Published in final edited form as:
Phys Chem Chem Phys. 2014 April 14; 16(14): 6508–6518. doi:10.1039/c4cp00110a.

NMR Mapping of Protein Conformational Landscapes using Coordinated Behavior of Chemical Shifts upon Ligand Binding

Alessandro Cembran[1,2], **Jonggul Kim**[2], **Jiali Gao**[2], and **Gianluigi Veglia**[1,2,*]

[1]Department of Biochemistry, Biophysics & Molecular Biology

[2]Department of Chemistry – University of Minnesota 55455.

Abstract

Proteins exist as an ensemble of conformers that are distributed on free energy landscapes resembling folding funnels. While the most stable conformers populate low energy basins, protein function is often carried out through low-populated conformational states that occupy high energy basins. Ligand binding shifts the populations of these states, changing the distribution of these conformers. Understanding how the equilibrium among the states is altered upon ligand binding, interaction with other binding partners, and/or mutations and post-translational modifications is of critical importance for explaining allosteric signaling in proteins. Here, we propose a statistical analysis of the chemical shifts (CONCISE, COordiNated ChemIcal Shifts bEhavior) for the interpretation of protein conformational equilibria following linear trajectories of NMR chemical shifts. CONCISE enables one to quantitatively measure the population shifts associated with ligand titrations and estimate the degree of collectiveness of the protein residues' response to ligand binding, giving a concise view of the structural transitions. The combination of CONCISE with thermocalorimetric and kinetic data allows one to depict a protein's approximate conformational energy landscape. We tested this method with the catalytic subunit of cAMP-dependent protein kinase A, a ubiquitous enzyme that undergoes conformational transitions upon both nucleotide and pseudo-substrate binding. When complemented with chemical shift covariance analysis (CHESCA), this new method offers both collective response and residue-specific correlations for ligand binding to proteins.

Keywords

Protein Kinase A; Allostery; NMR relaxation; SAR by NMR

INTRODUCTION

To perform their functions, proteins undergo conformational transitions among different states that are distributed in funnel-like free energy landscapes similar to the classical folding funnels[1, 2]. Allosteric effectors, such as co-factors, drugs, and other binding partners, modulate the conformational equilibrium by shifting the population of these conformers among different energy basins [3-5]. Understanding how the equilibrium between these states

[*]**Corresponding Author** Department of Biochemistry, Biophysics, and Molecular Biology, University of Minnesota, 6-155 Jackson Hall, MN 55455. Telephone: (612) 625-0758. Fax: (612) 625-2163. vegli001@umn.edu. .

is altered upon ligand binding is of critical importance for elucidating allosteric signaling and regulation of proteins [6].

NMR spectroscopy is emerging as the technique of choice to map allosteric phenomena at the atomic level [7-9], unveiling mechanisms underlying allostery[10, 11]. Heteronuclear correlation experiments [12] provide NMR amide "fingerprints" of proteins and protein complexes. For amide groups, the [^{1}H,^{15}N]-HSQC experiment correlates the ^{1}H and ^{15}N frequencies (chemical shifts); amide chemical shifts are sensitive probes for structural changes and are used to assess folded state and monitor the effects of ligand binding on protein structures[13-16]. Resonance chemical shifts reflect the weighted average of different conformer populations within the sample and are affected by the exchange regime between the different conformational states. Under a slow exchange regime in the NMR time scale ($k_{ex} << \Delta\omega$), the amide resonances display distinct peaks. Under intermediate exchange ($k_{ex} \sim \Delta\omega$), the resonances broaden out, while for a fast exchange regime ($k_{ex} >> \Delta\omega$), the resonances display narrow lines, reflecting the weighted average of the populations present in the sample. For ligand binding studies, it is customary to carry out titration experiments. If a protein exists in fast equilibrium between different states, and ligand binding shifts the equilibrium toward a single state, the trends of the chemical shifts upon ligand titration follow linear paths that reflect ligand dissociation constants (reviewed in [17]). Similarly, if a protein exists in equilibrium between different states in a fast exchange regime and one were to promote the population of other states by mutations or posttranslational modifications, the trajectories of the chemical shifts follow linear trends [17-24].

These linear trends can be used to quantitate both the stability and functional states of proteins [25]. Amide chemical shift linear trends provide the basis for a quantitative method recently proposed by Melacini's group to correlate long-range chemical shift changes to allosteric regulation, as well as to differentiate between active and inactive states of the regulatory subunit of cAMP dependent Protein Kinase A [26-29]. More recently, the Forman-Kay group used a similar approach to analyze the allosteric coupling in the cystic fibrosis transmembrane conductance regulator [30] and quantify the differential engagement of peptide complexes [31]. Finally, chemical shift trajectories have been utilized to determine the affinities and the number of binding sites in protein ligand interactions [32].

Inspired by this body of work, we developed a complementary analysis method (CONCISE, COordiNated ChemIcal Shifts bEhavior) that estimates the density of the states of a protein in different bound forms, providing a degree of collective response, or cooperativity, of the protein residues upon ligand binding. The method can be generally employed to investigate a variety of perturbations including mutations, post-translational modifications, and protein-protein interactions. Moreover, such an approach may be very useful in cases where chemical shifts are the only means to gain insight into protein function, such as in molten globule proteins[33]. When combined with thermocalorimetric data, the density of populations obtained by this analysis defines a free energy landscape of the protein's ligated states. We tested this method for the conformational transitions of the C-subunit of the cAMP-dependent protein kinase A (PKA-C) upon nucleotide and pseudo-substrate binding, and constructed the free energy landscape along the enzymatic reaction coordinates.

PKA-C is a ubiquitous enzyme involved in many signaling pathways, playing a fundamental role in the pathophysiology of several different cellular events. PKA-C has a bilobal fold, with a small lobe (N-lobe), comprising mostly β-sheets that harbor the nucleotide binding pocket, and a large lobe (C-lobe), including mostly helical segments that host the substrate binding groove[34, 35]. During turnover, PKA-C is thought to interconvert between three major conformational states: *open* (apo), *intermediate* (nucleotide-bound), and *closed* (nucleotide and substrate bound)[34, 35]. While the *apo* form explores mainly an open conformation, with the two lobes disengaged and conformational dynamics uncommitted to catalysis, nucleotide binding shifts the enzyme conformational ensemble toward a dynamically committed state that is able to bind the substrate with enhanced affinity (positive cooperativity)[22-24]. Finally, substrate binding shifts the equilibrium toward a new basin, where the conformational dynamics are redistributed throughout the entire protein, priming it for phosphoryl transfer[22]. Binding of an inhibitor peptide (PKI_{5-25}), however, traps the kinase in a fully closed state, quenching the conformational dynamics throughout the enzyme and restricting the opening and closing motions required for product release[22]. To map these equilibria, we utilized previously measured chemical shifts of the PKA-C fingerprints of the apo, ATP-γ-N (AMP-PNP)-bound, ATP-bound, and AMP-PNP/PKI-bound forms, and applied both CONCISE and the chemical shifts covariance analysis (CHESCA)[26]. We found that our method is able to distinguish between the multiple states and quantitate the density of each state, and, when combined with binding free energy data, define the free energy landscape of the kinase. While site-specific correlations among the residues are attainable through the CHESCA analysis[26], CONCISE is able to estimate the collective response of the enzyme.

METHODS

Calculation of the Population Density

In order to validate our method, we generated synthetic [^{1}H,^{15}N]-HSQC spectra corresponding to four virtual states (A, B, C, and D) similar to those obtained experimentally for PKA-C, with an identical number of peaks, positions, and range of chemical shift perturbations. Then, we simulated four different cases. In the first case (*ideal*, Figure 1A), we placed the chemical shifts of the resonances in four virtual states according to perfect linear trajectories, with equilibrium positions: 0 (state A), 1/3 (state B), 2/3 (state C), and 1 (state D), going from A to the D state. For the second case (*ideal+noise*, Figure 1B), we added random noise to both the nitrogen (standard deviation, SD = 0.20 ppm) and proton (SD = 0.03 ppm) dimensions of the ideal spectrum in Figure 1A. In the third case (*random*, Figure 1C), the positions of the resonances were completely randomized, and in the fourth case (*mixed random+ideal*, Figure 1D), we set half of the peaks following linear trajectories with added random noise (with nitrogen and proton SD of 0.10 and 0.02 ppm, respectively), while the other half of the peaks represented a random distribution of the states, as in Figure 1C. In order to analyze the chemical shift trajectories, we utilized principal component analysis (PCA). This approach is similar to the one proposed by Sakurai and Goto [25, 36] and allows one to filter out nearest-neighbor effects and perturbations that do not reflect the shifts in the conformational equilibrium upon ligand binding (see also Melacini [27] and Forman-Kay[31]). We model the conformational

equilibrium by a two-state conformational exchange, and the four states can be described as population weighted states. In Figure 2, we report a typical example of the PCA analysis filtering procedure for two selected resonances in the *ideal+noise* case (Figure 1B). To give equal weight to both proton and nitrogen chemical shift perturbations, ^{15}N chemical shifts were scaled by a factor of 0.154. PCA is applied separately to each residue, and rotates the residue's resonances from the ^{1}H, ^{15}N plane to a new coordinate system, defined by PC1 and PC2, which is a linear combination of the original axes. The new axes (PC1/PC2) maximize the variance along PC1 and minimize it along PC2. As a result, a perfectly linear correlation would perfectly align along PC1 and show no spread along PC2, and thus we consider the PC1 projection as a measure of the equilibrium position as reported by each individual residue. The ratio between the SD along PC1 and that of PC2 defines the degree of linearity for each residue, which is used to filter out those resonances whose chemical shift changes do not report on the conformational equilibrium or for which the nearest-neighbor effects trump the sensitivity to the conformational equilibrium.

After linearizing the trajectories for each residue, the per-residue information is averaged together into a global descriptor – the average PC score – that reports on the position of each state along the equilibrium. To average the PC1 projections, all PCs must be oriented in the same direction. Since the PCA analysis orients the trajectories of the peaks either toward the positive, or rotated by 180° toward negative values (Figures 2B and 2E), we imposed a common orientation for all of the trajectories, applying a 180° rotation for those trajectories in the opposite direction (Figures 2E and 2F). The two states that define the directionality correspond to the extremes (initial, A, and final, D, states) of the conformational equilibrium. The scores of different residues along PC1 are then normalized, dividing by the standard deviations of PC1 (see scale on top of Figures 2B, 2C, 2E, and 2F). Once oriented and normalized, the PC1 projections of the selected subset of residues are averaged and the standard deviations computed (Figure 3). The average PC1 score indicates the position of each state along the equilibrium, while the spread reported by the standard deviation bars is a measure of the collective response of the selected subset of residues to ligand binding. In this context, narrow distributions are indicative of a collective behavior of the protein residues reaching the ligated state, while broad distributions indicate the presence of uncoordinated behavior. To filter out resonances not reporting on conformational equilibrium, the average is performed on a reduced set of residues with the highest degree of linearity. We found that a SD_{PC1}/SD_{PC2} ratio ≥ 3 gives a reasonable threshold for linearity, while at the same time ensuring a sufficient number of residues for a statistically significant sampling. The experimental error was taken into account by discarding all of the residues with a PC1 range below 0.05 ppm, which is analogous to the range used in the projection analysis method [26].

Covariance Analysis (CHESCA)

The CHESCA analysis of the chemical shift covariance was carried out in agreement with the protocol proposed by Melacini and co-workers [26]. While the original CHESCA method uses the ^{15}N and ^{1}H chemical shifts weighted sum ($\delta = w^N\delta^N + w^H\delta^H$, with $w^N = 0.2$ and $w^H = 1$) to calculate the covariance matrix, we used instead the PC1 scores, which more directly report on the conformational equilibrium.. While our approach gives a concise view

of the conformational transitions, CHESCA gives a residue-specific view of the correlations among the residues in different protein domains [26]. Therefore, we calculated the correlation coefficients between the PC1 scores of the different residues using the following relationship:

$$c_{ij}=\frac{\langle (S_i-\langle S_i\rangle)(S_j-\langle S_j\rangle)\rangle}{\sqrt{\left(\langle S_i^2\rangle-\langle S_i\rangle^2\right)\left(\left\langle S_j^2\right\rangle-\langle S_j\rangle^2\right)}}$$

where *s* indicates the normalized PC1 scores for residues *i* and *j*, and the brackets indicate the mean over all of the states. The correlation coefficients between residues are plotted in a matrix format to identify the residues with highly correlated chemical shift changes. Furthermore, we clustered the residues using a hierarchical clustering step as described by Selvaratnam *et al.* [26]. The largest cluster of correlated residues was identified using a correlation coefficient of 0.99 as threshold for the clustering trees.

Experimental HSQC titration data of PKA-C

The HSQC data used in our analysis were taken from the previous published work by Masterson *et al.* [22], and all the samples were obtained under the same buffer, pH, and temperature conditions. A crucial element for the successful application of this method is the correct referencing of the resonances in the [^{1}H,^{15}N]-HSQC spectra, since the chemical shifts are highly sensitive, even to small structural perturbations. Therefore, the corrections to the chemical shifts ($\Delta\delta^{15}N^j$, $\Delta\delta^1H^j$) for a state *j* were calculated minimizing a distance function, D^j, defined as:

$$D^j=\sum_{i=1}^{Nres}\left|\left(\delta^{15}N_i^j+\Delta\delta^{15}N^j-\delta^{15}N_i^{ref}\right)\right|+\left|\left(\delta^1H_i^j+\Delta\delta^1H^j-\delta^1H_i^{ref}\right)\right|$$

where the superscript *ref* indicates the reference state (in our case the apo form), the subscript *i* runs over all of the *N* residues, and *j* is the state to be re-referenced.

RESULTS

CONCISE analysis of synthetic data

The CONCISE calculations carried out on the synthetic data are shown in Figure 3, where the average position and spread of the PC1 scores for each state are displayed as bar plots and normal distributions. For the ideal case illustrated in the first panel of Figure 3A-B (perfect linear correlation), the equilibrium values for the four states were −1.34, −0.45, 0.44, and 1.35 standard units, which correspond to the equilibrium positions of 0, 1/3, 2/3, and 1. Since no random noise was added to the spectrum, there is no spread in the equilibrium positions. In contrast, the introduction of random noise (Figure 3A-B, second panel) causes a slight shift of the average PC score, broadening the distribution by approximately one standard deviation. When a random distribution of the chemical shifts is used (Figure 3A-B, third panel), the B and C states are indistinguishable (SD −0.04 and

−0.01), while states A and D are still distinguishable (SD −0.63 and 0.68). The latter originates from the rotation of the principal components, which requires a specific peak order and the assignment of the two reference states. Therefore, if the average values of the two reference states differ by ~1.3 standard deviations and the broadening is greater than 2 standard deviations, the analysis cannot be applied. Finally, the fourth panel in Figure 3A-B illustrates the results obtained for the mixed case, where the first half of the peaks (1-175) are ordered along the linear trajectory, while the second half (176-350) are randomly distributed. As expected, the random data introduce a broadening of the distributions, which can be eliminated using the filtering method described above (Figure 3C-D). While the filtering improves the accuracy of the average position and reduces the broadening of the distribution reported in the fourth panel of Figure 3, it produces no effect in the case of random chemical shift distribution reported in the third panel of Figure 3.

CHESCA analysis of synthetic data

Figure 4 shows the calculated correlation matrices of PC1 scores for the synthetic data, with a threshold for displaying correlation coefficients higher than 0.9. For the ideal case, the changes in chemical shifts are all strongly correlated, since the residues follow perfect linear trajectories. Upon addition of random noise (Figure 4B), the values of the correlation coefficients throughout the protein sequence lower substantially. Nonetheless, it is still possible to observe that several residues throughout the proteins are highly correlated. For the random case (Figure 4C), the degree of correlations among the different residues is much lower. However, some of the residues still show correlation coefficients greater than 0.9. The latter is due to a limited number of samples (4 states). If more states are included, these correlations will become much weaker or filtered out. When only half of the protein resonances have linear changes in the chemical shifts, (mixed case, Figure 4D) it is possible to see a higher density of correlations in the map for the first part of the protein, while weaker correlations are observed for the remaining residues. To identify the large correlation clusters, we applied a hierarchical clustering analysis to the data (see dendrograms in **Figure S2**). The largest clusters obtained from these calculations are shown in the correlation matrices of Figure 5. To select the clusters, we used a threshold of 0.99. For the ideal case (Figure 5A), there is only one cluster, which includes all of the protein's residues. Addition of noise reduces the size of the largest cluster (Figure 5B), while in the random case (Figure 5C), the correlations among the residues through the protein are sparse, with only a handful of residues contributing to the largest correlated cluster. In the mixed case (Figure 5D), the method shows more dense correlations for the linear trajectories of the residues, while the fictitious correlations are rather sparse.

Mapping Protein kinase A Intramolecular Allostery using CONCISE and CHESCA

We applied CONCISE to the conformational transitions of PKA-C as mapped by [^{1}H,^{15}N]-HSQC spectra for four different states of the enzyme: apo, binary (AMP-PNP bound), binary (ATP-bound), and ternary (AMP-PNP and PKI_{5-25} bound). Peak positions for the four experimental data sets are shown in Figure 6A. The distributions of the residue positions in the different states are reported in Figure 6B. The two limits of the linear trajectories are represented by the apo and the ternary complex, corresponding to the open and closed conformations, respectively [35]. The average PC scores obtained using the full set

of residues (dashed lines in Figure 6B) are within reliable thresholds, with the two extremes of the linear relationship separated by 2.22 standard units and with a spread of about one standard deviation. After filtering out the residues characterized by low linearity and small perturbations, the results obtained with the reduced set (solid lines in Figure 6B) improved significantly, *i.e.*, the distance between the two extremes increases and the spread is reduced to less than one standard deviation. Overall, the positions of the four states (−1.11, −0.50, 0.28, and 1.33) correspond to equilibrium positions of 0, 0.25, 0.57, and 1, defining the fractional populations of the different states. The binding of the two different nucleotides induces two intermediate conformations: one (AMP-PNP bound) slightly shifted toward the open state and the second (ATP-bound) shifted toward the closed state. The spread of the distribution indicates the collective behavior of the residues, since it measures the extent to which each residue approaches a defined state. Since the distributions of the chemical shifts are directly related to the populations of the different states, it is possible to combine this information with thermocalorimetric measurements and define a conformational energy landscape for PKA-C. To construct the free energy basins for the open, intermediate, and closed states of the kinase along the open to close reaction coordinate, we used a harmonic potential underlying the Boltzmann distribution for each state:

$$E(s) = \left(\frac{RT}{2\sigma^2}\right)(s-\mu)^2$$

where *s* is the position along the average PC score in standard deviation units, and μ and σ are the mean and standard deviation for each distribution. Nucleotide and substrate binding were mapped along the “Ligand” reaction coordinate (see Figure 7), with the states uniformly distributed along this coordinate and widths fitted to reproduce the experimental transition states barriers. The transition from apo to ATPγN-bound was mapped using the dissociation constant K_d (39 μM) and the k_{on} (0.7 10^6 to 2.8 10^6 $M^{-1}s^{-1}$) from ATPγN binding [24], which resulted in an energy barrier ranging from 8.7 to 9.5 kcal mol^{-1}, and a free energy difference between the two minima of −6.0 kcal mol^{-1}. In addition, using the K_d for PKI binding to the ADP:PKA binary form (0.03 to 0.06 μM)[24], we determined an upper limit (−9.9 to −10.3 kcal mol^{-1}) for the free energy associated to the transition from the binary to the ternary complex. For the transition state between the binary and ternary forms, we were able to set an upper limit of about 7 kcal mol^{-1} based on the following considerations: first, the transition state barrier for PKI and PLN binding should be comparable (Hammond’s postulate); second, the rate limiting step is associated with ADP release, which sets an upper limit of about 15 kcal mol^{-1} to S16p-PLN dissociation; and third, the K_d for PLN binding to ADP:PKA (10±4 μM) brackets its binding free energy between −6.7 and −7.2 kcal mol^{-1}. Based on these considerations, we modeled the transition state (TS) to be about 3 kcal mol^{-1}. Figure 7 shows the free energy surface with three energy basins, corresponding to the apo, intermediate and closed state. The shapes of the minima correspond to the density of the populations calculated from the chemical shifts. These results are in qualitative agreement with the molecular dynamics calculations[22] where the ternary complex was shown to sample a narrow minimum in the closed conformation, while the nucleotide-bound binary form explored a wider range of conformations along the open/closed reaction coordinate.

To analyze the extent of the chemical shift correlations between residues in the different part of the kinase, we mapped the largest cluster of correlated residues on the correlation matrix (Figure 8A). These clusters are indicative of the effects of the ligands on the conformational equilibrium of the enzyme and on the propagation of the allosteric signals throughout the structure[26]. We found that a large number of residues report on the open-to-close transitions of the enzyme. Interestingly, these correlations span the entire enzyme structure (Figure 8B), indicating a collective response of the protein to the binding events. Several residues with large chemical shift changes ($\Delta\delta > 0.15$ ppm, red spheres in Figure 8C) do not follow linear paths and are mostly localized near the ligand binding sites. For these residues it is not possible to discern the perturbations associated with the opening and closing equilibrium from the electrostatic and structural changes induced locally by ligand binding. Within the largest correlated cluster, we observe that residues on both the small and large lobes respond to ligand binding in a correlated fashion (Figure 8B), and among these only a couple of residues are in direct contact with the nucleotide. Within the cluster of correlated residues, we identified several that belong to loops that play key roles in catalysis. Notably, we observe a group of residues in the Mg^{2+}-positioning loop (Arg-190, Val-191, and Gly-193), which we have previously found to allosterically coupled to the active site perturbations caused by myristoylation of the PKA-C N-terminus[37]. Other residues characterized by correlated behavior include Trp-196 (activation loop), Ile-209, and Leu-211 (peptide positioning loop), and Glu-334 (acidic patch). Interestingly, there are also residues that trace the inner core of the enzyme, including Ala-40 that contributes to forming the C-spine [38], Val-104, and Phe-100 of the C-helix/β4 loop (peripheral to the C-spine).

DISCUSSION

Proteins exert their function via binding of ligands or other partners. Binding of a ligand to rigid protein scaffolds follows a mechanism reminiscent of the classical lock-and-key [39] or induced-fit mechanisms [40]. However, many proteins undergo conformational interconversions involving domains or loops, and in more extreme cases, the entire protein. For instance, signaling proteins such as protein kinases are notoriously dynamic molecules [41, 42], with conformational isomers distributed among high and low energy states resembling classical folding funnels [1, 2, 4]. The extent of conformational dynamics manifested by a protein defines the shape of these free energy basins as well as the heights of the energy barriers between the different basins [43]. Dynamic proteins display broad basins and low energy barriers between them, whereas more rigid proteins populate low energy basins and possess high energy barriers between the different states [44]. A revised view of allostery suggests that binding events are interpretable as conformational shifts of proteins among the different energy basins [45], with ligands *selecting* those conformers that are complementary to their structural and dynamic features, shifting the population of the biomacromolecules from the unligated to the more stable ligated state [46-48].

NMR chemical shifts are atomic reporters of the conformational equilibria that the proteins undergo upon ligand binding. Although the extent of the chemical shift perturbations has been routinely used to identify binding hotspots and estimate the populations of the conformers, more recent analyses of the chemical shift trajectories for two-state equilibria suggest that statistical correlations can give a more complete view of the conformational

transitions of proteins and identify possible allosteric networks [26, 27]. The latter is particularly important to rationalize the conformational transitions of PKA-C [49, 50], where we observed chemical shift changes interspersed throughout the entire enzyme. When analyzed using the compounded $^{1}H/^{15}N$ chemical shifts of the enzyme fingerprint, we were able to identify contiguous paths localized in the proximity of the nucleotide and ligand binding sites [24]. However, in several instances we were not able to interpret the long-range chemical shift perturbation sparsely distributed across the entire enzyme. Nonetheless, many of these changes followed linear trajectories upon ligand binding [22, 23].

The CONCISE approach presented here enables us to filter out chemical shift changes occurring in the immediate proximity of the ligand binding crevices from those that report on the conformational transitions, identifying the progressive conformational transitions of the enzyme from the apo (open), intermediate (partially closed) to the ternary complex (fully closed state) upon nucleotide and substrate binding. Importantly, our analysis shows that the residues involved in the opening and closing transitions act collectively, i.e., with a coordinated behavior upon ligand binding (Figure 6C). Interestingly, CONCISE discriminates between the positions of the equilibrium reached upon binding two nucleotides, ATP and ATPγN (AMP-PNP). ATPγN has been considered a non-hydrolyzable nucleotide, which mimics the effects of ATP binding. Our analysis reveals that ATPγN induces a conformational state slightly more open than the corresponding ATP-bound enzyme. The density distribution of the amide resonances upon addition of ATPγN is shifted toward the apo state of the enzyme, with an average position for the ATPγN-bound of ~27%; whereas the corresponding distribution of the population of the ATP-bound kinase is shifted toward the closed state with an average value of ~62%. A possible explanation is that the geometry around the β- and γ-phosphates is crucial to define the intermediate state, and small changes in the geometry of the γ-phosphate may cause the enzyme to adopt a more open conformation. The latter suggests that ATPγN is not a perfect mimic for ATP.

How does the kinase reach a defined conformational state? Are there preferential paths activated by ligand binding that modulate allosteric response (networks)? Are the perturbations localized or does the protein respond collectively?

For PKA-C, we found that the major cluster of residues reporting on the opening and closing of the enzyme is not localized in a specific area; rather it is distributed throughout the enzyme across the small and large lobes. Although it is possible that some of the missing links that would physically connect all correlated residues are unassigned or filtered out due to large nearest neighbor effects, from these data it is not possible to explain all the observed correlations in terms of spatial proximity. Perhaps, the best explanation for allosteric behavior in PKA-C is a combination of well-defined pathways [51, 52] together with non-structure-based allosteric phenomena, reflecting thermal fluctuation (local folding and unfolding) in agreement with ensemble-based model [53], where allostery is formulated as thermodynamic coupling between protein domains in the absence of physical couplings. It should be noticed that the chemical shift analysis provided here gives only a partial view of signal propagation throughout the enzyme. A definite answer will require analysis of further isotopic probes, including those of methyl side chain groups, that can trace the changes in the hydrophobic residues populating the core of the kinase [54].

CONCLUSIONS

In conclusion, we presented a statistical interpretation of the linear trajectories of the chemical shifts that gives a quantitative and concise view of the transitions associated with ligand binding. In the case of PKA-C, our method describes the collective behavior of the resonances through the structural transitions upon ligand binding, and shows that it is possible to discern between different nucleotides, nucleotide analogs, and competitive inhibitors in the conformational transitions of the enzyme. This method is also applicable to other structural perturbations, such as post-translational modifications or mutations, to identify the average conformational state of a protein. The combination of the population densities derived from the chemical shifts with thermocalorimetric and kinetic data enables the description of the free energy landscape of proteins. Therefore, this method gives a concise view of the way in which different ligands, such as drugs and peptides, can affect the conformational equilibrium and will be instrumental in the design of positive or negative allosteric effectors.

Acknowledgments

This work is supported by the National Institute of Health (GM100310 and GM72701 to G.V.). We thank Prof. Davide Ferrari (U. of Melbourne) and Fa-An Chao for helpful discussions. Prof. Giuseppe Melacini (McMaster University) for critical reading of the manuscript. The NMR experiments were carried out at the Minnesota NMR Center (MNMR) and the calculations were performed at the Minnesota Supercomputing Institute.

ABBREVIATIONS

PKA	protein kinase A
PKA-C	C-subunit of PKA
PKI	protein kinase A inhibitor peptide
PCA	principal component analysis

References

1. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Proteins. 1995; 21:167–195. [PubMed: 7784423]
2. Oliveberg M, Wolynes PG. Q Rev Biophys. 2005; 38:245–288. [PubMed: 16780604]
3. Tsai CJ, Kumar S, Ma BY, Nussinov R. Protein Science. 1999; 8:1181–1190. [PubMed: 10386868]
4. Kumar S, Ma BY, Tsai CJ, Sinha N, Nussinov R. Protein Science. 2000; 9:10–19. [PubMed: 10739242]
5. Csermely P, Palotai R, Nussinov R. Trends in Biochemical Sciences. 2010; 35:539–546. [PubMed: 20541943]
6. Monod J, Wyman J, Changeux JP. Journal of molecular biology. 1965; 12:88–118. [PubMed: 14343300]
7. Kay LE. Journal of Magnetic Resonance. 2005; 173:193–207. [PubMed: 15780912]
8. Tugarinov V, Hwang PM, Kay LE. Annual Review of Biochemistry. 2004; 73:107–146.
9. Tzeng SR, Pai MT, Kalodimos CG. Methods in molecular biology. 2012; 831:133–140. [PubMed: 22167672]
10. Tzeng S-R, Kalodimos CG. Nature. 2009; 462:368–U139. [PubMed: 19924217]
11. Tzeng SR, Kalodimos CG. Nat Chem Biol. 2013; 9:462–465. [PubMed: 23644478]

12. Bodenhausen G, Ruben DJ. Chem Phys Letters. 1980; 69:185–189.
13. Zhuravleva A, Gierasch LM. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:6987–6992. [PubMed: 21482798]
14. Revington M, Zhang YB, Yip GNB, Kurochkin AV, Zuiderweg ERP. Journal of Molecular Biology. 2005; 349:163–183. [PubMed: 15876376]
15. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A. Proc Natl Acad Sci U S A. 2008; 105:4685–4690. [PubMed: 18326625]
16. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:9615–9620. [PubMed: 17535901]
17. Williamson MP. Prog Nucl Magn Reson Spectrosc. 2013; 73:1–16. [PubMed: 23962882]
18. Tzeng SR, Kalodimos CG. Nature. 2012; 488:236–240. [PubMed: 22801505]
19. Li P, Martins IRS, Amarasinghe GK, Rosen MK. Nature Structural & Molecular Biology. 2008; 15:613–618.
20. Strickland D, Yao X, Gawlak G, Rosen MK, Gardner KH, Sosnick TR. Nature Methods. 2010; 7:623–U618. [PubMed: 20562867]
21. Das R, Melacini G. The Journal of biological chemistry. 2007; 282
22. Masterson LR, Shi L, Metcalfe E, Gao J, Taylor SS, Veglia G. Proc Natl Acad Sci U S A. 2011; 108:6969–6974. [PubMed: 21471451]
23. Masterson LR, Cheng C, Yu T, Tonelli M, Kornev A, Taylor SS, Veglia G. Nat Chem Biol. 2010; 6:821–828. [PubMed: 20890288]
24. Masterson LR, Mascioni A, Traaseth NJ, Taylor SS, Veglia G. Proc Natl Acad Sci U S A. 2008; 105:506–511. [PubMed: 18178622]
25. Sakurai K, Goto Y. Proc Natl Acad Sci U S A. 2007; 104:15346–15351. [PubMed: 17878316]
26. Selvaratnam R, Chowdhury S, VanSchouwen B, Melacini G. Proc Natl Acad Sci U S A. 2011; 108:6133–6138. [PubMed: 21444788]
27. Selvaratnam R, VanSchouwen B, Fogolari F, Mazhab-Jafari MT, Das R, Melacini G. Biophys J. 2012; 102:630–639. [PubMed: 22325287]
28. Akimoto M, Selvaratnam R, McNicholl ET, Verma G, Taylor SS, Melacini G. Proc Natl Acad Sci U S A. 2013; 110:14231–14236. [PubMed: 23946424]
29. Selvaratnam R, Mazhab-Jafari MT, Das R, Melacini G. PLoS One. 2012; 7:e48707. [PubMed: 23185272]
30. Dawson JE, Farber PJ, Forman-Kay JD. PLoS One. 2013; 8:e74347. [PubMed: 24058550]
31. Stollar EJ, Lin H, Davidson AR, Forman-Kay JD. PLoS One. 2012; 7:e51282. [PubMed: 23251481]
32. Arai M, Ferreon JC, Wright PE. J Am Chem Soc. 2012; 134:3792–3803. [PubMed: 22280219]
33. Chen L, Balabanidou V, Remeta DP, Minetti CA, Portaliou AG, Economou A, Kalodimos CG. Mol Cell. 2011; 44:734–744. [PubMed: 22152477]
34. Knighton DR, Zheng JH, Eyck L. F. Ten, Xuong NH, Taylor SS, Sowadski JM. Science. 1991; 253:414–420. [PubMed: 1862343]
35. Johnson DA, Akamine P, Radzio-Andzelm E, Madhusudan, Taylor SS. Chemical Reviews. 2001; 101:2243–2270. [PubMed: 11749372]
36. Konuma T, Lee YH, Goto Y, Sakurai K. Proteins. 2013; 81:107–118. [PubMed: 22927212]
37. Cembran A, Masterson LR, McClendon CL, Taylor SS, Gao J, Veglia G. Biochemistry. 2012; 51:10186–10196. [PubMed: 23205665]
38. Kornev AP, Taylor SS. Biochim Biophys Acta. 2010; 1804:440–444. [PubMed: 19879387]
39. Fischer E. Berichte der deutschen chemischen Gesellschaft. 1894; 27:2985–2993.
40. Koshland DE. Proceedings of the National Academy of Sciences. 1958; 44:98–104.
41. Huse M, Kuriyan J. Cell. 2002; 109:275–282. [PubMed: 12015977]
42. Pellicena P, Kuriyan J. Current Opinion in Structural Biology. 2006; 16:702–709. [PubMed: 17079130]
43. Ma B, Kumar S, Tsai CJ, Nussinov R. Protein Eng. 1999; 12:713–720. [PubMed: 10506280]

44. Tsai CJ, Ma BY, Sham YY, Kumar S, Nussinov R. Proteins-Structure Function and Genetics. 2001; 44:418–427.
45. Cui Q, Karplus M. Protein Science. 2008; 17:1295–1307. [PubMed: 18560010]
46. Kuzu G, Keskin O, Gursoy A, Nussinov R. Methods in molecular biology. 2012; 819:59–74. [PubMed: 22183530]
47. Nussinov R, Ma B. BMC biology. 2012; 10:2. [PubMed: 22277130]
48. Nussinov R, Ma B, Tsai CJ. Biophys Chem. 2013 DOI: 10.1016/j.bpc.2013.10.002.
49. Masterson LR, Cembran A, Shi L, Veglia G. Advances in protein chemistry and structural biology. 2012; 87:363–389. [PubMed: 22607761]
50. Veglia G, Cembran A. The FEBS journal. 2013; 280:5608–5615. [PubMed: 23902454]
51. Lockless SW, Ranganathan R. Science. 1999; 286:295–299. [PubMed: 10514373]
52. Suel GM, Lockless SW, Wall MA, Ranganathan R. Nature Structural Biology. 2003; 10:59–69.
53. Hilser VJ, Wrabl JO, Motlagh HN. Annual review of biophysics. 2012; 41:585–609.
54. Taylor SS, Ilouz R, Zhang P, Kornev AP. Nature reviews. Molecular cell biology. 2012; 13:646–658.

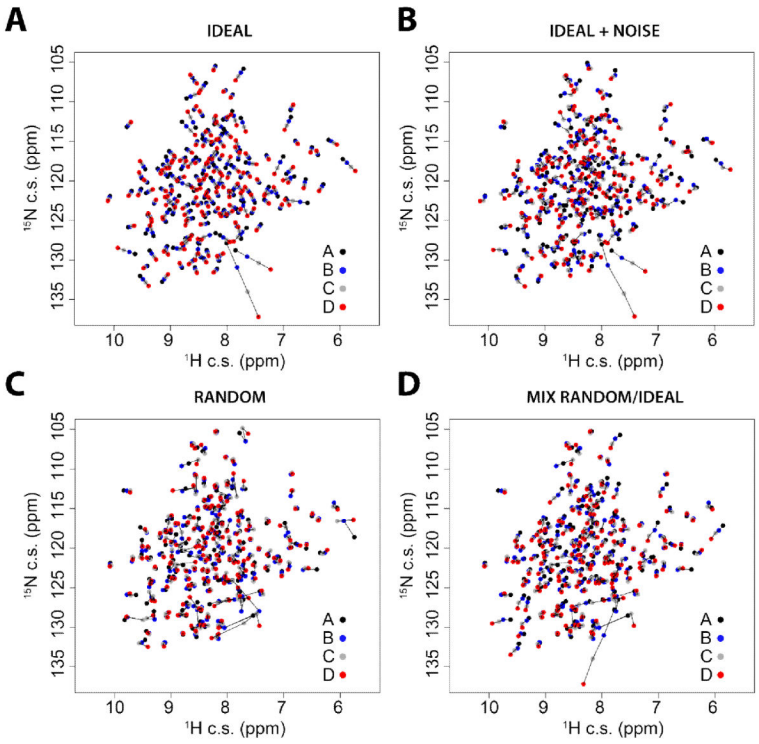

Figure 1. Synthetic [^{1}H,^{15}N]-HSQC spectra mimicking possible chemical shift perturbations upon ligand binding

A) ideal case (perfect linear correlations); **B)** ideal case with the addition random of noise; **C)** completely random behavior of the chemical shifts; **D)** mixed case, where half of the peaks follow random chemical shift trajectories and the other half follow linear behavior with noise added. For each spectrum, the trajectories of the chemical shifts for the four states (A, B, C, D) are connected with black lines.

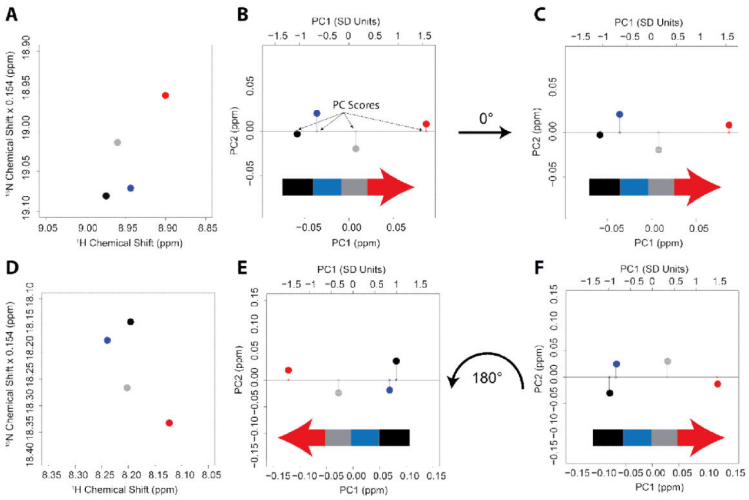

Figure 2. Graphical representation of the PCA of the linear chemical shift trajectories

PCA is applied to the chemical shifts of two residues for the *ideal+noise* case. Panels **A** and **D** show a close-up of the peaks positions in the synthetic [^{1}H,^{15}N]-HSQC spectra. The data projected along the two principal components are shown before (panels B and E) and after (panels C and F) orientation. For PC1, the top scale shows the projection normalized to one standard deviation.

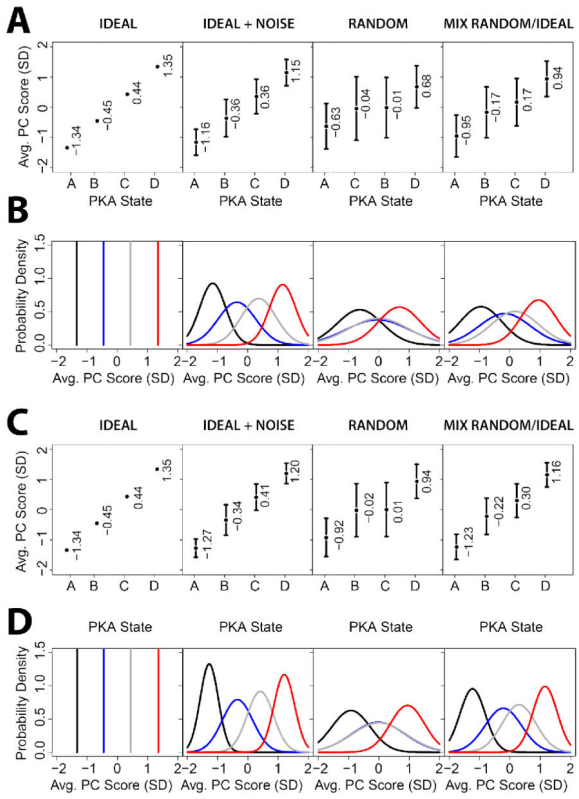

Figure 3. Application of the CONCISE method to four synthetic data sets represented in Fig. 1

A) Average PC scores versus the different states (A, B, C, and D) for all residues. The mean equilibrium values (circles) and standard deviations (vertical bars). **B)** Normal distributions for the residues around the mean values. **C)** Average PC scores versus the different states plotted for the subset of residues showing linear trajectories. **D)** Normal distributions for the case C for linear residues.

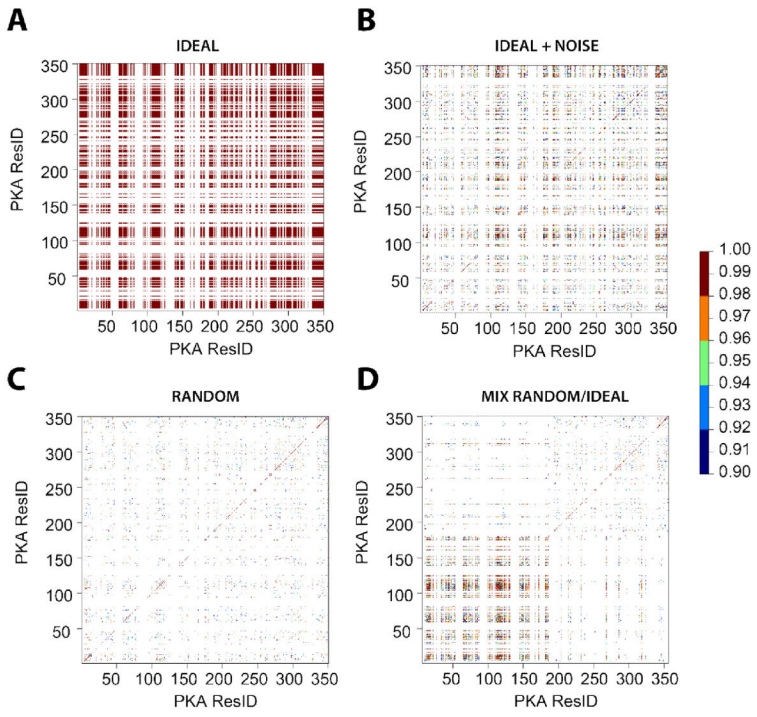

Figure 4. PCA correlations for the four synthetic states

The correlation coefficients between the PC1 scores for all residues are plotter using a scale showing only correlations higher than 0.9.

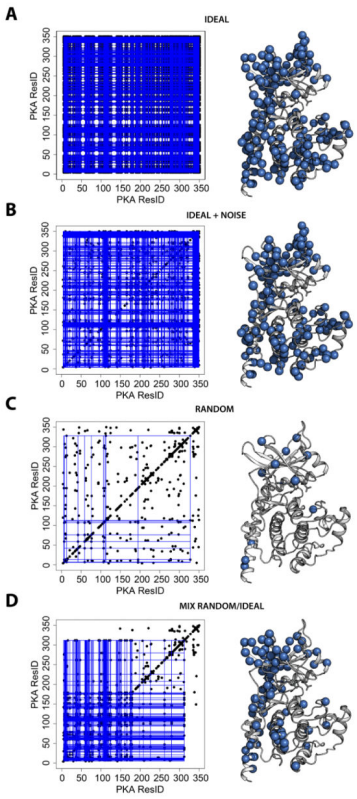

Figure 5. Correlation-Matrix as part of the CHESCA analysis for four synthetic states

On the matrix (left panel), the largest cluster resulting from hierarchical clustering (correlation cutoff of 0.99) is indicated with blue lines and mapped on the crystal structure of the kinase using blue spheres.

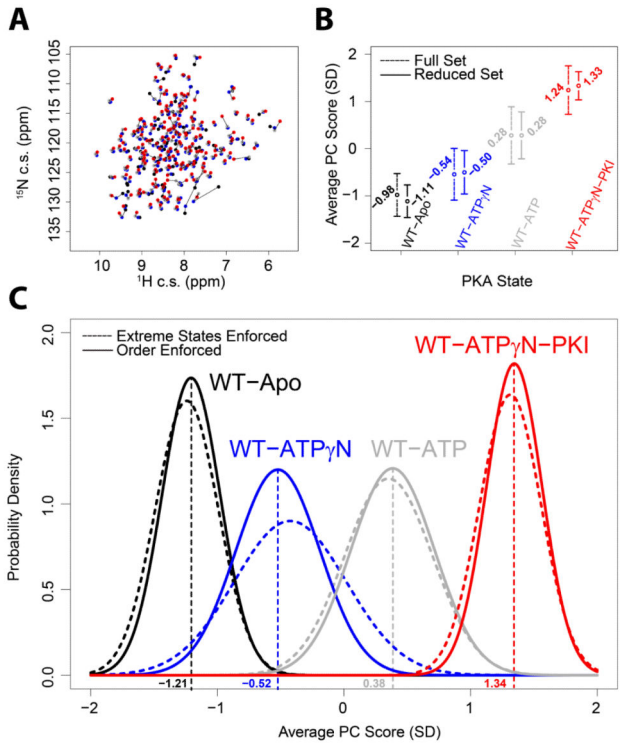

Figure 6. CONCISE analysis applied to the kinase experimental data

A) Superposition of the kinase resonances of [^{1}H,^{15}N]-HSQC spectra for the four states with peak positions shown as dots of different colors connected by black lines. **B)** The equilibrium position analysis for PKA-C four states is shown as circles and vertical bars, representing averages and standard deviations, respectively. Dotted lines show the results obtained using all available residues, and solid lines refer to the reduced set of residues. **C)** The equilibrium positions for the reduced set of residues are plotted as normal distributions centered in the average and with width given by the standard deviation. Dotted lines display the results obtained by discarding all the residues for which the apo and closed states are not the two extremes; solid lines show the results obtained by discarding all residues in which the states are not in following the apo-ATPγN-ATP-closed order.

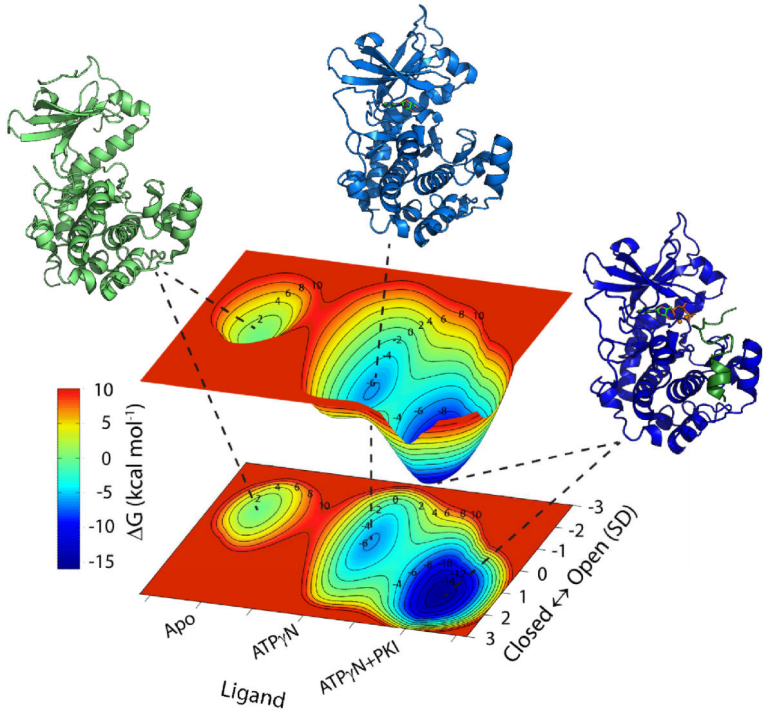

Figure 7. PKA-C free energy landscape

The PKA-C free energy profile is shown along the ligand binding and along the open/closed reaction coordinates. The Apo state is shown in green (pdb code 1J3H), the binary form in cyan (1BKX), and the ternary complex in blue (1ATP).

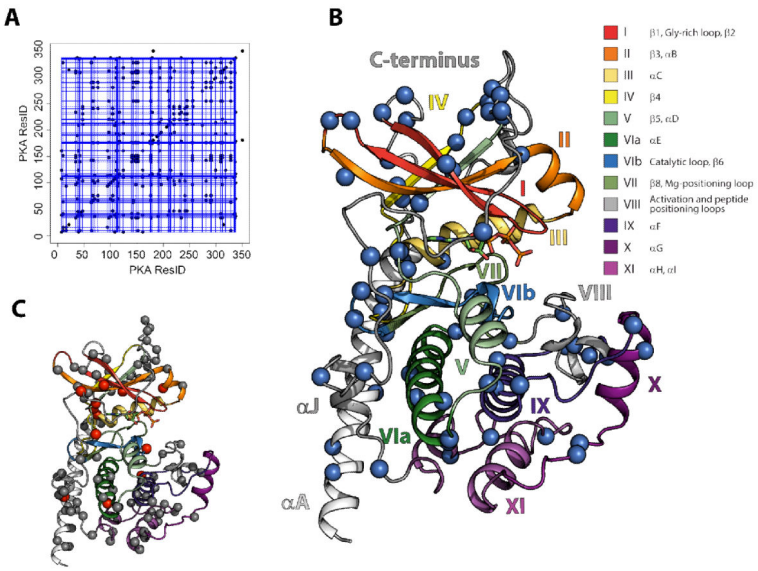

Figure 8. PKA-C Collective Behavior

A) The largest cluster of correlated residues in the linear data set obtained with a cutoff of 0.99 is shown as blue lines on a matrix representation. **B)** The largest correlated cluster is mapped onto the kinase structure as blue spheres. **C)** The residues that do not follow linear paths are shown as grey spheres, and in red are highlighted those characterized by large chemical shifts perturbations.