Dynamics Fingerprints of Active Conformers of Epidermal Growth Factor Receptor Kinase

German P. Barletta, Marcia Anahi Hasenhuer, Maria Silvina Fornasari, Gustavo Parisi, and Sebastian Fernandez-Alberti*

Epidermal growth factor receptor (EGFR) is a prototypical cell-surface receptor that plays a key role in the regulation of cellular signaling, proliferation and differentiation. Mutations of its kinase domain have been associated with the development of a variety of cancers and, therefore, it has been the target of drug design. Single amino acid substitutions (SASs) in this domain have been proven to alter the equilibrium of pre-existing conformer populations. Despite the advances in structural descriptions of its so-called active and inactive conformations, the associated dynamics aspects that characterize them have not been thoroughly studied yet. As the dynamic behaviors and molecular motions of proteins are important for a complete understanding of their structure–function relationships we present a novel procedure, using (or based on) normal mode analysis, to identify the collective dynamics shared among different conformers in EGFR kinase. The method allows the comparison of patterns of low-frequency vibrational modes defining representative directions of motions. Our procedure is able to emphasize the main similarities and differences between the collective dynamics of different conformers. In the case of EGFR kinase, two representative directions of motions have been found as dynamics fingerprints of the active conformers. Protein motion along both directions reveals to have a significant impact on the cavity volume of the main pocket of the active site. Otherwise, the inactive conformers exhibit a more heterogeneous distribution of collective motions. © 2018 Wiley Periodicals, Inc.

DOI:10.1002/jcc.25590

Introduction

The epidermal growth factor receptor (EGFR) is a key protein in cellular signaling, and regulation of cell proliferation, differentiation, and migration.[1] EGFR is part of the family of ERBB cell membrane receptor proteins. It is composed by, following the human EGFR canonical amino acid sequence numbering (Universal Protein Resource, UniProtKB, P00533, isoform 1), an extracellular receptor (residues 25–641), a single-pass transmembrane helix (642–668), and a cytoplasmic region harboring a membrane proximal juxtamembrane segment (669–711) followed by a kinase domain (N-lobe: 712–792, C-lobe: 793–979) and a 231 residue long C-terminal intrinsically disordered tail (980–1210) that works as a recruiter of cell signaling molecules through its many phosphorylatable sites. Mutations and overexpression of EGFR have been associated to different human cancers[2,3] with kinase activity enhancement, reason why the EGFR kinase region is an intensely pursued target of small-molecule drugs.[4–7] Single amino acid substitutions (SASs), deletions and insertions in this domain can alter the equilibrium of pre-existing conformer populations.[8–12] In this way, certain EGFR kinase cancer mutants become enzymatically more active than the wild-type.[13,14] These SAS are denominated as activating, due to the stabilization of the conformer required to drive the phosphorylation (active form).

The current accepted mechanism of activation of wild-type EGFR,[15–17] proposes that the binding of factors like EGF (epidermal growth factor) to the extracellular receptor, shifts the conformational equilibrium to conformers that are able to establish specific interactions between each other in extracellular, transmembrane, and juxtamembrane regions, approaching the kinase domains in an asymmetric dimer way, facing the C-lobe of one kinase (activator) with the juxtamembrane and N-lobe of the other (activated). This way, the last becomes catalytically active, being able to phosphorylate the Tyrosines present in the C-terminal tail of the activator. Besides, the presence of intrinsic disorder in some regions of the EGFR kinase domain has been proposed to play a significant role in the EGFR dimerization.[18]

The detailed understanding at a molecular level of the structural and dynamical differentiation among wild-type and cancer mutants of EGFR kinase is a very important goal in the development of personalized cancer treatments.[19] In particular, it would be most enlightening to be able to elucidate structural and dynamical features that represent fingerprints of kinase activation.

There are common sequence and structural characteristics shared among most of active EGFR kinase domain have been identified[20,21] (see Fig. 1): (a) α C-helix rotated inwards against the N-lobe and toward the active site, allowing a salt-bridge

G. P. Barletta, M. A. Hasenhuer, M. S. Fornasari, G. Parisi, S. Fernandez-Alberti
Departamento de Ciencia y Tecnologia, Universidad Nacional de Quilmes/CONICET, Roque Saenz Peña 352, B1876BXD, Bernal, Argentina
E-mail: sfalberti@gmail.com

Contract Grant sponsor: Universidad Nacional de Quilmes; Contract Grant number: PUNQ 1402/15

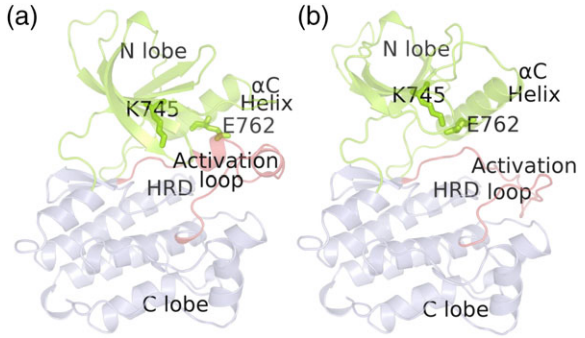

Figure 1. a) Inactive, and b) active EGFR kinase conformers indicating the main structural elements that commonly distinguish them. [Color figure can be viewed at wileyonlinelibrary.com]

between residues E762 and K745[20]; (b) a more extended and open conformation of the activation loop (residues 855–886), so-called DFG-in conformation with its aspartate pointing to the ATP-binding site. Besides these two main features, the triad HRD (residues 835–837)[22,23] and different amino acid networks seem to contribute to the ATP active site stabilization.[24–26] Nevertheless, not all of active conformers fulfill all of this requirements,[27] and several reported controversies and ambiguous conformation classification can be found.[28] In a previous article,[29] we have extensively analyzed these structural parameters in a large set of human EGFR kinase domains previously classified as active or inactive forms. We have focused our attention on changes in the size and shape of the main pocket among the EGFR kinase active and inactive conformers. For this purpose, a hierarchical clustering based on RMSD (root mean square difference) of α-C belonging to main pocket positions has been performed. Our main pocket structural comparison allows us to analyze effects that changes in the main pockets' structural features have on the EGFR kinase activity.

Despite advances in structural descriptions of the so-called active and inactive conformations,[21,30–33] the associated dynamics aspects that characterize them have not been thoroughly studied yet. As the dynamic nature of proteins is essential to their function, a complete description should consider their motions.[34] Using unbiased long-timescale molecular dynamics simulations, Shan et al.[35] have explored the transitions from the active to the inactive conformations of a single EGFR mutant. Besides, the intrinsic disorder of EGFR kinase domain has been shown to play a significant role in the EGFR dimerization.[18] Differences between active and inactive states have been also studied using MD followed by principal component analysis (PCA),[36,37] finding that EGFR is more flexible in the active state than in the inactive state. Other MD studies have been centered on the analysis of SASs on the relative stability of the active versus the inactive state,[38] and also on the conformational transition between active and inactive conformers.[39,40]

In the present study, we use normal mode analysis (NMA) to explore global dynamics aspects that differentiate these conformers. NMA based on a coarse-grained model of proteins has been extensively proved as a useful technique to explore the intrinsic dynamics within a folded state.[41–45] The complex motions of proteins are decoupled into a linear combination of independent harmonic oscillators, that is, the normal modes. Low-frequency normal modes correspond to the most collective ones, involving the concerted motions of many atoms. These modes are commonly related to functional aspects of a protein[46–49] and they have been proven to be robust to sequence variations, that is, they are evolutionary conserved.[50–54]

The fact that normal modes provide a decoupled harmonic description of protein vibrations is fundamental to identify common dynamics aspect shared by a set of related mutant proteins.[53] Nevertheless, small local structural perturbations, introduced by SASs, can significantly modify the global dynamics and functionality of proteins.[55]

Herein, we present a novel procedure that allows us to identify collective dynamics shared among different conformers in EGFR kinase. The method allows the comparison of patterns of low-frequency vibrational modes defining representative directions of motions that can be considered as dynamics fingerprints of the active conformers.

The article is organized as follows: theoretical methods are described in **methods** section. Our results are presented and discussed in **results and discussion** section. Finally, **conclusions** section summarizes our findings and conclusions.

Methods

Set of active and inactive EGFR kinase mutant structures

Our set of active and inactive EGFR kinase structures has been selected from CoDNAS[56] corresponding to the canonical sequence of the human EGFR (UniProtKB ID: P00533). EGFR kinase is a domain with $N = 277$ residues plus the activation loop. Structures presenting more than 16 missing residues were not included in our dataset, otherwise missing residues have been included using the MODELLER software package with loop optimization.[57] Most of the missing residues were located in the activation loop and a loop between the β-sheets of the N lobe. Each model was evaluated using the DOPE scoring function to assert that each amino acid scored a negative energy value. Theoretical *B*-factors profiles have been calculated using NMA on each protein of the initial dataset and compared them with the experimental *B*-factor profiles. Then, we removed from the dataset all structures with optimal Spearman rank correlation coefficient between experimental and theoretical *B*-factors <0.6 Å. As a result of these filters in the original dataset, we were left with a final dataset composed of 41 EGFR kinase conformer's structures, 26 of which are classified as active conformers and the rest of 15 structures as inactive conformers. The complete set of EGFR kinase structures used in our final dataset is given in Table S1 (Supporting Information).

NMA background

Normal mode analysis has been performed using the elastic network model (ENM) that considers the protein as an elastic network of N α-carbons (nodes) linked by springs within a cut-off distance *rc*. For each protein in the dataset, the value of *rc* varies to obtain the best match between theoretical and experimental *B*-factors.

The interaction potential between residues is defined as.[41,58,59]

$$E(\mathbf{r}_i,\mathbf{r}_j)=\frac{1}{2}k_{ij}\left(|\mathbf{r}_{ij}|-\left|\mathbf{r}_{ij}^0\right|\right)^2 \quad (1)$$

being $\mathbf{r}_{ij}\equiv\mathbf{r}_i-\mathbf{r}_j$ the vector connecting residues i and j, and the zero superscript indicates the equilibrium position that corresponds to the coordinates of the α-carbons in the experimental structure. The value of the force constant k_{ij} is determined according to the type of interaction between nodes[60]:

if $|i-j|=1 \Rightarrow k_{ij}=\gamma$.
else
if $\left|\mathbf{r}_{ij}^0\right| \le r_c$ then
  if i and j are connected by disulphide bridge
  $\Rightarrow k_{ij}=\gamma$
    if i and j interact by hydrogen bond or salt bridge $\Rightarrow k_{ij}=\gamma$ x 0.1
    otherwise $\Rightarrow k_{ij}=\gamma$ x 0.01
if $\left|\mathbf{r}_{ij}^0\right| \ge r_c \Rightarrow k_{ij}=0$

being γ a scaling constant to match the theoretical result to experimental data. HBPlus program[61] has been used to obtain the connectivity information related to hydrogen bonds, salt bridges, and disulphide bridges.

Normal modes are obtained by solving the eigenvalue equation

$$\Lambda=\mathbf{q}^{\mathbf{T}}\mathbf{H}\mathbf{q} \quad (2)$$

being the Hessian matrix **H** a 3*N* x 3*N* matrix of second-order partial derivatives of the potential energy, **q** is an orthogonal 3*N* x 3*N* matrix whose columns $\mathbf{q}_k$ are the eigenvectors of **H**, that is, the normal modes, and **Λ** is the diagonal matrix of eigenvalues λ_k of **H**. The theoretical *B*-factor B_i of residue *i*, associated to its thermal fluctuation, is calculated as[62]

$$B_i=\frac{8\pi^2}{3}\left\langle\Delta\mathbf{r}_i^2\right\rangle \quad (3)$$

with $\left\langle\Delta\mathbf{r}_i^2\right\rangle=\left\langle\left(\mathbf{r}_i-\mathbf{r}_i^0\right)^2\right\rangle$ as the mean square displacement from its $\mathbf{r}_i^0$ equilibrium position. $\left\langle\Delta\mathbf{r}_i^2\right\rangle$ that can be expressed as the sum of contributions from the 3*N*-6 internal modes of motion $\{\mathbf{q}_k\}_{k\,=\,1,\,3N\,-\,6}$ as[63]

$$\left\langle\Delta\mathbf{r}_i^2\right\rangle=3k_BT\sum_{k=1}^{3N-6}\left[\lambda_k^{-1}\mathbf{q}_k\mathbf{q}_k^T\right]_{ii} \quad (4)$$

where k_B is the Boltzmann constant, *T* is the absolute temperature.

Comparison of weighted normal mode spaces

Vibrational motions associated to structural fluctuations on the main pocket of the active site of EGFR kinase structures have been compared through the calculation of the corresponding weighted Gramian matrix.[55,64,65] Being **q** and **q**′ the eigenvector 3*N*x(3*N* − 6) matrices (see NMA background subsection) associated to two conformers **A** and **B**, respectively, the vector projection of each $\mathbf{q}_j'$ onto the set of modes $\{\mathbf{q}_k\}_{k\,=\,1,\,3N\,-\,6}$ is defined as

$$\mathbf{p}_j^{\mathbf{AB}}=\sum_{k=1}^{3N-6}\left(\mathbf{q}_j'\cdot w_k\mathbf{q}_k\right)\mathbf{q}_k \quad (5)$$

where the weight w_k associated to normal mode $\mathbf{q}_k$ is defined as the normalized accumulation of contributions of $\mathbf{q}_k$ to *B*-factors B_i of each *i*th residue belongs to the main pocket of the active site of the EGFR kinase

$$w_k=\frac{\sum_{i\,\in\,\text{pocket}}\left[\lambda_k^{-1}\mathbf{q}_k\mathbf{q}_k^T\right]_{ii}}{\sum_{j=1}^{3N-6}\sum_{i\,\in\,\text{pocket}}\left[\lambda_j^{-1}\mathbf{q}_j\mathbf{q}_j^T\right]_{ii}} \quad (6)$$

The weighted Gramian matrix **G** (3*N* − 6 x 3*N* − 6) of the set of vectors $\left\{\mathbf{p}_j^{\mathbf{AB}}\right\}_{j=1,N}$ is calculated as the matrix of inner products with elements

$$G_{kl}=\left(w_k\mathbf{p}_k^{\mathbf{AB}}\cdot w_l\mathbf{p}_l^{\mathbf{AB}}\right) \quad (7)$$

The diagonalization of **G**

$$\mathbf{L}_{\mathbf{G}}^T\mathbf{G}\mathbf{L}_{\mathbf{G}}=\mathbf{\Omega}_{\boldsymbol{G}} \quad (8)$$

allows us to use the eigenvalues $\mathbf{\Omega}_{\boldsymbol{G}}$ of **G**, $\{\Omega_{G_k}\}_{k=1,3N-6}$, as a measure of the similarity between the two vibrational motions. Eigenvalues of **G** varies between 0 and 1.[64] The higher the value of Ω_{G_k}, the more the associated direction of motion, given by the corresponding $\mathbf{L}_{\mathbf{G}k}$, is common to both structures. Therefore, we define the similarity between the two weighted vibrational spaces can be defined as

$$\zeta^{\mathbf{AB}}=\frac{\sum_{k}^{3N-6}\Omega_k}{3N-6} \quad (9)$$

Moreover, an effective number of dimensions shared by two structures can be expressed as the Kirkpatrick index[66] rounded to the nearest higher integer

$$n_D=\sum_{k=1}^{3N-6}\frac{\Omega_k}{\Omega_1} \quad (10)$$

SVD representative vectors

The similarity among directions of fluctuations shared by different pairs of structures can be analyzed as follows. Matrices $\mathbf{A}^k$ of dimension 3 *N* x *L* are built with columns representing the $\mathbf{L}_{\mathbf{G}k}$ directions of each of the *L* pairs of structures compared as

described in comparison of weighted normal mode spaces subsection.

($\mathbf{L}_{Gk}$, pair # 1) ($\mathbf{L}_{Gk}$, pair # 2) …. ($\mathbf{L}_{Gk}$, pair # L)

$$A^k = \begin{Bmatrix} \text{residue\#1,x} \\ \text{residue\#1,y} \\ \text{residue\#1,z} \\ \text{residue\#2,x} \\ \\ \\ \text{residue\#N,z} \end{Bmatrix} \quad (11)$$

Singular value decomposition (SVD)[67] of each $\mathbf{A}^k$ matrix is performed. That is, each $\mathbf{A}^k$ is written as the product of an $3N \times L$ column-orthogonal matrix $\mathbf{U}^k$, an $L \times L$ diagonal matrix $\mathbf{W}^k$ with positive or zero elements (the singular values), and the transpose of an $L \times L$ orthogonal matrix $\mathbf{V}^k$:

$$(A^k) = (U^k) \cdot \begin{pmatrix} w_l^k & & \\ \dots & w_i^k & \dots \\ & & w_l^k \end{pmatrix} \cdot \left((V^k)^T \right) \quad (12)$$

Thus, the a_{ij}^k elements of matrix $\mathbf{A}^k$ can be expressed as the sum of products of columns of $\mathbf{U}^k$ and rows of $(\mathbf{V}^k)^T$, with the "weighting factors" being the singular values w_j^k.

$$a_{ij}^k = \sum_{z=1}^{l} w_z^k \cdot u_{iz}^k \cdot v_{jz}^k \quad (13)$$

Because of this, in the present work, the $\mathbf{U}_1^k$ vector with the highest w_1^k is considered the *representative mode* for the $\mathbf{L}_{Gk}$ *directions* of the matrix $\mathbf{A}^k$.

Results and Discussion

The human EGFR kinase domain is depicted in Figure 1. Following the definition given by Hasenahuer et al.[29] using structural an biological information, the main pocket of the active site is selected by manually inspection of the active conformers PDBids 1M14 (apo form) and 2GS6, and considering residues within a 5 Å radius from each atom of the ATP analog substrate-peptide conjugate in 2GS6.[68] Residues belonged to the pocket related to the active site are pointed out in Figure 2. The total number of residues in the pocket, N_{pocket}, is 53 (see Table S2 Supporting Information). Active and inactive conformers (Fig. 1) present distinguishable structural features concerning relative orientation of αC-helix, the N-lobe, and the active site, and changes in the main conformation of the activation loop. Nevertheless, the classification is long from been definitive since not all the currently available structures completely fulfill these requirements.

In the present work, common dynamical features among active conformers are explored to achieve unique fingerprints that enlighten us on minimum functional mechanisms within them. We first analyze differences in contributions of protein vibrations to thermal fluctuations of main pocket residues. Figure 3 shows the distributions of relative values of the weight w_k [see eq. 6)] associated to each normal mode $\mathbf{q}_k$. The lowest frequency normal mode $\mathbf{q}_1$ is the one that contributes the most to thermal fluctuations of main pocket residues. Therefore, w_k/w_1 represents the corresponding relative contribution of kth normal mode. Small values of w_k/w_1 indicate that the contribution of the kth normal mode to thermal fluctuations of main pocket residues can be neglected compared to the corresponding contribution of the first mode. Figure 3 shows that, in the case of active conformers, thermal fluctuations of main pocket residues are mainly covered by the first two lowest frequency normal modes. The contribution of the third normal mode results approximately 50% of the contribution of the first mode. On the contrary, in the case of inactive conformers, at least the six lowest frequency modes present larger contributions to pocket residue fluctuations. That is, fluctuations of the main pocket of active conformers are restrained to less number of low-frequency collective motions than inactive conformers.

Second, we investigate differences in the nature of low frequency normal modes. To do that, we have calculated the residue participation number of lowest frequency normal modes for EGFR kinase active and inactive conformers defined by

$$P_k = \left(\sum_{i=1}^{N} r_{ik}^4 \right)^{-1} \quad (14)$$

where $r_{ik}^2 = (q_{ik}^x)^2 + (q_{ik}^y)^2 + (q_{ik}^z)^2$, and $q_{ik}^j = (j = x,y,z)$ are the components of the ith atom in the kth normal mode. Values of $P_k \approx N$ correspond to vibrations equally distributed throughout all the residues of the protein, and $P_k \approx 1$ corresponds to vibrations involving the displacement of a single residue. In Figures 4a and 4b, we show the distribution of the fraction of residues involved in the motion of the two lowest frequency normal modes, calculated as values P_k/N, for active and inactive EGFR kinase conformers, respectively. We note first that, on

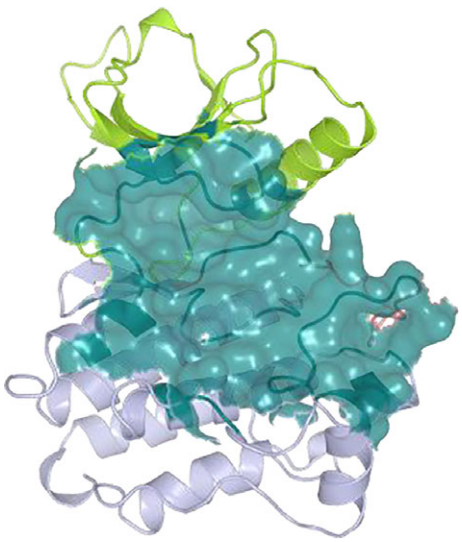

Figure 2. Human EGFR kinase domain. Residues belonged to the main pocket of the active site are pointed out (green) following the definition given by Hasenahuer et al.[29] [Color figure can be viewed at wileyonlinelibrary.com]

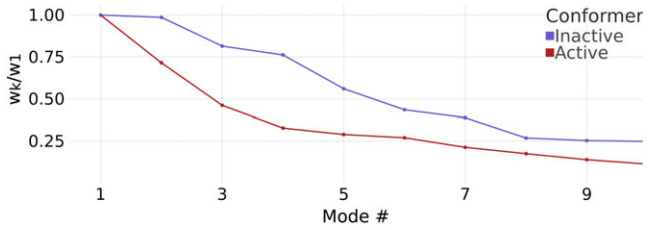

Figure 3. Distributions of relative values of normalized accumulation of contributions of modes $\mathbf{q}_k$ to *B*-factors B_i of residues belong to the main pocket of the active site of the EGFR kinase, that is, w_k/w_1 [eq. 6)]. [Color figure can be viewed at wileyonlinelibrary.com]

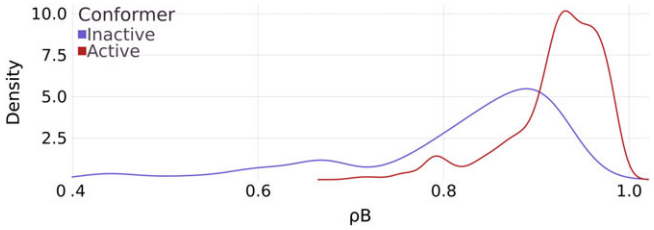

Figure 5. Histograms of backbone flexibility similarity ρ_B. The two histograms show the frequency distributions of ρ_B for active and a reference set of inactive conformers. [Color figure can be viewed at wileyonlinelibrary.com]

average, inactive conformers comprise a disparity of effects on the collective motion of residues involved in these vibrations, spreading the distributions over a larger range of values. On average, first normal modes of active conformers are more delocalized ($\langle P_1 \rangle_A = 0.21 \pm 0.03$) than first normal modes of inactive conformers ($\langle P_1 \rangle_I = 0.10 \pm 0.05$), involving almost the double of residues. The effects on second normal modes are more scattered (($\langle P_2 \rangle_A = 0.17 \pm 0.06$) and ($\langle P_2 \rangle_I = 0.21 \pm 0.10$)) without showing a clear tendency to disrupt the collectivity. That is, P_1 indicates a large impact of nonactivating mutations on the original collectivity of the lowest normal mode observed on active conformations. It is, therefore, expected that functional aspects of EGFR kinase involve coordinated motions between residues that are mainly reflected in the lowest frequency normal mode of active conformations.

Normal mode conservation has been shown to increase linearly with collectivity, so that the slowest most collective modes are the most conserved ones.[54] Therefore, our findings could point out toward a relative conservation of unique flexibility patterns among active conformers compared to inactive ones. To analyze this feature, we have calculated Pearson linear correlation ρ_B coefficients to quantify the similarity between C_α *B*-factor profiles of pairs of active and inactive conformers. In Figure 5, we compare the ρ_B distributions obtained for active and inactive conformers. Mean values of 0.92 and 0.81 are obtained for active and inactive conformers, respectively. The difference between both distributions is statistically validated by the Kolmogorov–Smirnov (KS) statistic value of 0.57. This result indicates that thermal fluctuations are more conserved between active conformers than between inactive conformers. Nevertheless, *B*-factors flexibility patterns do not result accurate enough to separate both types of conformers.

Vibrational motions associated to structural fluctuations of pairs of EGFR kinase conformers can be further compared through the calculation of the corresponding weighted Gramian **G** matrix (see Comparison of weighted normal mode spaces section). The diagonalization of **G** provides a set of new directions of motion $\{\mathbf{L_G}k\}_{k\,=\,1,\,N}$, given in decreasing order of their corresponding eigenvalues $\{\Omega_k\}_{k\,=\,1,\,3N\,-\,6}$. The values of Ω_k varies in a [0:1] range. On one hand, a value of $\Omega_k \approx 1$ indicates that the direction of motion defined by the corresponding vector $\mathbf{L_G}k$ is shared by both conformers. On the other hand, a value of $\Omega_k \approx 0$ indicates that the corresponding direction $\mathbf{L_G}k$ is specific to a single conformer. The value of $\zeta^{\mathbf{AB}}$, defined as the average $\{\Omega_k\}_{k\,=\,1,\,3N\,-\,6}$ values [see eq. 9)], provides our final similarity measure between weighted vibrational spaces of **A** and **B** conformers. Figure 6a displays the distribution of values of $\zeta^{\mathbf{AB}}$ obtained over all pairs of active and inactive conformations. The difference between both distributions is statistically validated by the KS statistic value of 0.97. This

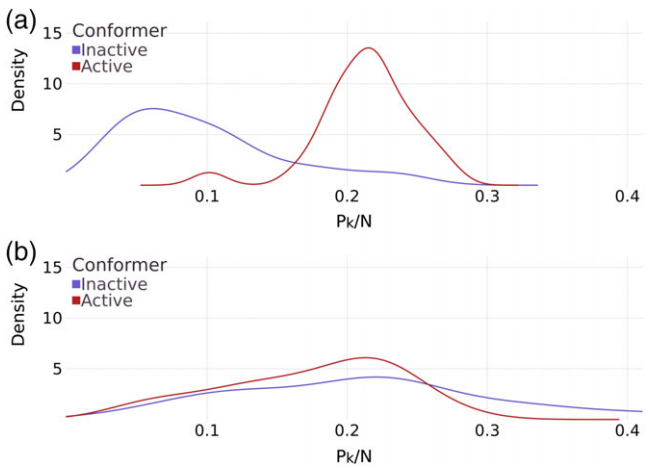

Figure 4. Distribution of the fraction of residues involved in the motion of the a) first and b) second lowest frequency normal modes, calculated as values P_k/N, for inactive and active EGFR kinase conformers. [Color figure can be viewed at wileyonlinelibrary.com]

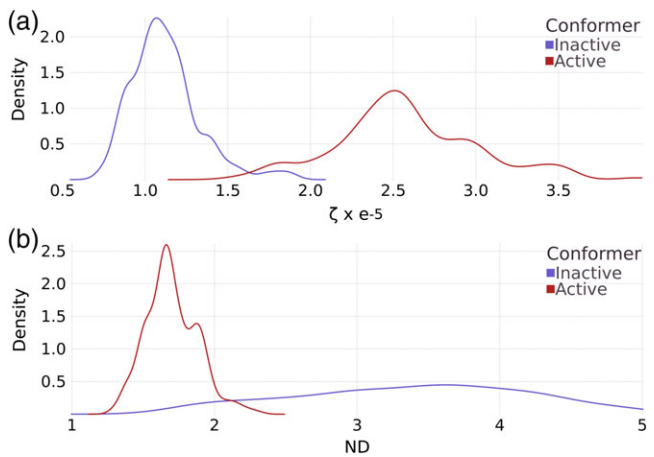

Figure 6. Histograms of values of a) the similarity meassure $\zeta^{\mathbf{AB}}$ between pairs of weighted vibrational spaces, and b) the effective number of dimensions n_D shared by pairs of mutants calculated over all pairs of active and inactive mutants. [Color figure can be viewed at wileyonlinelibrary.com]

means that ζ^{AB} separates active and inactive distributions better than Pearson linear correlation ρ_B coefficients between α-carbon *B*-factor profiles. That is, dynamics information concerning the directionality of collective motions allows a significant improvement in the distinction between active and inactive conformations respect to previous results using flexibility patterns (see Fig. 5).

Furthermore, the effective number of dimensions shared by pairs of conformers, given by the Kirkpatrick index n_D [eq. (10)], also allows a clear differentiation between active and inactive conformers. Figure 6b shows the distributions of values of n_D for both active and inactive conformers. An average value of 1.7 between pairs of active conformers indicates that dynamics similarities between them can be efficiently reduced to 1–2 common directions of motions. On the contrary, dynamics similarities between inactive conformers are scattered among 3–4 (average value of 3.4) different directions, less shared among pairs of conformers. The corresponding KS statistic value between both distributions is 0.89. That is, n_D results to be slightly less effective than ζ^{AB} to separate active and inactive dynamics.

The identity of the two lowest common directions $\mathbf{L_{G1}}$ and $\mathbf{L_{G2}}$ for pairs of active conformers can be explored by analyzing their projections on the basis of the original normal modes. The average contribution of the first lowest normal mode to $\mathbf{L_{G1}}$ is 0.73 ± 0.41 and the contribution of the second lowest normal modes to $\mathbf{L_{G2}}$ is 0.6 ± 0.41. Therefore, active conformers share one direction of motion mainly represented by the lowest frequency normal modes. This is in agreement with previous results obtained by Coveney and coworker.[36,37] using MD and PCA, that shows that the first PCA mode distinguish between active and inactive states. On the contrary, values of 0.59 ± 0.35 and 0.35 ± 0.30 are obtained for contributions of first and second normal modes to $\mathbf{L_{G1}}$ and $\mathbf{L_{G2}}$ of inactive conformers. That is, while the dynamics of pairs of active conformers reveal a common direction of motion that correspond to their natural lowest frequency of vibration, the dynamics of pairs of inactive conformers do not present unique patterns that can be assigned to individual original normal modes.

As we have shown, pairs of active conformers share common directions of motion represented by direction $\mathbf{L_{G1}}$. To obtain a fingerprint that characterizes it among the complete set of active conformers, further analysis is required. SVD, as a data compression technique, highlights the main common features of the original $\mathbf{L_{Gk}}$ directions within a few SVD representative modes U_i^k. Figure 7 depicts distributions of the overlap between SVD representative modes, U_1^k, and original $\mathbf{L_{Gk}}$ directions. As it can be seen, our complete set of active conformers share common directions represented by the corresponding SVD representative mode. In contrast, the corresponding U_1^1 in inactive conformers cannot be used as representative modes of the ensemble. Finally, directions $\mathbf{L_{G2}}$ and $\mathbf{L_{G3}}$ are different among either pairs of active or inactive conformers. In summary, all active conformers of our dataset share common dynamics that can ultimately associated to their lowest frequency mode. On the contrary, the dynamics of inactive conformers result heterogeneous and no unique dynamics patterns that include all of them can be found.

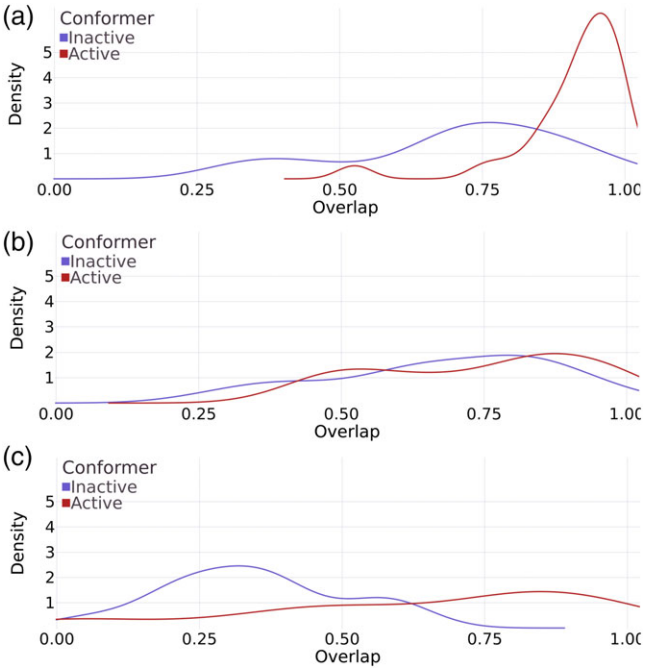

Figure 7. Histograms of the overlap between the first SVD representative modes, U_1^k, and original $\mathbf{L_{Gk}}$ directions for a) $U_1^1 \cdot \mathbf{L_{G1}}$; b) $U_1^2 \cdot \mathbf{L_{G2}}$; c) $U_1^3 \cdot \mathbf{L_{G3}}$ for both active (red) and inactive (blue) mutants. [Color figure can be viewed at wileyonlinelibrary.com]

Figure 8 shows that fingerprint mode U_1^1, shared by all active conformers, and also the less representative mode U_1^2. Both modes describe relative displacements between the two lobes (N and C lobes) of the kinase domain. Particularly, motions of N lobe, and αG helix. Local rearrangements of the αG helix have been previously related to the conformational change of c-Src kinase associated to ligand binding.[69] Besides, a relative suppression of active loop deformations also characterizes modes U_1^1 and U_1^2 respect to motions of the corresponding modes of inactive conformations, in agreement with previous MD results that reveal a relatively more rigid alpha-C helix and active loop in the active than inactive states of wild-type and active EGFR mutants.[36–38] For the sake of comparison, we also show shows the corresponding modes for inactive conformers.

To evaluate effects of modes U_1^1 and U_1^2 on the active conformation of EGFR kinase domain we calculate the corresponding

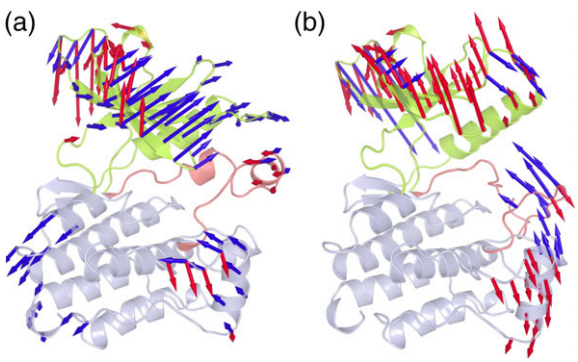

Figure 8. SVD representative modes U_1^1 (red) and U_1^2(blue) of a) inactive and b) active conformers. [Color figure can be viewed at wileyonlinelibrary.com]

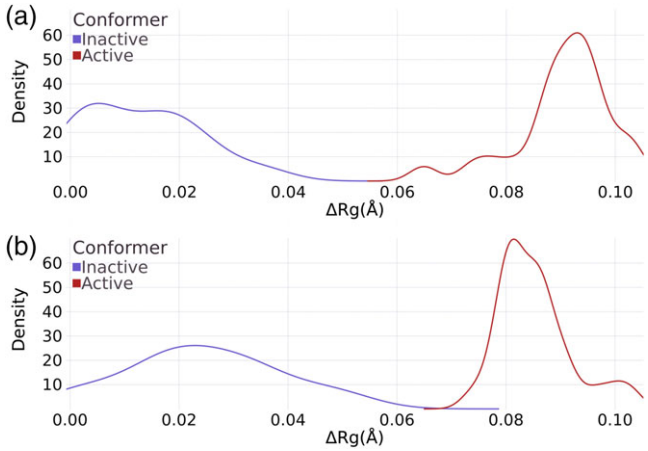

Figure 9. Histograms of changes of the radius of gyration, ΔRg, of the EGFR kinase domain due to displacements in the direction of SVD representative modes a) U_1^1 and b) U_1^2 for both inactive and active conformers. [Color figure can be viewed at wileyonlinelibrary.com]

changes of radius of gyration (ΔRg). Figure 9 shows the distribution of values of ΔRg due to displacements in the direction of these modes. As it can be seen, their effect on active conformers significantly differs from displacements of the corresponding modes of inactive conformers. Modes U_1^1 and U_1^2 of active conformers lead to more extended conformations involving a hinge-like separation of the N and C lobes (see also Supporting Information).

The motion in the direction of U_1^1 and U_1^2 modes in active conformers is consistent with previously reported motions[18,35] analyzed using molecular dynamics simulations, that describe the active–inactive conformational transition. Shan et al.[35] reported an opening of the two lobes to allow local unfolding at the hinge region, prior to the close of the lobes to re-stabilize in its inactive conformation. Besides, Coveney et al.[35,36] reported hinge and shear motions between the C- and N-lobe associated to first and second PCA modes. Our analysis suggests that these findings are actually a hallmark of active conformers dynamics when it comes to differentiate them from the inactive conformers.

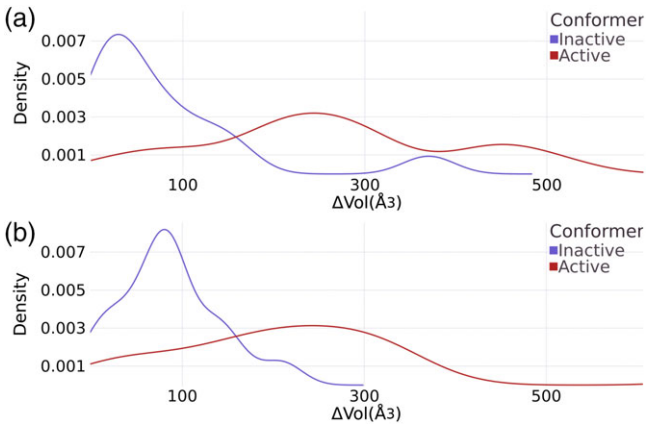

Figure 10. Histograms of differences in cavity volumes (ΔVol) between structures of active and inactive conformers previously displaced a magnitude –Ai and + Ai, being Ai(Å) the corresponding amplitude of the *i*th mode at room temperature. a) $i = 1$, and b) $i = 2$. [Color figure can be viewed at wileyonlinelibrary.com]

We finally explore functional aspects of motions in the direction of U_1^1 and U_1^2 modes. For this purpose, we calculate changes in the cavity volume of the main pocket of the active site due to structural displacements in the direction of these modes. Cavity volumes are calculated using our previously developed method,[70] particularly suited to measure changes in cavity volumes due to small atomic coordinate displacements in the direction of specific normal modes. Figure 10 displays the distribution of differences in cavity volumes between structures of active and inactive conformers previously displaced a magnitude $-A_i$ and $+A_i$, being A_i(Å) the corresponding amplitude of the *i*th mode at room temperature. That is, $A_i = ((2k_BT)/\lambda_i)^{1/2}$ where k_B is the Boltzmann constant and T is the absolute temperature (300 K). λ_i corresponds to the eigenvalue associated to the *i*th normal mode, as defined in NMA background section, scaled to best fit the theoretical residue fluctuations with the corresponding experimental temperature factors. As can be seen, motions in the direction of the two lowest frequency normal modes of active conformers lead to conformational changes that involve larger changes in cavity volumes than motions in the direction of the corresponding modes of inactive conformers. Changes in cavity volumes could have a subsequent impact on ligand affinity and, therefore, regulation of the biological function of the active site pocket of the EGFR kinase a protein. Herein, we have identified common dynamics shared only by the active conformers of the EGFR kinase. Since these dynamics are associated to significant changes in the volume of the main pocket of the active site of EGFR kinase, their functional impacts should be expected. Previous works have shown that inhibitors can act by hindering and/or change the direction of specific protein motions.[71] Furthermore, the identification of dynamically important residues associated to these motions can be performed using previously developed methods.[55,72,73] Therefore, strategies for the development of new inhibitors that hinder or modify our dynamics fingerprints can be faced. Particularly, compounds that disrupt relative displacements between the two lobes (N and C lobes), focusing on motions of the N lobe, and αG helix. Furthermore, inhibitors particularly designed to interact with the gly-rich loop could strongly modify these concerted motions.

Conclusions

We have performed an extensive comparative analysis of global dynamics features shared by existing conformers in EGFR kinase. We have identified clear directions of motions that can be used as fingerprints to differentiate active and inactive conformers. A novel procedure has been applied that can be used for sets of conformers in other proteins. The method allows the comparison of patterns of vibrations rescuing common representative directions of motions shared among the ensemble.

Two representative directions of motions have been identified and characterized among active conformations in EGFR kinase. These motions represent fingerprints of active conformers that can be added to previously reported structural

features. Their conservation among the complete set of active conformers enlighten us on minima functional mechanisms within them. Inactive conformers have shown a general tendency to disrupt the collective motion of residues involved in these vibrations. It is, therefore, expected that functional aspects of EGFR kinase involve coordinated motions between residues that are mainly reflected in the two lowest frequency normal modes of active conformers. These modes allow active conformers to reach more extended conformations involving a hinge-like separation of the N and C lobes, a prior requirement to achieve a relatively more stable inactive conformation. Besides, they lead to larger changes of cavity volumes compare to the corresponding modes of inactive conformers. This differential dynamics between active and inactive conformers can impact on ligand affinity and, therefore, regulation of the biological function of the active site pocket of the EGFR kinase region. The importance of protein dynamics to explain biological function has been extensively recognized using the ensemble of conformations in dynamic equilibrium to represent the native state of the protein.[74] Protein studies considering dynamics are important and central to improve our comprehension of protein function in normal as well in disease-related states.

The importance of protein dynamics to explain biological function has been extensively recognized using the ensemble of conformations in dynamic equilibrium to represent the native state of the protein.[74] Protein studies considering this dynamics are important and central to improve our comprehension of protein function in normal as well in disease-related states. Moreover, it is important in the study and design of inhibitors, given that their selectivity and specificity depend on protein sequence, structural, or conformational differences as well on protein motions being a challenge the recognition of specific mechanisms to modulate protein activity.

Keywords: normal modes · protein conformers · Epidermal growth factor receptor · dynamics fingerprints

How to cite this article: G. P. Barletta, M. A. Hasenhuer, M. S. Fornasari, G. Parisi, S. Fernandez-Alberti. *J. Comput. Chem.* **2018**, 00, 1–9. DOI: 10.1002/jcc.25590

Additional Supporting Information may be found in the online version of this article.

[1] Y. Yarden, M. Sliwkowski, Nat. Rev. Mol. Cell Biol. **2001**, 2, 127.
[2] C. Arteaga, J. Engelman, Cancer Cell **2014**, 25, 282.
[3] S. Forbes, D. Beare, E. Al, Nucleic Acids Res. **2015**, 43, D805.
[4] H. Cheng, E. Al, J. Med. Chem. **2016**, 59, 2005.
[5] G. Lurje, H. Lenz, Oncology **2009**, 77, 400.
[6] S. Müller, E. Al, Nat. Chem. Biol. **2015**, 11, 818.
[7] Q. Wang, J. Zorn, J. Kuriyan, Methods Enzymol. **2014**, 548, 23.
[8] S. Kumar, B. Ma, C. Tsai, N. Sinha, R. Nussinov, Protein Sci. **2000**, 9, 10.
[9] S. Osorio, R. Alba, Z. Nikoloski, A. Kochevenko, A. Fernie, J. Giovannoni, Plant Physiol. **2012**, 159, 1713.
[10] L. James, D. S. Tawfik, Trends Biochem. Sci. **2003**, 28, 361.
[11] K. Fergurson, Biochem. Soc. Trans. **2004**, 32, 742.
[12] A. Kornev, S. Taylor, Biochim. Biophys. Acta **2010**, 1804, 440.
[13] X. Zhang, K. Pickin, R. Bose, N. Jura, P. Cole, J. Kuriyan, Nature **2007**, 450, 741.
[14] C. Yun, T. Boggon, Y. Li, M. Woo, H. Greulich, M. Meyerson, M. Eck, Cancer Cell **2007**, 11, 217.
[15] N. Jura, N. F. Endres, K. Engel, S. Deindl, R. Das, M. H. Lamers, D. E. Wemmer, X. Zhang, J. Kuriyan, Cell **2009**, 137, 1293.
[16] N. Endres, R. Das, A. Smith, A. Arkhipov, E. Kovacs, Y. Huang, J. Pelton, Y. Shan, D. Shaw, D. Wemmer, J. Groves, J. Kuriyan, Cell **2013**, 152, 543.
[17] M. R. Brewer, S. Choi, D. Alvarado, K. Moravcevic, A. Pozzi, M. Lemmon, G. Carpenter, Mol. Cell **2009**, 34, 641.
[18] Y. Shan, M. P. Eastwood, X. Zhang, E. T. Kim, A. Arkhipov, R. O. Dror, J. Jumper, J. Kuriyan, D. E. Shaw, Cell **2012**, 149, 860.
[19] M. Kalia, K. Madhy, Metabolism **2015**, 64, S16.
[20] N. Jura, E. Al, Mol. Cell **2011**, 42, 9.
[21] A. Dixit, G. Verkhivker, Comput. Math. Methods Med. **2014**, 2014, 653487.
[22] D. Knighton, E. Al, Science **1991**, 253, 407.
[23] W. Hemmer, E. Al, J. Biol. Chem. **1997**, 272, 16946.
[24] S. Taylor, A. Kornev, Trends Biochem. Sci. **2011**, 36, 65.
[25] K. James, G. Verkhivker, PLoS One **2014**, 9, e113488.
[26] J. Hu, E. Al, Mol. Cell. Biol. **2015**, 35, 264.
[27] K. Gajiwala, E. Al, Structure **2013**, 21, 209.
[28] D. Fabbro, Mol. Pharmacol. **2015**, 87, 766.
[29] M. A. Hasenahuer, G. P. Barletta, S. Fernandez-Alberti, G. Parisi, M. S. Fornasari, PLoS One **2017**, 12, e0189147.
[30] S. Kumar, R. Nussinov, J. Mol. Biol. **1999**, 293, 1241.
[31] A. Tramontano, T. Anna, FEBS Lett. **2006**, 580, 2928.
[32] H. Du, J. R. Brender, J. Zhang, Y. Zhang, Methods **2015**, 71, 77.
[33] M. Hasenahuer, G. Parisi, M. Gautier, A. Lazarowski, G. Bramuglia, M. Fornasari, Ann. Hum. Genet. **2015**, 79, 385.
[34] K. Henzler-Wildman, D. Kern, Nature **2007**, 450, 964.
[35] Y. Shan, A. Arkhipov, E. T. Kim, A. C. Pan, D. E. Shaw, Proc. Natl. Acad. Sci. USA **2013**, 110, 7270.
[36] S. Wan, P. V. Coveney, J. Comput. Chem. **2011**, 32, 2843.
[37] S. Wan, D. W. Wright, P. V. Coveney, Mol. Cancer Ther. **2012**, 11, 2394.
[38] L. Sutto, F. L. Gervasio, Proc. Natl. Acad. Sci. USA, Vol. 110, **2013**, p. 10616.
[39] A. Dixit, G. M. Verkhivker, PLoS Comput. Biol. **2009**, 5, e1000487.
[40] S. Lovera, M. Morando, E. Pucheta-Martinez, J. L. Martinez-Torrecuadrada, G. Saladino, F. L. Gervasio, PLoS Comput. Biol. **2015**, 11, e1004578.
[41] M. M. Tirion, Phys. Rev. Lett. **1996**, 77, 1905.
[42] I. Bahar, B. Erman, R. L. Jernigan, A. R. Atilgan, D. G. Covell, J. Mol. Biol. **1999**, 285, 1023.
[43] I. Bahar, R. L. Jernigan, Biochemistry **1999**, 38, 3478.
[44] K. Hinsen, R. G. Kneller, J. Chem. Phys. **1999**, 24, 10766.
[45] A. Emperador, O. Carrillo, M. Rueda, M. Orozco, Biophys. J. **2008**, 95, 2127.
[46] M. Levitt, C. Sander, P. S. Stern, J. Mol. Biol. **1985**, 181, 423.
[47] O. Marques, Y. H. Sanejouand, Proteins **1995**, 23, 557.
[48] C. Xu, D. Tobi, I. Bahar, J. Mol. Biol. **2003**, 333, 153.
[49] L. Yang, G. Song, R. L. Jernigan, Biophys. J. **2007**, 93, 920.
[50] O. Keskin, R. L. Jernigan, I. Bahar, **2000**, 78, 2093–2106.
[51] A. Leo-Macias, P. Lopez-Romero, D. Lupyan, D. Zerbino, A. R. Ortiz, Biophys. J. **2005**, 88, 1291.
[52] W. Zheng, B. R. Brooks, D. Thirumalai, Proc. Natl. Acad. Sci. USA **2006**, 103, 7664.
[53] S. Maguid, S. Fernandez-Alberti, L. Ferrelli, J. Echave, Biophys. J. **2005**, 89, 3.
[54] S. Maguid, S. Fernandez-Alberti, J. Echave, Gene **2008**, 422, 7.
[55] T. Saldaño, A. Monzon, G. Parisi, S. Fernandez-Alberti, PLoS Comput. Biol. **2016**, 12, e1004775.
[56] A. M. Monzon, E. Juritz, M. S. Fornasari, G. Parisi, Bioinformatics **2013**, 29, 2512.
[57] A. Sali, T. Blundell, J. Mol. Biol. **1993**, 234, 779.
[58] K. Hinsen, Proteins **1998**, 33, 417.
[59] A. Atilgan, S. Durell, R. Jernigan, M. Demirel, O. Keskin, I. Bahar, Biophys. J. **2001**, 80, 505.
[60] J. I. Jeong, Y. Jang, M. K. Kim, J. Mol. Graph. Model. **2006**, 24, 296.
[61] I. McDonald, J. Thornton, J. Mol. Biol. **1994**, 238, 777.
[62] I. Bahar, A. R. Atilgan, B. Erman, Fold. Des. **1997**, 2, 173.
[63] F. Tama, F. X. Gadea, O. Marques, Y. H. Sanejouand, Proteins **2000**, 41, 1.

[64] W. J. Krzanowski, J. Am. Stat. Assoc. **1979**, 74, 703.
[65] M. Grosso, A. Kalstein, G. Parisi, A. Roitberg, S. Fernandez-Alberti, M. Grosso, A. Kalstein, G. Parisi, A. E. Roitberg, J. Chem. Phys. **2015**, 142, 245101.
[66] M. Kirkpatrick, Genetica **2009**, 136, 271.
[67] M. E. Wall, A. Rechtsteinner, L. M. Rocha, Singular Value Decomposition and Principal Component Analysis. In ; D. P. Berrar, W. Dubitzky, M. Granzow, Eds., Kluwer, Norrwell, MA, p. 2003.
[68] X. Zhang, J. Gureasko, K. Shen, P. A. Cole, J. Kuriyan, Cell **2006**, 125, 1137.
[69] M. A. Morando, G. Saladino,N. D'Amelio, E. Pucheta-Martinez, S. Lovera, M. Lelli, B. López-Méndez, M. Marenchino, R. Campos-Olivas, F. L. Gervasio, Sci. Rep. **2016**, 6, 1–9.
[70] G. P. Barletta, S. Fernandez-Alberti, J. Chem. Theory Comput. **2018**, 14, 998.
[71] N. A. Temiz, I. Bahar, Proteins **2002**, 49, 61.
[72] W. Zheng, B. R. Brooks, S. Doniach, D. Thirumalai, **2005**, 13, 565–577.
[73] T. E. Saldaño, G. Zanotti, G. Parisi, S. Fernandez-Alberti, PLoS One **2017**, 12, e0181019.
[74] G. Wei, W. Xi, R. Nussinov, B. Ma, Chem. Rev. **2016**, 116, 6516.

Received: 15 June 2018
Revised: 6 August 2018
Accepted: 19 August 2018