−19] generating novel opportunities for the development of therapeutics, possibly with minimal side effects. For these purposes, we have applied a combination of recently developed computational methods aimed at uncovering the intramolecular energetic networks and coordination mechanisms, and their mutation-induced modulations, defining internal protein properties that can be linked to functional states.

MATERIALS AND METHODS

MD simulations were carried out using the Gromacs software package[20] (v.4.5.5) with the Amber99 force field.[21] Selected starting structures for protein kinases are summarized in Table 1. All EGFR kinase domains are simulated in the apo form with neither ATP nor Mg^{2+}.

The proteins were centered in triclinic boxes at 0.9 nm distance from each box edge and solvated with TIP3P water molecules.[22] Counterions were randomly added to ensure overall charge neutrality. The system was first energy minimized using the steepest descent approach, followed by a 5 ns simulation in which the positions of the protein heavy atoms were restrained by a harmonic potential. Production trajectories were run for 100 ns at constant temperature of 300 K and a constant pressure of 1 atm. All simulations were run in two replicas. A cutoff radius of 0.9 nm for nonbonded van der Waals interactions was used in all simulations. Bond lengths involving hydrogens were restrained by the LINCS algorithm.[23] Electrostatic interactions were treated using the particle mesh Ewald method.[24] A time step was set to 2 fs and periodic boundary conditions were applied in all three dimensions.

Energy Decomposition Method. In the energy decomposition method (EDM)[25−27] attention is focused on the protein nonbonded interaction energy matrix $\mathbf{E}^{nb}$ whose elements represent the nonbonded (namely, van der Waals and electrostatic) interaction energies between residues. According to the EDM approach, the eigenvector ($\mathbf{v}_1$), also called the "first eigenvector", associated with the lowest eigenvalue (λ_1) of $\mathbf{E}^{nb}$ allows one to identify most of the crucial amino acids necessary for the stabilization of a certain protein conformation. Previous results showed that using that eigenvector, it is possible to obtain a filtered nonbonded interaction energy matrix ($\mathbf{E}^{conf}$) that recapitulates the main essential residue−residue interactions responsible for the protein conformation:

$$\mathbf{E}^{conf} = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T \quad (1)$$

where $\mathbf{v}_1^T$ is the transpose of the vector.

In the modification developed by Scarabelli et al. the solvation effects are taken into account by means of the molecular mechanics Poisson−Boltzmann surface area (MM-PBSA) method.[17] The generic elements E_{ij} of the protein nonbonded interaction energy matrices are computed as

$$E_{ij} = E_{ij}^{el} + E_{ij}^{vdW} + G_{ij}^{solv}$$

with E_{ij}^{el}, E_{ij}^{vdW}, and G_{ij}^{solv} as the electrostatic, van der Waals, and solvent contributions, respectively.[25] Energy components were calculated by averaging the energy terms along the calculated trajectories after discarding the first 10 ns considered for system equilibration.

Local Flexibility Parameter. We computed distance fluctuations DF along simulations to assess the intrinsic

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

flexibility of proteins. Distance fluctuation DF is defined as the time-dependent mean square fluctuation of the distance $\mathbf{r}_{ij}$ between Cα atoms of residues i and j:

$$DF_{ij} = \langle (r_{ij} - \langle r_{ij} \rangle)^2 \rangle$$

where brackets indicate the time-average over the trajectory.

Local flexibility (LF) was obtained by calculating the fluctuation in the distance r_{ij} between its Cα atom and the Cα atoms of neighboring residues j comprised in the interval $(i - 2, i + 2)$ along the sequence. The graphs corresponding to the LF calculations are reported in Figure S1.

Internal strain or deformation (p) is the local flexibility weighted by the average distance between residue pairs determined via MD.

Internal strain (p) of a given amino acid is defined as

$$p_i(t) = \sum_j DF_{ij} f(\langle r_{ij} \rangle)$$

where f is a sigmoidal function weighting the contribution to the sum to residues within a given cutoff, and not just along the sequence (r_{cut} = 6 Å in our calculations) from amino acid i.[28,29]

$$f(r) = 1/2(1 - \tanh(r - r_{cut}))$$

Definition of Cutoffs for Energy Decomposition and Local Distance Fluctuations. All the quantities described above have been calculated on the two replicas run for each system. The profiles obtained in the two replicas for each system showed high similarity in the absence of major conformational changes (see Supporting Information Figure S1). As a consequence, the errors over the energy and dynamic components turned out to be very low (see Supporting Information Figure S1c,d). On the basis of these observations, the results will be discussed in terms of the single system, rather than in terms of different simulation for the same system.

To establish a common reference for all systems, we define reference cutoff values for both energy decomposition analysis (energy cutoff, E_{ij}) and local distance fluctuations (dynamic cutoff, D_{ij}). The cutoff for each system is obtained by evaluating the contribution c_i of each residue i to the stabilization energy $\mathbf{E}^{conf}$ by summing all the products of the value for the i component of the $\mathbf{v}_1$ eigenvector defined in eq 1 multiplied by the values of all other j components (the components, their average values and relative fluctuations are reported in Supporting Information Figure S1a,c). The same calculation is repeated for all residues and the procedure is applied to both replicas.

$$c_i = \lambda_1 \sum_j v_i v_j$$

Finally, an overall average over all single residue contributions for both replicas is calculated and used to define the energy cutoff.

The energy cutoff is used to recapitulate the main determinants of the stabilization energy: the residues giving a contribution higher than the cutoff will thus define the networks of interactions that are most important to define the global stability of the 3D fold of the protein and/or to define regions of high local structural stability.

The same protocol is applied to the profiles of local flexibility calculated for the 2 replicas of each protein (the components, their average values and relative fluctuations are reported in Supporting Information Figure S1b,d). In physical terms, this calculation is aimed to discern the protein regions endowed with high conformational flexibility, defined by residues with local fluctuation higher than the cutoff, from those endowed with high rigidity, defined by residues with local fluctuation below the cutoff.

Reference values for simulated proteins are indicated in Table 2.

Moreover, averaged residue energies and fluctuations per single replica are provided in supplementary Table S1, with the corresponding errors (see Figure S1).

The definition of these thresholds is aimed at classifying amino acids based on their energetic and dynamic properties. We describe three clusters of residues: stability core (SC), ligand regulatory core (LRC), and dynamic domain. *Stability core* residues are characterized by high energy stabilization (E_{ij} < energy cutoff) and high rigidity (r_{ij} < dynamic cutoff); *ligand regulatory core* are amino acids with high energy stabilization and high mobility (E_{ij} < energy cutoff ; r_{ij} > dynamic cutoff); *dynamic domain* is the domain with low energy stabilization and large dynamical fluctuations (E_{ij} > energy cutoff; r_{ij} > dynamic cutoff).

RESULTS AND DISCUSSION

Analysis of the Residue−Residue Energetic Couplings in the Wild Type Protein: Rationalizing the Positions and Frequencies of Activating Oncogenic Mutations. Here, we address the question of whether the patterns of interaction that are most relevant to define the three-dimensional structure of the wild type protein and underlie its conformational properties may determine the tendency of oncogenic mutations to localize in specific regions. In this analysis, we focus on the 3D structure of the inactive conformation of the wild type.

We first analyze the distribution of the residue pair couplings that are most important in the stabilization of the 3D structure of the WT protein through the Energy decomposition method.[26] In this approach, the residue−residue interaction energy matrix, containing all pairs of nonbonded electrostatic and van der Waals terms, is simplified by eigenvalue decomposition. It was previously shown that the combination of the principal eigenvectors constitutes a simple vectorial representation of the sequence reporting on the contribution of each residue to the stabilization of the fold.[25−27] Moreover, the method provides information on the mean coupling energy between any two amino acids in the native state, revealing the networks of most stabilizing interactions, whose mutation would expectedly have an impact on the stability of protein conformations. From this analysis, an approximation to the global stabilization energy can also be recovered, which was proven to correlate with the relative stabilities of different mutants in several test proteins.[17−19] Herein, we compare the distribution of the most stabilizing residues along the sequence to the positions of oncogenic mutations.

The energy profile for WT EGFR is shown in Figure 2a. Projecting this information on the 3D structure highlights a clear distinction between the N- and C-lobes: the former appears to provide a minimal contribution to the overall stabilization energy, whereas the latter includes most of the residues determining structural stability (i.e., stability core). In particular, most of the stabilization energy peaks are clustered in secondary structure elements of the C-lobe, corresponding to helices αD/F/H/I (around a.a. ∼817/892/942/962) and to the terminal triad of the activation loop (Ala882, Leu883,

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

Table 2. Stability Core and Ligand Regulatory Core Residues per Protein

protein	stability core[a]	ligand regulatory core	energy cutoff (kJ/mol)	dynamic cutoff (nm^2)
2ITP	W817, C818, V819, Q820, I821, A822, M825, N826, L828, E829, L833, V834, H835, R836, D837, L838, A839, R841, N842, W880, M881, A882, L883, E884, Q894, S895, D896, V897, W898, S899, Y900, G901, V902, T903, V904, W905, E906, L907, R932, V943, Y944, M945, I946, M947, V948, K949, C950, W951, M952, I953, D954, A955, S957, R958, P959, K960, F961, R962, E963, L964, I965, I966, E967, F968, S969	K867, Y869, I878, K879, I886, L887, H893, I923, S924, S925, I926, L927, E928, D956	−17.8	322.4
2ITP*G719S		E868, Y869, H870, I878, I886, L887, H888, R889, I890, Y891, T892, I923, S924, S925, I926, L927, E928	−21.0	299.1
2GS7		I878, I923, S924, S925, I926, L927, E928	−15.5	247.5
2GS7*G719S		E865, I886, L887, R889, I923, S924, S925, I926, L927, E928	−14.7	208.5

[a]Stability core is common to all proteins, as defined in the main text. Energy and dynamic cutoffs are reported here as the mean value along the averaged two replicas distribution (see also Table S1 and Figure S1 in Supporting Information).

Glu884). In this context, the ALE motif of EGFR kinases (APE in other kinases) at the end of the activation loop emerges as a key energetic hub in the nonbonded interaction network (Figures 1 and 2a and c). This triad and the following αF-linker anchor the activation segment to the F-helix. This highly interacting network can arguably play a role in organizing the active structure of the protein.[8] In contrast, the N-lobe is characterized by the absence of highly stabilizing peaks.

Next, we compared this profile to the sequence distribution and frequencies of annotated kinase-activating oncogenic mutations (Figure 2a, red dots)[12] (EGFR oncogenic mutants are listed in Supporting Information Table S2). Figure 2a clearly shows that oncogenic mutations localize mainly in the region featuring residues that provide minimal energetic contribution to the overall stability of EGFR.

To further investigate this aspect, we also retrieved annotated somatic mutations for EGFR from the Cosmic database (www.sanger.ac.uk/genetics/CGP/cosmic) and performed the same analysis as described above. It is worth noting that 847 single mutations are listed in Cosmic collection: among these, 164 are repeatedly mutated amino acids out of 257 of the full catalytic domain. The height of the histograms represents the number of occurrences of individual mutations observed for a given residue (Figure 2b, Table S3). The comparison between the Cosmic-profile and the stabilization-energy profile confirms that mutations tend to cluster in the regions sharing the minimal amount of stabilization energy. Projecting this information on the EGFR structure highlights two main areas with distinct mutational propensities: most mutations concentrate on the N-terminal lobe, which is the part of the protein providing a minimal contribution to the overall 3D stabilization. Specifically, the positions with the highest occurrences of individual mutations localize around the P-loop, strands β1,2,3 together with αC-helix. Some mutations are observable also around position ∼840, corresponding to strands β6−β7, which support the activation loop (also defined as P+1 loop) (Figure 2c).

The observation that the positions most prone to oncogenic mutations correspond to the residues providing the minimal contribution to fold stability in WT EGFR suggests a functional relevance for this correlation (Figure 2): the residues that are essential for the stability of the native and active state of the protein are arguably the most conserved ones, since their modification may result in the destabilization of the structure with a consequent fall of the catalytic activity and ultimately a negative impact on cell-viability. On the other hand, positions that play a minimal part in stabilization, such as the ones constituting the N-lobe, could aptly be mutated and possibly modulate the dynamics underlying substrate recognition, binding and ultimately enzymatic activity. Indeed, many mutations cluster in or near modules of the protein that are involved in substrate recognition such as the C-helix, the P-loop, and the activation loop regions (P+1 loop) (Figure 1 and Figure 2c). Modulation of the conformational properties of these substructures can effectively tune protein functions.

The observed separation between the domains implicated in structural stability and functional dynamics may represent an efficient way to evolve improved protein activities, a concept linked to protein-polarity, recently proposed by Tawfik and co-workers to rationalize the role of mutations in protein evolvability and innovability.[30−32] Oncogenic mutations are, after all, enzyme-activating and mutated EGFRs can reasonably be considered catalytically improved, gain-of-function versions

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

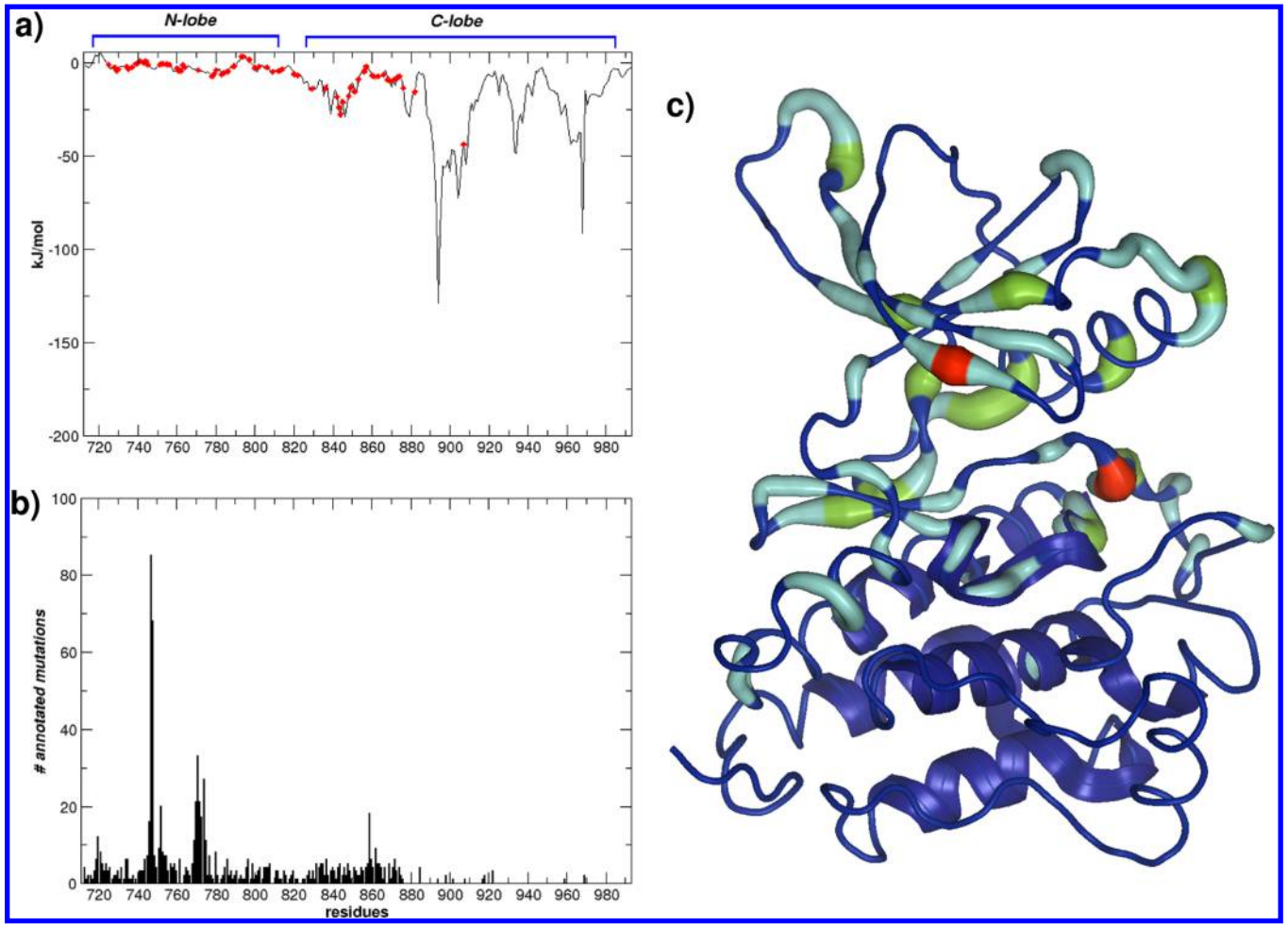


Figure 2. Energy profile and oncogenic mutations distribution. (a) Sequence-based alignment of the nonbonded interactions energy. The *wt*EGFR energy profile is derived by the energy decomposition method. Oncogenic mutations are mapped on *wt*EGFR energy profile as red diamonds [12]. The hinge region between N- and C-lone is ~800. (b) Somatic mutations collected from www.sanger.ac.uk/genetics/CGP/cosmic/. The incidence of mutational events along the sequence (a.a. 712−968 from the EGFR numeration) is reported as the counts of single mutations annotated per residue (see main text and Table S3 in the Supporting Information). For comparison, kinase catalytic domain is graphed. (c) Projection of oncogenic mutations on the three-dimensional structure of EGFR kinase. Ribbon thickness and color codes are based on the number of mutational occurrences, going from the least (blue and thin) to the most mutated (red and thick) residues. Note that most of the mutations clusterize within the N-domain. Blue cartoons within the C-domain represent the stability core (see main text).

of the native enzyme. In this framework, the coexistence of a flexible recognition and active site region and a stable structural core favors the selection of mutations that do not perturb the global structural stability of the system, determined by the C-lobe and thus have a higher probability of being tolerated. Mutations outside the structural core, targeted to the loosely stabilized N-lobe, modulate the conformational plasticity of this domain, which may in turn facilitate the conformational rearrangements leading to kinase activation. The modular combination of domains with different properties in one single molecule has also been observed as a key contributor to evolution, as seen in the assembly of multidomain proteins by recombination of existing units.[30−32] Hence, we hypothesize that the effect of an oncogenic mutation might be reflected by a modulated conformational dynamics of protein regions involved in recognition, and on the other hand by a fairly unperturbed structural core. In order to test this hypothesis, we set out to describe the energetics and dynamics of WT kinase and, for comparison, of one activated mutant, with the aim of highlighting the aspects of EGFR conformational dynamics that can be linked to functional activation and might differ in the two systems.

Functional Dynamics of the EGFR: Effects of Activating Mutation and Characterization of the Allosteric Regulation Mechanisms. In order to describe the conformational dynamics of EGFR, we focus first on the rigid dynamics properties of the protein, by evaluating residue−residue coordination throughout the whole structure. This analysis can provide information on global motions related to function and also possible mechanisms of allosteric control in EGFR kinases.[28] The latter represent an important aspect of functional regulation determined by coordinated protein dynamics, which in particular can be modulated by sequence variations.

To this end, we carried out a comparative analysis of the dynamics of the active and inactive conformations of both the wild type protein and a selected activating mutant (G719S), thus taking into account four systems, to highlight possible modulations of the internal dynamics that can be linked to the activation propensity.

We calculated the distance fluctuations (DF) map considering pairs of C-alpha atoms (see Methods). The results are

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

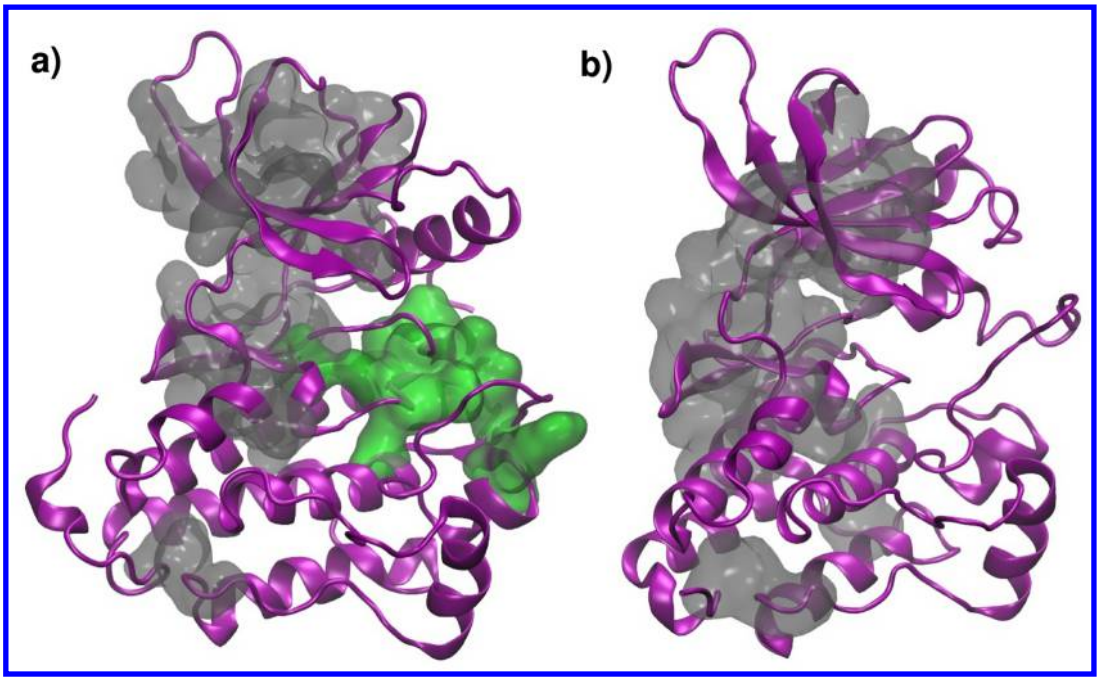


Figure 3. Long-range coordination. Inter-residue coordination is calculated by means of distance fluctuations. (a) Active and (b) inactive state EGFR kinase. Only high coordinated residues (DF < 0.04 nm^2) and not contiguous residues are displayed. Kinase structure is rendered in cartoon, and the coordination path is in ghost surface representation. The common route is displayed in gray whereas the additional path in the active kinase is in green (a).

shown in Figure S2 of the Supporting Information. The matrices exhibit the block character typical of multidomain proteins, with regions of low distance fluctuations indicating coordinated subdomains. A detailed analysis of this matrix identifies patterns of coordination between physically distant residues. Specifically, we focused on residue pairs whose DF is lower than the threshold of 0.04 nm^2 and that are separated by at least 10 residues in the primary structure.

A set of residues that are consistently coordinated long-range with the active site, in all the simulations, coincides with the myristyl-binding site in the C-lobe, at the interface of the αI−αH−αF bundle. Interestingly, energy decomposition has shown that these helices are part of the stability core of the protein (Figure 2). Binding of ligands to the myristyl site has been shown to allosterically regulate kinase functions:[11,13,33] in our model, the stability and preorganization of the scaffold helices provides an efficient route to relay the allosteric signal to the distal active site, through well structured, stable semirigid structural subdomains. In all four simulated systems, coordinated motions are also observed between the hydrophobic component of the N-lobe β-sheet and the C-lobe through the connecting hinge. This coordination appears essential in ensuring the correct structural organization independent of activation-related changes.

In contrast, in the conformations representing the active forms of both sequences (WT and G719S), this analysis reveals coordinated fluctuations also for portions of the short helices of the myristyl site, which are absent in the WT inactive conformation. Strikingly, this additional coordination is also found in the activating mutant starting from the inactive structure. Projecting the information on the 3D structures, one can visualize a continuous pathway connecting the N-domain and the myristyl site of the C-lobe (Figure 3).

As a general model, we speculate on the existence of multiple allosteric coordination pathways in EGFR structure. One such pathway is common to both the inactive and active states. The second one pertains to the systems linked to a functionally activated profile of the kinase: indeed, the second allosteric pathway is observed in the presence of the gain-of-function mutation (in the inactive and active structures) and in the WT active conformation. Interestingly, this involves the activation loop (P+1 loop). The presence of the second allosteric communication on the one hand may represent an additional layer of regulation, and, on the other hand, may render the communication between the two lobes in the activated state more efficient providing additional routes for the transmission of a ligand-encoded chemical signal. Interestingly, comparisons of the determinants of stability between the open and closed conformations show that the mutation stabilizes the regions around the myristyl site (Table 2).

Overall, the description of long-range coordination suggests an activation-related behavior and could support the idea of a sophisticated mechanism of allosteric regulation. In this framework, coordinated and highly interactive patches in active proteins can sense local modifications (such as myristoylation) and orchestrate the conformational changes underlying activity. The mutation modulates the intensity of these coordination patterns, which ultimately translates in variation of the levels of activity of the kinase. Moreover, inter-residue coordination analysis captures amino acids from two hydrophobic regions, called respectively catalytic and regulatory spines, first described by Taylor and co-workers as discriminatory elements in the activation state of the kinase.[8] The catalytic spine is always coordinated, connecting β2 and β7, and anchoring the C-terminus of the F-helix. The regulatory spine, composed by two amino acids from the N-lobe (β4 and C-helix) and two from the C-lobe (activation and catalytic loops) is well coordinated only in the active kinases (activated mutant and active states), due to the relative displacement of C-helix and activation loop.

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

Defining Local Substructures with Functional Relevance. To gain further insight into the links between structural organization, functional dynamics and the impact of activating mutations, we computed the fluctuations of pairwise amino acid distances in the MD trajectories of the WT protein. Here, we specifically concentrated on characterizing the local flexibility parameter (see Methods): for each residue i, local flexibility was obtained by calculating the fluctuation in the distance r_{ij} between its Cα atom and the Cα atoms of neighboring residues j comprised in the interval $(i - 2, i + 2)$ along the sequence.[28,29] The local flexibility profiles for the WT protein (inactive and active structures) and for the gain-of-function mutant (inactive and active structures) are shown in Figure 4.

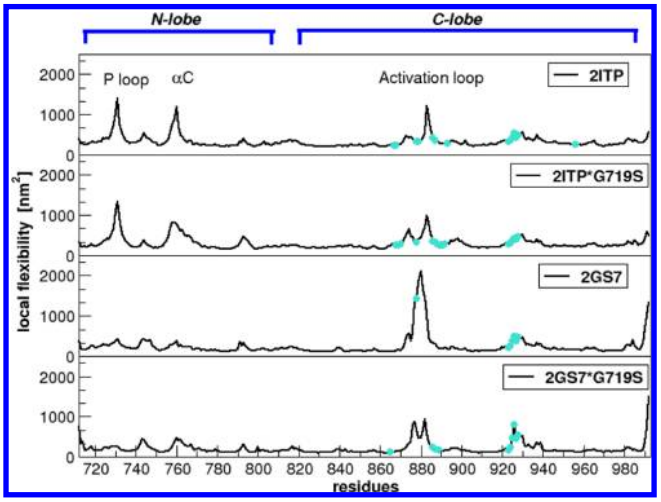


Figure 4. Local distance fluctuation profiles for the simulated EGFR kinase systems (WT and mutant, active, and inactive forms). Peaks correspond to the highly flexible loop regions (P-loop and activation loop) and to the αC helix. Turquoise spheres indicate the LRC residues.

Focusing first on the inactive WT system, by combining the results of this calculation with the energy decomposition analysis (see Methods), we can define three different areas within the fold. The cutoffs for the definition of stabilizing and flexible regions are described in the Methods section (see also Figure S1 and Table S1 in the Supporting Information). The first region is defined as the **stability core**, which corresponds to the stable nucleus previously identified and consists of residues with high stabilization energy contributions and minimal conformational flexibility (high rigidity), including WT residues spanning regions 817−842, 880−907, and 943−969 (detailed residues lists are given in Table 2). The second region is defined by residues with high stabilization energy contributions and high conformational flexibility, which entails WT residues 867, 869, 878, 789, 886, 887, 893, 923−928, and D956; we labeled it the LRC because it localizes near the catalytic area of the protein, namely within the activation loop, close to the ALE αF helix linker acting as an anchor to the C-lobe and spanning the solvent exposed part of the αG helix (Figure 5). This substructure is involved in recognition and binding of substrates. According to our analysis, the LRC acts as a small, structurally preorganized unit, as defined by the observed strong residue-couplings that determine its structural stability. We hypothesize that the LRC controls the mechanisms of conformational selection and adaptation to the substrate by means of its high local flexibility.

The third region we identify consists of residues with minimal stabilization energy contributions and high flexibility, the **dynamic domain**: it includes mainly residues of the N-domain and very flexible and solvent exposed fragments of the C-lobe (Figures 4 and 5).

Overall, the integration of dynamic information with the previously described energy decomposition analysis highlights the relevance of a compact and rigid substructure for the overall stability of the 3D fold, which is well separated from amino acids mostly endowed with functional roles that require structural plasticity for recognition and processing of substrates. The location of oncogenic mutations along the protein (Figures 1 and 2) in the dynamic domain suggests that the stability of the protein is in fact unaffected. Yet, given that such mutations have a functional activating effect, they might modify the structural dynamics of the functional parts of the protein defining the machinery devoted to recognize and process substrates, namely the LRC. Indeed, the modulation of the rigid dynamics of this region was observed in the G719S mutant as discussed in the previous section. Here we address the question of how the decomposition into stability core, LRC, and dynamic domain is affected in the G719S system, again combining the energetic and local flexibility profiles of EGFR, considering the active and inactive conformations and comparing the results to the respective conformations of the wild type.

As expected, the mutation does not affect the energetic and dynamic properties of the scaffold (stability core) region, neither in the active nor in the inactive conformations. This result confirms the role of this substructure as a common nucleus dedicated to determine the three-dimensional architecture of the protein. In contrast, we observe an extension of the area defined as the ligand regulatory core (Figures 5 and 6 and Table 2), in which high intramolecular energetic couplings are paralleled by high flexibility. Interestingly, in the mutant, the substructure defined as LRC partially overlaps with the activation (P+1) loop previously shown to play a role in allosteric control. These results indicate that the mutation has a specific impact on all the protein machinery dedicated to the interaction with substrates.

From the point of view of chemical properties, the mutation causes the replacement of the small glycine within the P-loop for a bulkier and more hydrophilic serine. This impacts the overall dynamics of the β-sheet core of the N-lobe and in particular of the $\beta1-\beta2$ loop, leading to the reordering of the surrounding interactions. The rearrangement of the N-lobe propagates toward the C-domain, ending with increased coordination of the activation loop: serine creates a novel interaction pattern that blocks the displacement of the C-helix and favors the open shape of the ATP-binding pocket. In particular, hydrogen bonds with T725, V726, R841, and D855 around serine contribute to stabilize this conformation. This conformation, compatible with the active state, is further stabilized by a network of salt bridges that keep the P+1 loop fully solvated and extended, so that it can recruit the substrate and simultaneously leave ATP and binding cavity accessible. In the WT protein, the two salt bridges between D855 and K745/R841, E872 and R836, and between E758 and K875 orchestrate the folding of activation loop over the ATP binding pocket. In the mutant, D761 bonded to K860/K757 and E758 bonded to K860 favor the extension of this loop.

To evaluate the propensity of the mutated protein toward an activated state, we monitored the time-evolution of the salt

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

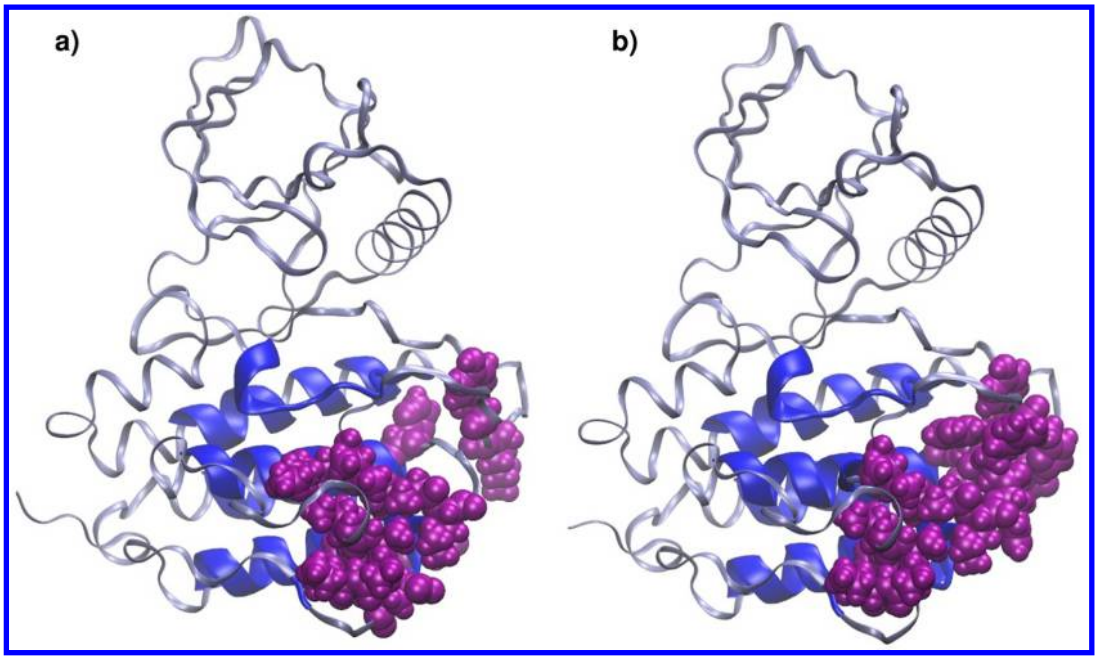


Figure 5. Ribbon 3D representation of active state kinase EGFR. Projection of the SC and LRC over *wild-type* (a) and mutated kinase (b). SC and LRC are shown in blue cartoons and magenta cpk, respectively. The stability core is common to all kinases (SC and LRC residues are listed in Table 2), while LRC depends on the activation state of the protein. It is worth noting that LRC is extended in the mutants. This increment is larger in the case of the open state kinase (2ITP*G719S), where it makes a continuous area extending over the activation loop (b).

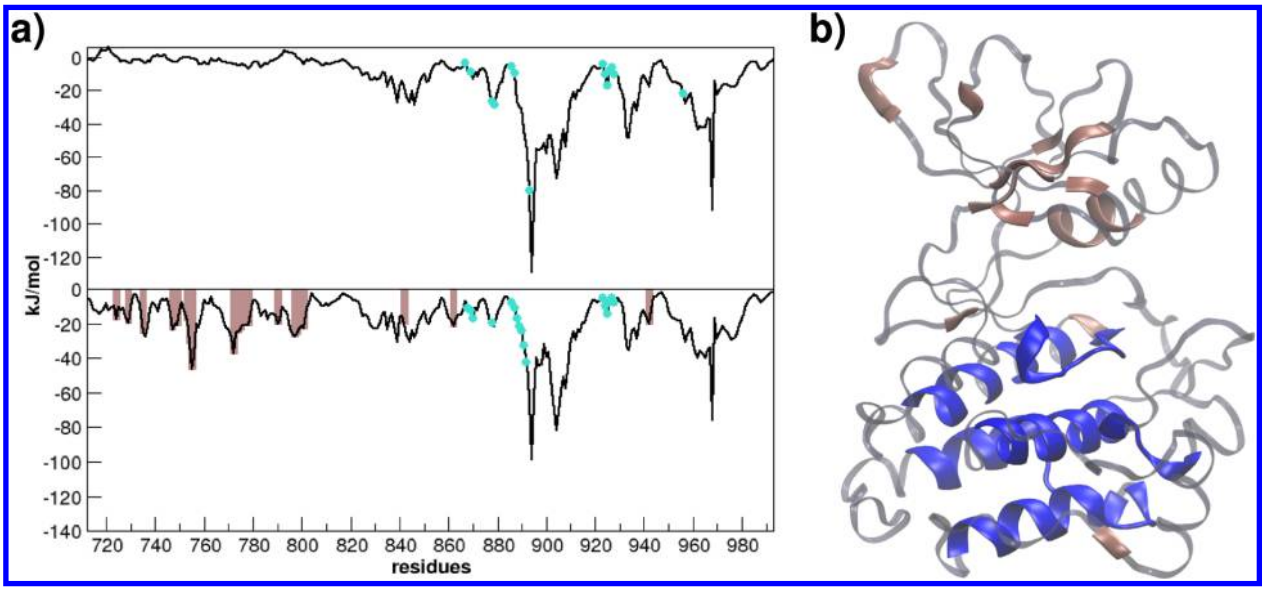


Figure 6. Effect of the mutation on the energy profile. (a) The upper panel displays the *wild type* 2ITP energy profile. Almost the totality of the strongest residue-couplings localize within the C-lobe. In the lower panel, the mutated kinase energy profile is plotted. The G719S mutation clearly affects the N-lobe distribution of nonbonded energy: peaks at ∼720 a.a. and ∼750 a.a. indicate P loop and αC helix stabilization, respectively. (b) 3D representation of newly stabilized regions. Enrichment is indicated by brown segments. The blue cartoon represents the scaffold, as already defined. Turquoise spheres indicate the LRC residues.

bridge between αC-helix Glu762 and β3 Lys745 (Figure S3 in the Supporting Information). The function of this interaction is to orient and stabilize the C-helix in the active state conformation. Indeed, the WT type and mutation induced evolutions show notable differences. In WT-EGFR this interaction displays a higher degree of fluctuation, while it is largely stable in the case of G719S mutation. It is worth mentioning at this point that for inactive structures, this salt-bridge is prevented by the hindrance of the small helix made by the activation loop.

Our discussion has been focused on the analysis of the energetics and dynamics of the folded states of kinases. However, it is important to point out that unfolded states can play a role in determining the observed properties of proteins. Specifically, in this case, the low propensity of Gly for beta-sheet conformations (present in the native state) may result in a higher relative stabilization of the unfolded state compared to the Ser mutant, assuming that the local conformation in the unfolded ensemble is non beta-sheet, with a consequent destabilization of the catalytically active conformation. Serine,

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

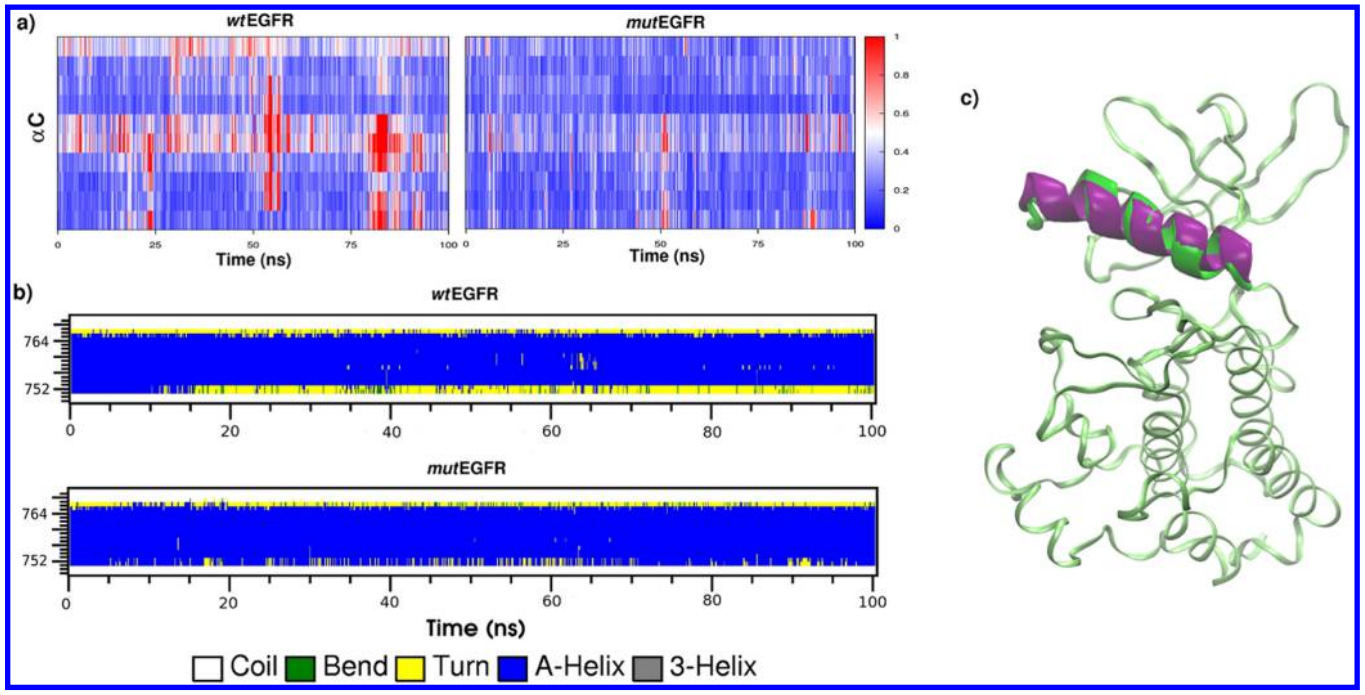


Figure 7. αC internal strain and secondary structure time-evolution for *wt*EGFR (2ITP) and *mut*EGFR (2ITP*G719S). (a) Local distortion along the simulation time refers to the αC sequence for wild type and mutated EGFR kinase. Red strips indicate increased flexibility or higher structural deformation, suggesting the starting unfolding of the helix. (b) αC helix secondary structure evolution along the simulation. Yellow stripes indicate coiled regions where the helix (blue) is lost in *wt*EGFR and *mut*EGFR kinases. Yellow bands approximately indicate a 3-residues turn of the helix. (c) 3D superposition of αC helix in *wt*EGFR (green) and *mut*EGFR (purple). An additional turn helix is visible at the top of the mutated kinase.

on the other hand, may stabilize the beta-sheet in the folded state to a higher extent thanks to its higher beta-sheet propensity, thus favoring the preferential population of the native conformation. To gain quantitative insight into these aspects of the stability properties of kinases, one should in principle be able to fully characterize the unfolded ensembles of these proteins. This is however still out of reach, even in the case of dedicated software or accelerated sampling methods.

As a caveat, one should clearly consider these points as potential limitations of our analysis.

Overall, based on the available data, we conclude that the mutation is adding structural coordination to the C-helix region, and this is compatible with a more stable activated state, consistent with previous research.[7,9,15,16,34]

To connect energetic stabilization, local structural rearrangement and dynamic modulation we calculated the internal strain and the evolution of secondary structure content for C-helix in the WT and in the mutant. The internal coordination of the helix as well as its ordered secondary structure content increase and are stable over the whole simulation time in the mutant compared to the WT (Figure 7). Consistent with these observations, the energy decomposition analysis of the G719S mutant in the active conformation shows a remarkable increase in the contribution of the N-lobe residues to the stability of the protein (Figure 6), with the αC-helix characterized by higher peaks, in line with the results obtained by Gervasio and collaborators through metadynamics-based free energy calculations.[16,35] In the inactive state simulation, the mutation also affects the stabilization of N-lobe elements, albeit to a smaller extent (Figure S1 in the Supporting Information).

These considerations suggest a possible mechanism by which the mutation favors the preferential population of the kinase active state. The stabilization of the C-helix and extension of the LRC to include part of the P+1 loop support activation through the onset of local interactions that reduce the overall conformational heterogeneity determined by local folding-unfolding events: these stabilized substructures can fluctuate as coherent rigid bodies and select conformations apt for substrate recognition, while avoiding potentially unproductive random conformational changes. In this picture, the effect of the mutation reverberates in a more efficient mechanism of exploration and selection of the structural ensembles necessary to assume a catalytically competent conformation.[9]

A second important point deserves consideration: it has been shown that EGFR dimerization upon ligand binding is crucial to regulate cellular activity. Experimental structures of closely related systems,[10,36−39] as well as detailed structural models obtained by means of microsecond long simulations,[15,34] have linked EGFR activation to the formation of an asymmetric dimer formed by the juxtaposition of the C-lobe of one kinase domain to the N-lobe of the other kinase domain. In this model, the αC-helix in the N-terminal domain of one of the two partners extensively interacts with the C-lobe of the second one.[10] The observed preorganization of the αC-helix in the monomer induced by the G719S mutation can arguably play a critical role in the dimerization process by decreasing the entropic penalty associated with the ordering of the protein−protein interaction interface.[15,16,34−39] In this case, the point of mutation is not directly located at the interaction interface, so that the effect on the interaction is indirect and allosteric in nature. This model is consistent with the results of the Shan group,[7] showing that some cancer-linked mutations, though distal to the interaction surface, facilitate EGFR dimerization by suppressing the local disorder at the dimerization interface which otherwise characterizes the wild type form of the protein. Thus, based on our findings, we can reasonably assume that

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

gained stabilization induces loss of disorder, which can ultimately preorganize EGFR for dimerization.[15,34−39]

Overall, our results indicate that the Gly to Ser mutation determines an increase of the relative contribution of the N-lobe and LRC residues to the global stability of the protein (see Figure 6), which can aptly translate into an increased structural preorganization that may favor substrate recognition and processing as well as the dimerization processes linked to kinase activation.

CONCLUSIONS

In conclusion, the results presented in this study provide support for the efficacy of strategies combining novel structural, dynamic and physicochemical investigations to shed new light on the determinants of protein stability and dynamics and their role in defining which regions of a protein may be more apt to host mutations with an impact on activity. Herein, we have used the EGFR kinase as a test system. In particular, the results provide a physical rationale for the observed clustering of activating mutations in the N-terminal region of the protein. Moreover, they define significant differences between the wild type and a representative activated mutant in their mechanisms of dynamic coordination, in possible allosteric mechanisms across the architecture of the protein, and in the energy and dynamic networks responsible for the stabilization of the active conformations. Overall, our analysis defines distinctive regions in the WT and mutant that can be helpful in designing allosteric modulators selective for the oncogenically mutated EGFR, by way of targeting substructures whose structural, dynamic and energetic properties are differentially perturbed by sequence variations.

ASSOCIATED CONTENT

Supporting Information

This material contains Figure S1, reporting the profiles on which the energy and dynamic properties have been calculated; Figure S2 reporting the DF matrices for the simulated systems; Figure S3 reporting on the time evolution of the distance between the αC Glu and β3 Lys; Table S1 with average energies, fluctuations and error estimations; and Table S2 describing annotated oncogenic mutations; and Table S3. Annotated somatic mutations for EGFR kinase catalytic domain (a.a. 712-968). The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00270.

AUTHOR INFORMATION

Corresponding Author

*E-mail: g.colombo@icrm.cnr.it. Tel.: +39-02-28500031. Fax: +39-02-28901239.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

G.C. acknowledges funding from AIRC (Associazione Italiana Ricerca sul Cancro) through the grant: IG 15420; from Fondazione Cariplo through grant 2011.1800 for the RST call—"Premio fondazione cariplo per la ricerca di frontiera"; the PRACE project 2012071270 for allocation of computing time.

REFERENCES

(1) Dixit, A.; Verkhivker, G. M. 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.* **2009**, *5*, e1000487.

(2) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298*, 1912−1934.

(3) Futreal, P. A.; Coin, L.; Marshall, M.; Down, T.; Hubbard, T.; Wooster, R.; Rahman, N.; Stratton, M. R. A Census of Human Cancer Genes. *Nat. Rev. Cancer* **2004**, *4*, 177−183.

(4) Greenman, C.; Stephens, P.; Smith, R.; Dalgliesh, G. L.; Hunter, C.; Bignell, G.; Davies, H.; Teague, J.; Butler, A.; Stevens, C.; Edkins, S.; O'Meara, S.; Vastrik, I.; Schmidt, E. E.; Avis, T.; Barthorpe, S.; Bhamra, G.; Buck, G.; Choudhury, B.; Clements, J.; Cole, J.; Dicks, E.; Forbes, S.; Gray, K.; Halliday, K.; Harrison, R.; Hills, K.; Hinton, J.; Jenkinson, A.; Jones, D.; Menzies, A.; Mironenko, T.; Perry, J.; Raine, K.; Richardson, D.; Shepherd, R.; Small, A.; Tofts, C.; Varian, J.; Webb, T.; West, S.; Widaa, S.; Yates, A.; Cahill, D. P.; Louis, D. N.; Goldstraw, P.; Nicholson, A. G.; Brasseur, F.; Looijenga, L.; Weber, B. L.; Chiew, Y. E.; DeFazio, A.; Greaves, M. F.; Green, A. R.; Campbell, P.; Birney, E.; Easton, D. F.; Chenevix-Trench, G.; Tan, M. H.; Khoo, S. K.; Teh, B. T.; Yuen, S. T.; Leung, S. Y.; Wooster, R.; Futreal, P. A.; Stratton, M. R. Patterns of Somatic Mutation in Human Cancer Genomes. *Nature* **2007**, *446*, 153−158.

(5) Huse, M.; Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **2002**, *109*, 275−282.

(6) Fedorov, O.; Müller, S.; Knapp, S. The (Un)Targeted Cancer Kinome. *Nat. Chem. Biol.* **2010**, *6*, 166−169.

(7) Shan, Y.; Eastwood, M. P.; Zhang, X.; Kim, E. T.; Arkhipov, A.; Dror, R. O.; Jumper, J.; Kuriyan, J.; Shaw, D. E. Oncogenic Mutations Counteract Intrinsic Disorder in the EGFR Kinase and Promote Receptor Dimerization. *Cell* **2012**, *149*, 860−870.

(8) Taylor, S. S.; Kornev, A. P. Protein Kinases: Evolution of Dynamic Regulatory Proteins. *Trends Biochem. Sci.* **2011**, *36*, 65−77.

(9) Marino, K. A.; Sutto, L.; Gervasio, F. L. The Effect of a Widespread Cancer-causing Mutation on the Inactive to Active Dynamics of the B-Raf Kinase. *J. Am. Chem. Soc.* **2015**, *137*, 5280−5283.

(10) Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell* **2006**, *125*, 1137−1149.

(11) Bastidas, A. C.; Pierce, L. C.; Walker, R. C.; Johnson, D. A.; Taylor, S. S. Influence of N-Myristylation and Ligand Binding on the Flexibility of the Catalytic Subunit of Protein Kinase A. *Biochemistry* **2013**, *52*, 6368−6379.

(12) Dixit, A.; Verkhivker, G. M. The Energy Landscape Analysis of Cancer Mutations in Protein Kinases. *PLoS One* **2011**, *6*, e26071.

(13) Mcclendon, C. L.; Kornev, A. P.; Gilson, M. K.; Taylor, S. S. Dynamic Architecture of a Protein Kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E4621−E4631.

(14) Schlessinger, J. Dimerization and Activation of EGF Receptor. *Cell* **2002**, *110*, 669−672.

(15) Shan, Y.; Arkhipov, A.; Kim, E. T.; Pan, A. C.; Shaw, D. E. Transitions to Catalytically Inactive Conformations in EGFR Kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 7270−7275.

(16) Sutto, L.; Gervasio, F. L. Effects of Oncogenic Mutations on the Conformational Free-Energy Landscape of EGFR Kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10616−10621.

(17) Scarabelli, G.; Morra, G.; Colombo, G. Predicting Interaction Sites from the Energetics of Isolated Proteins: A New Approach to Epitope Mapping. *Biophys. J.* **2010**, *98*, 1966−1975.

(18) Peri, C.; Gagni, P.; Combi, F.; Gori, A.; Chiari, M.; Longhi, R.; Cretich, M.; Colombo, G. Rational Epitope Design for Protein Targeting. *ACS Chem. Biol.* **2013**, *8*, 397−404.

(19) Moroni, E.; Zhao, H.; Blagg, B. S. J.; Colombo, G. Exploiting Conformational Dynamics in Drug Discovery: Design of C-Terminal Inhibitors of Hsp90 with Improved Activities. *J. Chem. Inf. Model.* **2014**, *54*, 195−208.

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387

(20) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(21) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules ? *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(23) Hess, B.; Bekker, H.; Herman, J. C.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(24) Darden, T.; Darrin York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(25) Genoni, A.; Morra, G.; Colombo, G. Identification of Domains in Protein Structures from the Analysis of Intramolecular Interactions. *J. Phys. Chem. B* **2012**, *116*, 3331−43.
(26) Tiana, G.; Simona, F.; De Mori, G. M. S.; Broglia, R. A.; Colombo, G. Understanding the Determinants of Stability and Folding of Small Globular Proteins from their Energetics. *Protein Sci.* **2004**, *13*, 113−124.
(27) Morra, G.; Colombo, G. Relationship between Energy Distribution and Fold Stability: Insights from Molecular Dynamics Simulations of Native and Mutant Proteins. *Proteins: Struct., Funct., Genet.* **2008**, *72*, 660−672.
(28) Morra, G.; Verkhivker, G.; Colombo, G. Modeling Signal Propagation Mechanisms and Ligand-Based Conformational Dynamics of the Hsp90 Molecular Chaperone Full-Length Dimer. *PLoS Comput. Biol.* **2009**, *5*, e1000323.
(29) Morra, G.; Potestio, R.; Micheletti, C.; Colombo, G. Corresponding Functional Dynamics Across the Hsp90 Chaperone Family: Insights from a Multiscale Analysis of MD Simulations. *PLoS Comput. Biol.* **2012**, *8*, e1002433.
(30) Dellus-Gur, E.; Toth-Petroczy, A.; Elias, M.; Tawfik, D. S. What Makes a Protein Fold Amenable to Functional Innovation? Fold polarity and stability trade-offs. *J. Mol. Biol.* **2013**, *425*, 2609−2621.
(31) Tokuriki, N.; Tawfik, D. S. Stability Effects of Mutations and Protein Evolvability. *Curr. Opin. Struct. Biol.* **2009**, *19*, 596−604.
(32) Tokuriki, N.; Tawfik, D. S. Protein Dynamism and Evolvability. *Science* **2009**, *324*, 203−207.
(33) Dixit, A.; Verkhivker, G. M. Computational Modeling of Allosteric Communication Reveals Organizing Principles of Mutation-Induced Signaling in ABL and EGFR kinases. *PLoS Comput. Biol.* **2011**, *7*, e1002179.
(34) Arkhipov, A.; Shan, Y.; Das, R.; Endres, N. F.; Eastwood, M. P.; Wemmer, D. E.; Kuriyan, J.; Shaw, D. E. Architecture and Membrane Interactions of the EGF Receptor. *Cell* **2013**, *152*, 557−569.
(35) Lovera, S.; Sutto, L.; Boubeva, R.; Scapozza, L.; Dölker, N.; Gervasio, F. L. The Different Flexibility of c-Src and c-Abl Kinases Regulates the Accessibility of a Druggable Inactive Conformation. *J. Am. Chem. Soc.* **2012**, *134*, 2496−2499.
(36) Zhang, X.; Pickin, K. A.; Bose, R.; Jura, N.; Cole, P. A.; Kuriyan, J. Inhibition of the EGFR Receptor by Binding to an Activating Kinase Domain Interface. *Nature* **2007**, *450*, 741−744.
(37) Jura, N.; Zhang, X.; Endres, N. F.; Seeliger, M. A.; Schindler, T.; Kuriyan, J. Catalytic Control in the EGF Receptor and its Connection to General Kinase Regulatory Mechanisms. *Mol. Cell* **2011**, *42*, 9−22.
(38) Jeffrey, P. D.; Russo, A. A.; Polyak, K.; Gibbs, E.; Hurwitz, J.; Massague, J.; Pavletich, N. P. Mechanism of CDK Activation Revealed by the Structure of a Cyclin-CDK2 Complex. *Nature* **1995**, *376*, 313−320.
(39) Roskoski, R., Jr. ErbB/HER Protein-Tyrosine Kinases: Structures and Small Molecule Inhibitors. *Pharmacol. Res.* **2014**, *87*, 42−59.

DOI: 10.1021/acs.jcim.5b00270
J. Chem. Inf. Model. 2015, 55, 1377−1387