PROBING ALLOSTERY IN EPAC BY NMR SPECTROSCOPY

PROBING ALLOSTERY IN THE EXCHANGE PROTEIN DIRECTLY ACTIVATED BY cAMP (EPAC) USING NMR SPECTROSCOPY

By RAJEEVAN SELVARATNAM, HONOURS, B.Sc.

A Thesis Submitted to

the School of Graduate Studies in Partial Fulfilment

of the Requirements for the Degree

Doctor of Philosophy

Doctor of Philosophy (2013) McMaster University

(Chemical Biology) Hamilton, Ontario

Thesis Title: Probing Allostery in the Exchange Protein directly Activated by cAMP (EPAC) Using NMR Spectroscopy

Author: Rajeevan Selvaratnam, Honours B.Sc. (University of Toronto)

Supervisor: Associate Professor Giuseppe Melacini

Number of Pages: xiv, 121

Thesis Abstract:

Exchange proteins directly activated by cAMP (EPAC) are guanine nucleotide exchange factors for the small GTPases, Rap1 and Rap2. The central regulatory module of EPAC is a cAMP binding domain (CBD), which in the absence of cAMP provides auto-inhibition of the catalytic guanine nucleotide exchange activity. Binding of the allosteric effector, cAMP, removes the auto-inhibition exerted by the CBD of EPAC. Herein, we investigate through NMR spectroscopy the structural and dynamical basis of auto-inhibition and cAMP-dependent allosteric activation in the CBD of EPAC. Specifically, the work described in this dissertation proposes novel methods that utilize NMR chemical shifts to define the network of residues that mediates long-range intra-molecular signalling, *i.e.* the chemical shift covariance analysis (CHESCA) and the chemical shift projection analysis (CHESPA). Using CHESCA as explained in Chapter 2, we identified an allosteric network that bridges the sites of cAMP-binding and cAMP-dependent structural changes to those of cAMP-dependent dynamical changes, which are critical for the release of auto-inhibition. The CHESCA results therefore rationalize how cAMP leads to activation through modulation of both structure and dynamics. In order to dissect the determinants of auto-inhibition in the absence of cAMP, several mutations along the signaling pathways identified by CHESCA were implemented and their effect on the auto-inhibitory conformational equilibrium of the apo-CBD was assessed through CHESPA, as outlined in Chapters 3 and 4. Overall, we have shown how CHESCA and CHESPA provide unprecedented insight into the allosteric networks underlying auto-inhibition and cAMP dependent activation in the CBD of EPAC. In addition, the methods employed here to map EPAC allostery are likely to be generally applicable to other systems.

Acknowledgements:

First and foremost, to my supervisor Dr. Giuseppe Melacini, to whom I am greatly indebted to for the great guidance and mentorship throughout both the high and low tides of the entire graduate journey. From you I learned the importance of paying attention to details and I take away a lot more than I state here. It has been a great pleasure working alongside you and I could not have asked for more.

I have been very fortunate to have always been in good company and a lot of my graduate work would not have been fruitful without the past and present lab members: Madoka Akimoto, thanks for helpful discussions relating to wet lab work; Bryan VanSchouwen, thanks for the Molecular Dynamics work that strengthened our experimental findings. Mostafa Almagal and Naeimeh Jafari, thanks for stimulating discussions. Raul Das, Mohammad Mazhab-Jafari, Somenath Chowdhury, Annie Raditsis, and Soumita Das – Thanks for all the help in getting me acquainted with the lab and my project during the early phase of my graduate ventures.

Tyler McNicholl, Julijana Milojevic, and Amir Bashiri: Thanks for the good times, both inside and outside the lab. I truly appreciated your friendship and the accompanying discussions and quarrels over the many lunch and coffee outings☺.

To my undergraduate thesis/summer students that I had the privilege to supervise – The learning was bidirectional, I learned a lot from you all: Amir Bashiri, Katherine Li, Lawrence Lin, and Nilay Shah.

My committee members, Drs. Alba Guarné and Joaquin Ortega. I very much appreciated your advice, criticism and enjoyed conversing with you during these past years. Dr. Alba Guarné, thanks for always providing access to your lab and being a great collaborator, in which regards, I am grateful for the opportunity to have worked with Lindsay Matthews and Melanie Gloyd.

Drs. Richard and Raquel Epand, thanks for allowing me to use your CD spectrometer, even on the weekends and at odd times. Dr. Murray Junop, thanks for the sub-cloning tips and discussions. The NMR Facility Staff, most notably, Drs. Bob Berno and Hilary Jenkins for always being available. I am also grateful to my undergraduate thesis advisor, Dr. Sonia Gazzarrini, for providing me a sound learning experience and foundation in molecular biology.

Lastly, but of great personal importance: I own an insurmountable of gratitude to my parents for all that has been possible. Most importantly, thanks for being supportive of my choice of study, even though you probably had only the faintest clue as to what I was “cooking” in the lab or where all this would lead

me to. To my beloved Periyamma and Ammama, for raising me in Jaffana and Toronto. Periyamma, thanks for always providing the hospitality and comfort, especially during the high school years, without which I am uncertain if I could have pursued post-secondary education so easily. Ravi Mama, to whom I am also deeply indebted: Thanks also for always providing the necessary tools and foundations, sparking my interest in science, and the mentorship throughout my childhood. To my dear Thangachi Inthuja, thanks for always listening, providing me with your amazing perspectives, and most importantly for challenging me on many different levels and contexts. Thambi Elilan, thanks for keeping me updated with the world when I was getting caught up in the lab. Rani Akka, "the apple does not fall far from the tree" – thanks for always having my interest at heart.

Dedicated to my Periyamma, Navamalar Sivasubramaniyam, 1945 – 2011.

For the calming hospitality, encouragement and love.

அன்பு காட்டி ஆர்வத்தை ஊட்டி அமைதியான படிப்பதற்கேற்ற சூழலை ஏற்படுத்தித் தந்த மறைந்த என் பெரியம்மாவுக்கு (நவமலர் சுப்ரமணியம் 1945 – 2011) இது சமர்ப்பணம்.

Table of Contents

List of Main Text Figures

Chapter 1: Introduction

Chapter 2: Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts

Chapter 3: The Projection Analysis of NMR Chemical Shifts Reveals Extended EPAC Auto-inhibition Determinants

Chapter 4: The Auto-inhibitory Role of the EPAC Hinge Helix as Mapped by NMR

List of Supplementary Figures

Chapter 2: Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts

Chapter 3: The Projection Analysis of NMR Chemical Shifts Reveals Extended EPAC Auto-inhibition Determinants

Chapter 4: The Auto-inhibitory Role of the EPAC Hinge Helix as Mapped by NMR

List of Tables

Chapter 2: Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts

Chapter 3: The Projection Analysis of NMR Chemical Shifts Reveals Extended EPAC Auto-inhibition Determinants

Chapter 4: The Auto-inhibitory Role of the EPAC Hinge Helix as Mapped by NMR

List of Abbreviations

AC	agglomerative clustering
BBR	base binding region
cAMP	adenosine 3’5’ cyclic monophosphate
CBD	cyclic-AMP binding domain
CDC25HD	cell division cycle 25-homology domain
cGMP	guanosine 3’5’ cyclic monophosphate
CHESCA	chemical shift covariance analysis
CHESPA	chemical shift projection analysis
cNTs	cyclic-nucleotides (cAMP and cGMP)
CNG	cyclic-nucleotide gated
CR	catalytic region
DEP	dishevelled, egl-10, pleckstrin
DSSP	define secondary structure of proteins
EPAC	exchange proteins activated by cAMP
GC	guanylyl cyclase
GEF	guanine nucleotide exchange factors
GPCR	G-protein coupled receptor
HCN	Hyperpolarization-activated cyclic nucleotide modulated
HSQC	hetero-nuclear single quantum coherence
IL	ionic latch
MD	molecular dynamics
NMR	nuclear magnetic resonance
NTHB	N-terminal helical bundle
PBC	phosphate binding cassette
PC	principal component
PKG	protein kinase G
PKA	protein kinase A
RA	ras association
REM	ras exchange motif
RR	regulatory region
SI	supplementary/supporting information
SVD	singular value decomposition
Wt	wild type

Chapter 1

Introduction

1.1 The Molecular Basis for Regulation - Allostery

Signalling cascades are regulated by "molecular switches" that bind effectors with high specificity (1, 2). Such molecular recognition results in conformational differences, often at sites remote from the effector binding site. This phenomenon of regulation at a distance, *i.e.* the intra-protein signal transduction event that dictates the state of the molecular switch subsequent to the effector binding, is commonly referred to as *allostery* (3–6). The concept of allostery was originally formulated in the context of large multi-subunit and/or oligomeric and symmetric proteins to explain regulatory and cooperative behaviour (6, 7). However, it is now clear that allostery is essentially an intrinsic property of all proteins, including those that are monomeric, asymmetric, and even of single regulatory domains (4, 8–12).

In the absence of the appropriate allosteric effector, a protein may be in an inactive and auto-inhibited state. The mechanism by which the auto-inhibition is relieved or reinforced is a consequence of an allosteric perturbation, which is typically due to ligand/effector binding, post-translational modifications, mutations and/or environmental changes in solution (3, 4, 8). In other words, allostery is an innate function of auto-inhibition and vice versa.

Currently, the two prominent mechanistic models of allosteric regulation are induced fit and conformational selection. The induced fit model proposes that it is the allosteric effector that drives the conformational change toward the active state. According to the induced fit model, no active conformation is sampled prior to binding. On the contrary, the conformational selection model posits that there is an ensemble of multiple conformers with a minor pre-existing population of the active species even in the absence of the allosteric effector. Preferential binding of the effector to this minor population results in a population redistribution, which underlies the equilibrium shift toward the now more stable and dominant active state. Experimental evidence from NMR studies illustrates the presence of this minor population in several regulatory systems (3, 5, 13, 14).

It has also been proposed that both the induced fit and conformational selection mechanisms can contribute to molecular recognition and removal of auto-inhibition. The mechanism that dominates is thought to be dependent on the kinetics of binding and conformational change as well as on the concentration of the allosteric effector (5). For example, at relatively low concentrations of the allosteric effector, or in the presence of substrates that promote the apo/active state, the conformational selection mechanism is thought to prevail(3–5). However, at relatively high concentrations of the allosteric effector, the induced fit pathway should be taken into consideration as well (5, 15).

1.1.1 Role of Allostery in Drug Design

Traditionally the active site, or the 'orthosteric site' as it has come to be called, has been the primary target for screening and rational drug design. However, allosteric sites have also attracted considerable attention in the last decade-and-a-half as potential drug targets. This is partly because allosteric sites are thought to be subject to lower evolutionary pressure to attain a highly specific function in comparison to orthosteric sites. As a result, the remote allosteric sites are often less conserved and provide greater subtype selectivity (16). Given this potential for enhanced selectivity, targeting allosteric sites is thought to be useful in the generation of novel agonists and antagonists or inhibitors, such as antibiotics that may be species specific (7).

In light of the benefits of allosteric sites, drug discovery endeavours have also resulted in the identification of novel allosteric sites. These novel sites are commonly referred to as the 'serendipitous allosteric sites' because they appear to have no endogenous allosteric effector, but drugs that sample the appropriate chemical space are able to bind in these "unconventional pockets" and remotely control protein function. This adaptive binding by inhibiting drugs occurs often between secondary structural

elements, in grooves and in cavities that are not part of previously known allosteric sites (16, 17). Structural investigations have revealed that drugs binding in the serendipitous allosteric sites often trap the protein in previously observed inhibited states (17). This suggests that having a map of the intra-molecular networks that bridge the sites underlying auto-inhibition can aid in the design of novel therapeutic inhibitors that down-regulate or "jam" the signal propagation between functional sites.

1.1.2 Role of Dynamics in Allostery

The structural comparison between the active and inactive end points as determined by X-ray or NMR structure determination methods, provides an initial framework for appreciating allosteric modulation. However, allostery is inherently a thermodynamic function of both structural and dynamic changes. The latter is not fully captured by the static representation of the average structures since functional events involve motions on a wide range of time-scales. That is, intra-protein signalling encompasses not only structural changes, but also the changes in intrinsic dynamics – which have been shown to relay information from one site to another, even without major structural changes (4, 5, 18–20). For example, in the case of the prokaryotic transcription factor, catabolite activator protein, that is activated by cAMP, Popovych et al., recently employed NMR measurements to show dynamics alone without major structural changes, modulates the negative cooperativity of the second unliganded subunit in the homodimeric protein (21). In this regard, NMR spectroscopy is then ideally suited for investigating allostery as it provides both structural and dynamic information at atomic resolution (18–20).

1.2 Second Messengers and cyclic-Nucleotides

Second messengers are essential endogenous allosteric effectors. These small molecules regulate intra-cellular responses to extra-cellular stimuli. One class of second messengers are cyclic-nucleotides (cNTs): guanosine 3'5' cyclic monophosphate (cGMP) and adenosine 3'5' cyclic monophosphate (cAMP). The generation of cAMP is initiated by the binding of a primary messenger molecule such as hormones and neurotransmitters to a G-protein coupled receptor (GPCR), which in-turn activates the appropriate α-subunit of the heterotrimeric G-protein. The stimulatory isoform of the α-subunit of the G-protein subsequently activates adenylyl cyclase, which catalyzes the conversion of ATP into cAMP (22, 23). The increase in concentration of cAMP leads to targeting of the archetypal cAMP binding domain (CBD) linked to various cellular function.

The CBD is an ancient regulatory module embedded in a number of multi-domain and supramodular proteins (Figure 1). These include the essential housekeeping proteins: tetrameric K^+ transporters, such as hyperpolarization modulated and cyclic-nucleotide gated (HCN) channel proteins; phosphotransferases, such as Protein Kinase A (PKA) and Protein Kinase G; and guanine nucleotide exchange factors (GEF), such as exchange proteins directly activated by cAMP (EPAC).

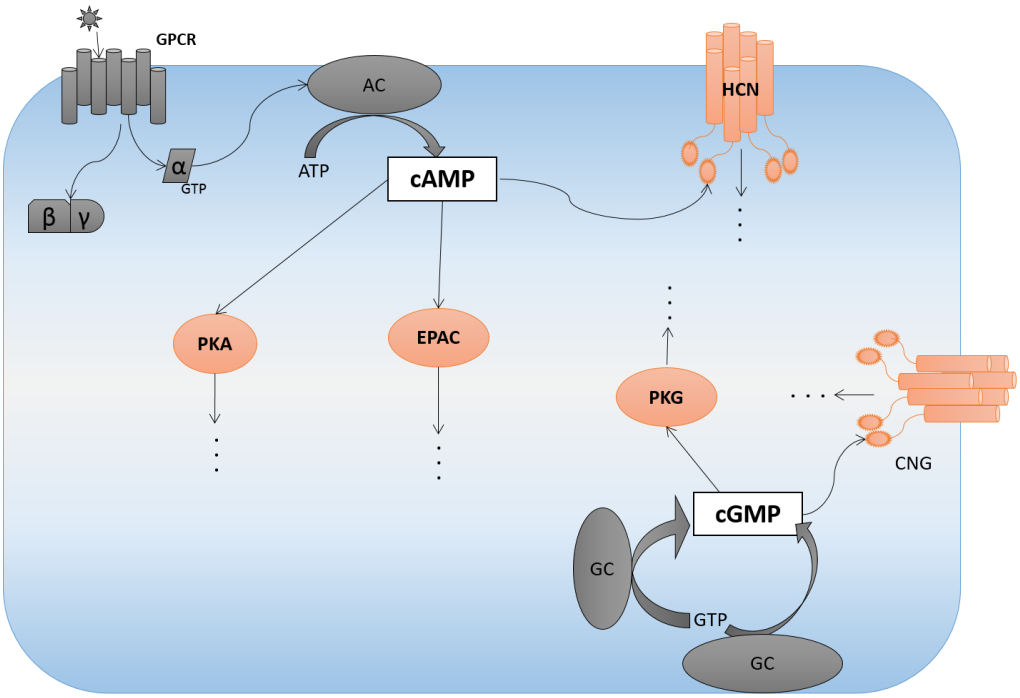

Figure 1: *Cyclic-nucleotide mediated signalling.* cAMP is generated through the binding of primary messenger molecule (shown as star) to GPCR (G-protein coupled receptor), which activates the α-subunit of the G-protein that binds adenylyl cyclase (AC). AC catalyzes the conversion of high energy molecule ATP into cAMP. Also shown is the production of cGMP from high energy molecule GTP by soluble or membrane bound guanylyl cyclase (GC). cGMP preferentially targets PKG and some cyclic-nucleotide gated channel (CNG) proteins.

1.3 EPAC – a primary mammalian receptor for cAMP

The CBD of EPAC is particularly important because it provides a cytosolic mammalian receptor for cAMP, different from HCN and PKA, which prior to the discovery of EPAC were believed to be the major target of cAMP in the cytosol (24, 25). EPAC is a guanine nucleotide exchange factor (GEF) for the small GTPases, Rap1 and Rap2. That is, EPAC is an allosterically controlled catalyst that upon activation by cAMP, greatly accelerates the dissociation of the Rap-bound GDP, so that GTP can then readily bind (24). Rap activity in-turn underlies fundamental cellular events such as growth, differentiation, secretion, and adhesion(2).

As an essential member of the cAMP signalling cascade, EPAC is expressed in nearly all human tissues (24, 26, 27) and exists as two well characterized paralogs (Figure 2a): EPAC1 and EPAC2, with the latter containing an additional CBD at the N-terminus that is nonessential for regulation and binds cAMP only weakly (24, 28). The sequence identity between the essential CBD of EPAC2 and that of EPAC1 is ~69% (Figure 2b). EPAC1 is more ubiquitous, and is involved in cAMP dependent immunoregulation and inflammatory response (26, 27). Hence EPAC1 regulation and expression has been suggested to have therapeutic implications in renal- and neuro-pathology as well as a vast number of other inflammatory diseases (26, 29). Therefore, the focus of this thesis is on EPAC1 (henceforth referred to as EPAC, unless otherwise noted) and specifically on its indispensable regulatory module, the CBD (Figure 2a).

Deletion of the CBD as a whole from EPAC results in constitutively active GEF activity (24). This suggests that the CBD includes the essential determinants (structural and dynamical) for auto-inhibition of apo-EPAC.

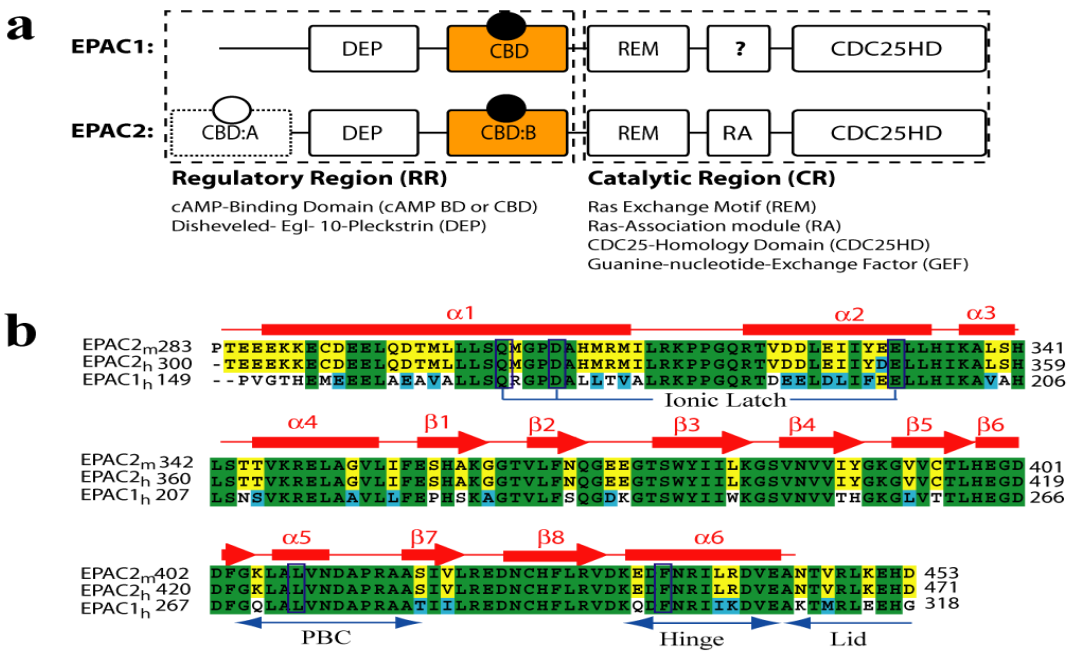

Figure 2: *EPAC1 vs. EPAC2.* a) Domain organization of EPAC1 and EPAC2. The catalytic region is composed of REM, RA, and CDC25HD. The CDC25HD is the central site of catalysis where the substrate protein Rap binds. The regulatory region is defined by the CBD and DEP domain. The latter is involved in membrane translocation (30, 31). Abbreviations are as defined in the figure. b) Sequence alignment between human EPAC1 and EPAC2 from both mouse and human. The secondary structure obtained from apo-EPAC2 (PDB ID: 2byv) is reported above the sequence. This figure was adapted from Das et al. (2008), ref. (9).

As in other characterized CBDs, the CBD of EPAC preserves the archetypal fold with eight β-strands forming a jelly-roll like topology (32–34) (Figure 2b and 3). Together, the β-strands constitute the β-subdomain of CBDs. The segment between β-strands 6 and 7 is the phosphate binding cassette (PBC) that contains a short helix called α5. This PBC segment interacts with the phosphate moiety of the cAMP molecule. The antiparallel strands β4-β5 typically interact with the adenine base and are referred to as the base binding region (BBR). The helices (α1-α4) at the N-terminus of the β-subdomain, are appropriately called the N-terminal helical bundle (NTHB), whereas the helix C-terminal to the β-subdomain is referred to as the α6 hinge helix. Together, the non-contiguous helices in the CBD are referred to as the α-subdomain.

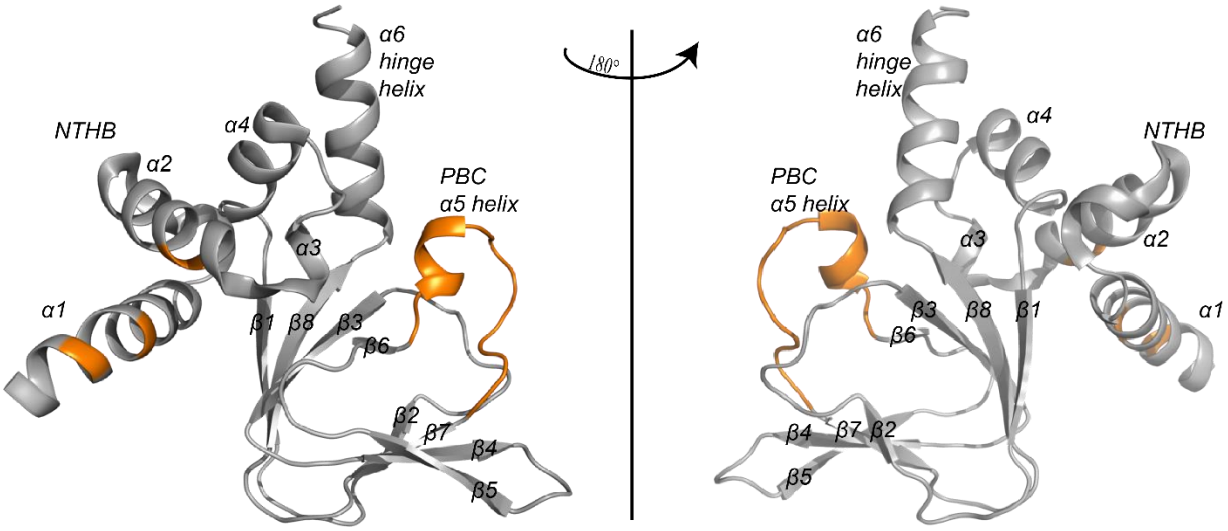

Figure 3: *The archetypal CBD of EPAC.* The N-terminal helical bundle (NTHB) includes helices α1-α4. The IL residues and the PBC segment where cAMP binds are colored in orange. This figure was generated using EPAC2 crystal structure with PDB ID: 2byv.

1.4 Understanding of EPAC Allostery Prior to this Dissertation

In the absence of the endogenous allosteric effector, *i.e.* cAMP, EPAC exists in a predominantly inactive/auto-inhibited state. Upon cAMP binding, the auto-inhibitory interactions mediated by the CBD is removed. This allosteric regulation of EPAC by cAMP-dependent release of auto-inhibition is effectively modeled by a four state thermodynamic cycle resulting from the coupling of the cAMP-binding and the EPAC-activation equilibria (Figure 4).

The available structures suggests that in the absence of cAMP, the dominant species of EPAC share an overall closed and auto-inhibited conformation, whereby the CBD is situated in the proximity of the catalytic site (CDC25HD; Figure 4) (30, 35, 36). This closed topology is stabilized by a cluster of four conserved salt-bridges between helices α1 and α2 in the NTHB of the CBD and the CDC25HD (Figure 4). These salt bridges are collectively referred to as the ionic latch (IL).

Upon cAMP binding, two key regions have previously been shown to be affected and to promote the transition toward an active and open conformation: 1) a conserved structural element in the CBD called

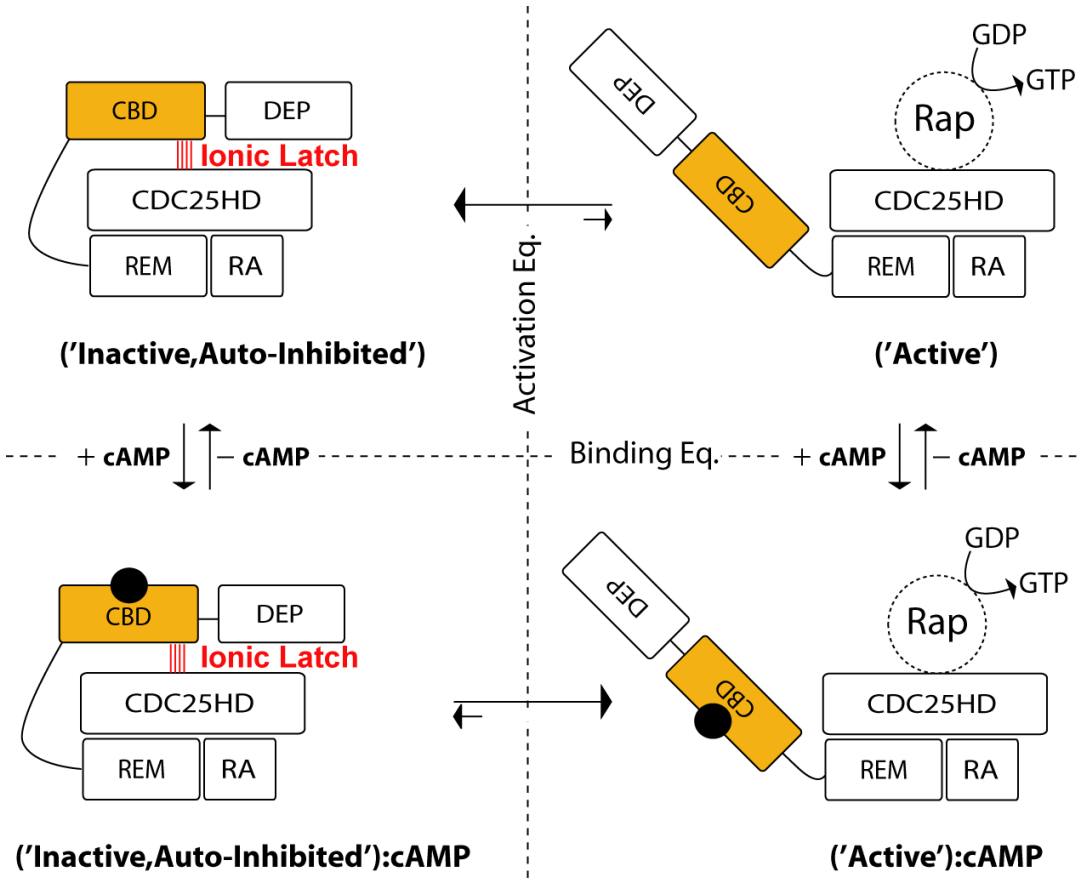

Figure 4: *Thermodynamic cycle of the EPAC.* In the absence of the cAMP, EPAC samples mostly the auto-inhibited and inactive state (top equilibrium). cAMP (black sphere) binding shifts the equilibrium, favorably toward the open and active conformation. Abbreviations are as in Figure 2a.

the α6 hinge helix is rotated towards the β-subdomain; and 2) the IL that bridges the CBD and catalytic site is disrupted (11, 30, 35, 36). These two cAMP-dependent events assist the transition to an open and active conformation, enabling the active site of the catalytic domain, CDC25HD, to be accessible by the

substrate protein Rap (Figure 4). This turn of events leads to up regulation of the GEF activity by EPAC (30, 35, 37).

1.4.1 The role of the conserved α6 hinge helix

Secondary chemical shifts from NMR indicate that upon cAMP binding, the α6 hinge helix becomes partially unstructured, suggesting that the rotation of the helix toward the β-subdomain is accompanied by a partial destabilization at the C-terminal end of the hinge helix. This has been confirmed by the crystal structure of holo-EPAC2 (35), illustrating a random-coil like preference for the last two turns of the hinge helix (30, 35, 36). Additionally, the cAMP analogue, Rp-cAMPS, an antagonist that binds the CBD, but fails to induce GEF activity, does not destabilize the C-terminal end of the hinge helix to the same extent as cAMP, even under saturating concentration of Rp-cAMPS (9, 11). Therefore the disordering of the hinge helix and the ensuing increase in flexibility is purely an allosteric effect induced by cAMP to relieve auto-inhibition. This is further supported by the enhanced dynamics observed in this region through NMR relaxation experiments for the backbone amides (9).

1.4.2 Dynamically Driven Control of the Ionic Latch (IL)

The importance of the IL for the control of EPAC function was illustrated for EPAC2 through a deletion mutation (EPAC2Δ306), which lacks one of the four salt bridges. The functional result of the mutation was a 5-fold increase in maximum GEF activity, suggesting that in the absence of cAMP, the IL promotes the inactive and auto-inhibited conformation (30).

Initial NMR studies on EPAC showed that upon cAMP binding, the region spanning the IL (α1 and α2 helices) was not subject to major structural variations as judged by chemical shift changes of the backbone amides (Figure 5) and by the secondary chemical shifts (9, 11, 38). The lack of major structural variation observed in solution was confirmed subsequently by the crystal structure of holo-EPAC2 (35) when compared to the apo-EPAC2 (30, 36).

However, the apo *vs*. holo structure comparisons by NMR chemical shifts and X-ray crystallography did not explain how cAMP allosterically modulates the IL. Since, no significant cAMP-dependent structural changes were observed for the IL, Das et al. hypothesized that the IL of the CBD was controlled by cAMP-dependent changes in dynamics and therefore employed NMR relaxation measurements to probe both

the ms-µs and ps-ns motions of the backbone amide. Comparison of the dynamic profiles between the apo and cAMP-bound forms showed an enhanced flexibility in the region spanning the IL. However, to ensure that the observed variation in the dynamics were caused by allosteric effects of cAMP and not simply due to binding, the dynamics of EPAC bound to an antagonist, Rp-cAMPS, was also assessed. The Rp-cAMPS antagonist which binds the CBD, but does not activate EPAC, failed to elicit a similar enhanced flexibility in the IL region, suggesting that cAMP controls the IL allosterically through dynamics. Therefore, it can be deduced that cAMP binding weakens the inhibitory IL interactions by increasing the conformational entropy in this region. That is, the entropic penalty imposed by cAMP on the IL promotes the transition to the active and open conformation of EPAC. Based on these findings, a refined model of EPAC allostery that encompasses also the dynamic features of the CBD was proposed (Figure 6) (9).

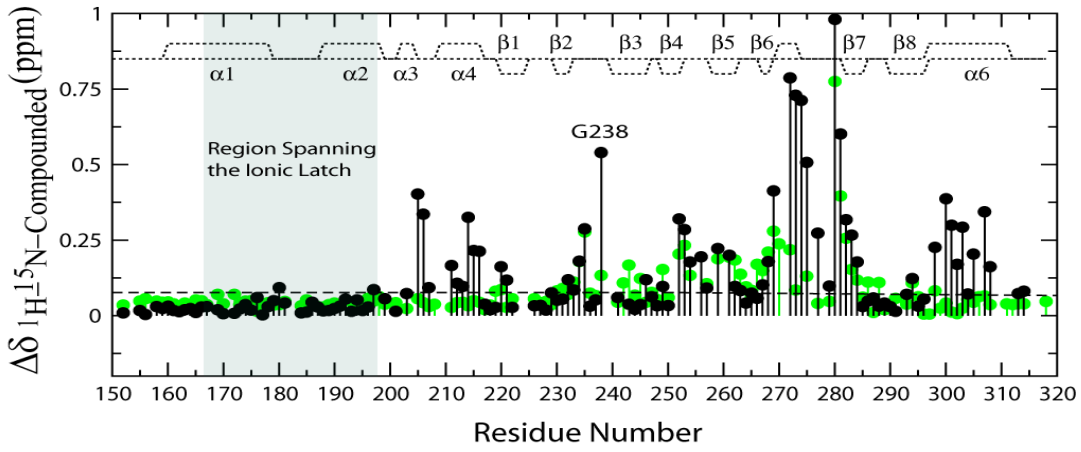

Figure 5: *The effect of cAMP on the IL.* Binding of cAMP (black) causes only minor chemical shift changes in the region spanning the IL as computed by the compounded chemical shifts for the backbone amides. Binding of Rp-cAMPS (green), an antagonist of cAMP, induces no major chemical shift changes in the α6 hinge helix. This figure was adapted from Das et al. (2008), ref. (9).

Interestingly, the dynamic enhancement of the IL was also not observed in the presence of cGMP, another second messenger molecule that binds EPAC, but does not induce activation, thereby being an antagonist, similarly to Rp-cAMPS (11). This selective modulation of dynamics in the IL rationalizes how the CBD of EPAC has evolved to garner an increased specificity for cAMP; in contrast to CBDs of other proteins that can be activated by both second messenger molecules (11, 39).

Although the importance of the IL dynamics in auto-inhibition emerged from the CBD fragment alone, the findings are likely to be even more relevant in the full-length EPAC where the IL interactions are further stabilized by the catalytic site in the apo inactive form (9). In support of this, dynamics in the context of

full-length EPAC2 has also been assessed by VanSchouwen et al. through NMR validated molecular dynamics (MD) simulations. Although MD simulations sample dynamics on a relatively short time scale (ps-ns), the results reaffirm the central role of dynamics in the IL of full-length EPAC, while implicating other conserved structural features of the CBD as also dynamic constituents of auto-inhibition (15).

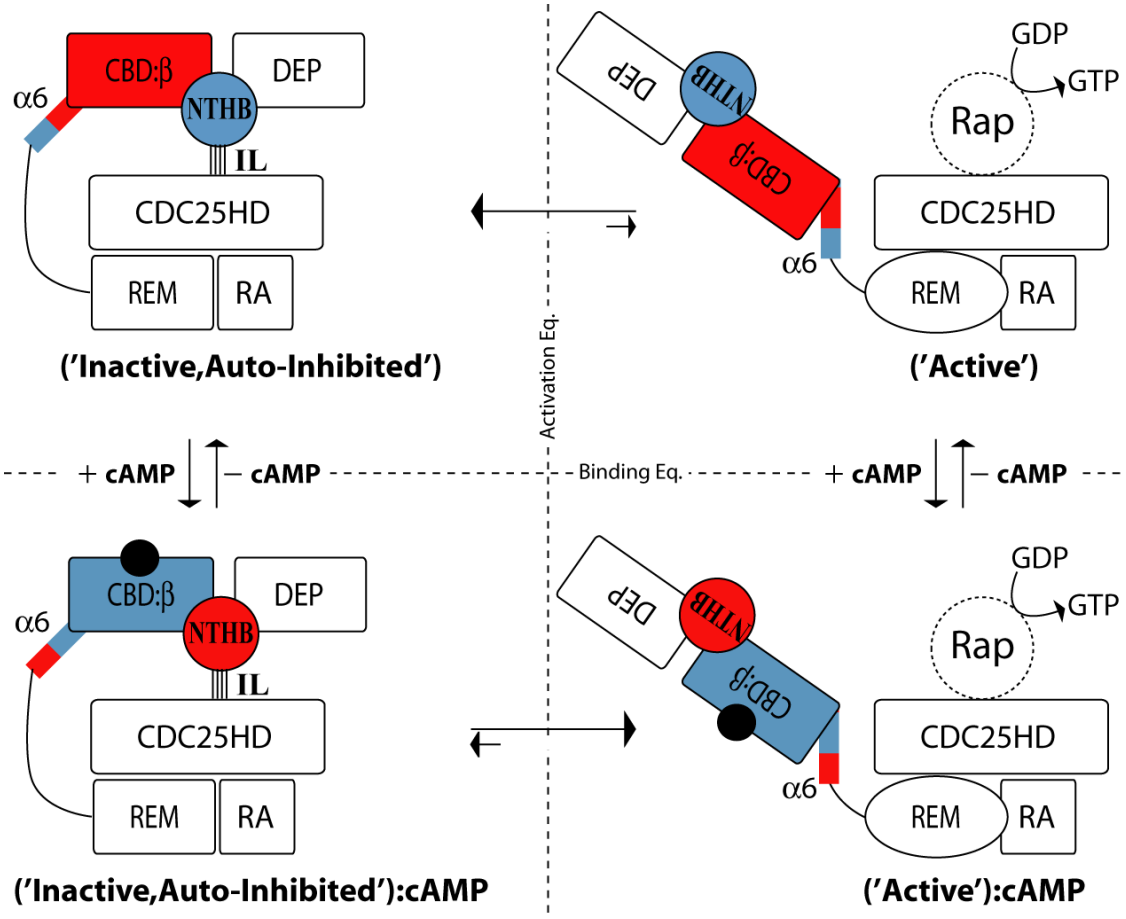

Figure 6: *Refined thermodynamic cycle of EPAC with incorporation of backbone dynamics.* Overall dynamics in the central core and in the β-subdomain (denoted as CBD:β) that are subject to quenching by cAMP (black sphere) are colored red in the apo-states (top auto-inhibitory equilibrium) and blue in the holo-states. Overall dynamics that are subject enhancement by cAMP, the C-terminal end of the hinge helix and the region spanning the IL, are colored as blue in the apo-states (top auto-inhibitory equilibrium) and red in the holo-states. This figure is adapted from Das et al. (2008), ref. (9).

1.4.3 The coupled role of the β2-β3 loop and PBC

A unique feature of the computational study utilizing MD simulations by VanSchouwen et al., was the ability to provide an initial insight into the experimentally unobservable transient states (*i.e.* cAMP-bound inactive state and apo yet active state in Figure 4). A comparative analysis of the dynamics of the

apo/inactive state to that of the transient apo/active states showed that an additional barrier toward allosteric activation in apo-EPAC is embedded in the dynamic nature of the flexible β2-β3 loop that is coupled to the PBC. That is, the MD simulations suggest that transition to the active state in the absence of cAMP involves a loss of conformational entropy not only in the PBC, but also in the β2-β3 loop (15). Additionally, NMR relaxation measurements have shown that binding of cAMP has an overall effect of reducing the flexibility of the distal β2-β3 loop, concomitant with quenching of dynamics in the PBC region (9–11).

The β2-β3 loop is a conserved structural motif (33) whose significance had been elusive until the advent of NMR studies on the first CBD of PKA (10, 40). Subsequent studies elaborated on the allosteric role of this region by comparing CBDs of diverse proteins from a wide array of species to illustrate that G238 within the β2-β3 loop region co-evolved with R279 of the PBC. These two highly conserved residues interact through the CH-π interaction between the guanidinium group of R279 and the Cα of the G238. The coupling between these two sites is also observed for the CBD-A of PKA, in which an overall loss of conformational entropy (*i.e*. rigidification) of the β2-β3 loop is necessary for inducing allosteric activation (10, 40, 41). Additionally, as in the case of PKA, the β2-β3 loop region, in particular G238 is a sensitive reporter of allosteric activation in EPAC, exhibiting pronounced chemical shifts changes in response to cAMP (Figure 5) and agonists alike (9, 10, 40).

1.5 Thesis Outlook and Key Questions Addressed by this Dissertation:

A common theme stemming from the previous sections is the functional role of dynamics in the conserved structural elements of the CBD in EPAC. Upon cAMP-induced allosteric activation, the C-terminal end of the α6 hinge helix and the IL of the NTHB exhibits enhanced dynamics as required for the closed-to-open transition, while the β2-β3 loop and the PBC regions become rigid (9, 11, 35, 38). Among these cAMP-dependent modulations of EPAC dynamics, the increased IL entropy upon activation is the longest-range effect observed for the CBD. However, it remains unclear how the signal originating at the site of cAMP-binding is transmitted to the distant IL site that lacks major structural changes (9, 11). Therefore one of the central questions addressed by this thesis is: What is the intra-molecular network of allosteric residues that bridges the sites affected by macroscopic cAMP-dependent structural variations, such as the hinge-helix, to the distant loci exhibiting cAMP-dependent modulations of dynamics, such as the IL? To address this question, Chapter two describes a method to identify the allosteric networks embedded in the CBD of EPAC by employing the covariance analysis of NMR chemical shifts, which are

sensitive probes of allosteric modulations. We call this method the chemical shift covariance analysis (CHESCA).

The main premise of CHESCA is that highly coupled residues that define an allosteric network exhibit concerted and correlated chemical shift movements for a given set of perturbations. Consequently, such correlations can be between residues with major chemical shift changes, reflecting well resolved structural variations, and residues subject to only minor chemical shift variations, reflecting more subtle alterations in the conformational ensemble, as in the case of the IL. The results of CHESCA suggest that the cAMP signal initiated at the site of binding are propagated through a network of contiguous residues situated in the flexible loops and helices of the α-subdomain that ultimately reach the IL, while leaving the β-sheets largely unaffected. Further details and background on the CHESCA method, including the theoretical link between the CHESCA approach and the NMR theory of fast chemical exchange, are available in the supplementary material for Chapter two.

A challenge in the implementation of the CHESCA approach is in the selection of an appropriate set of informative allosteric perturbations that simultaneously meet the following criteria: 1) induce different degrees of activation with minimal covalent modifications; 2) are spatially well co-localized within a single region of the protein structure. For EPAC, these criteria were fulfilled by a carefully selected set of cAMP analogs: Rp-cAMPS, Sp-cAMPS, and 2'-OMe-cAMP. Alternatively, criteria one and two can be met through mutations used as allosteric perturbations, which is often a more feasible strategy. However, to ensure that the perturbations caused by the mutations have a significant effect on the auto-inhibitory equilibrium (Figure 4, top equilibrium) requires a systematic pre-screening and assessment of the mutation-induced chemical shift changes. In other words, a quantitative analysis is required to gauge the extent of activation or inactivation achieved by a given mutation prior to a full CHESCA implementation. Chapter 3 addressed this pre-screening need and illustrates how a simple vectorial analysis, called chemical shift projection analysis (CHESPA), efficiently maps the effect of selected mutations on the auto-inhibitory equilibrium of EPAC. For example, the effect of mutations in flexible regions were assessed, since MD simulations suggested that in the absence of cAMP, mutations that reduce the flexibility in the β2-β3 loop and PBC regions sample more frequently the active state (15). In addition, since the β2-β3 loop is also part of the allosteric network proposed in Chapter 2, the mutations employed in Chapter 3 serve also as a validation of the signalling pathway identified by CHESCA.

The functional role of flexible regions is further probed in Chapter 4 with regard to the C-terminal end of the α6 hinge helix. The allosteric network identified by CHESCA also implicated a sub-cluster of residues

that form hydrophobic spines involving the C-terminal end of the hinge helix, which becomes flexible in the ps-ns time-scale upon cAMP-binding (9, 12). Since NMR studies have shown that allosteric activation by cAMP causes partial unfolding of the hinge helix (9, 11), we argued that mutations that destabilize the hinge helix could to some extent mimic the effect of cAMP in the apo-state. To test this hypothesis, Chapter 4 employs the CHESPA approach to assess how deletion mutations that induce hinge helix destabilizati(42)on, affect the auto-inhibitory equilibrium of the EPAC CBD. Interestingly, it was found that a deletion mutation without the destabilized end, bound cAMP more tightly. Additionally, the hinge helix localized mutations in the apo-state were used as a perturbation set for CHESCA to probe the allosteric role of the hinge helix in the absence of cAMP. The result of employing both chemical shift analysis methods, suggest that the integral hinge helix stability underlies auto-inhibition, at least partially and is tightly coupled to the other allosteric hot-spots mediating inhibitory interactions, even in the absence of cAMP (42).

Lastly, Chapter five of the thesis recapitulates the significance of the chemical shift analysis approach in providing a link between the "allosteric hotspots" within the CBD of EPAC and a brief discussion of alternative methods that have attempted to address similar questions in other systems. The final chapter also discusses possible extensions of the methods and application to other systems in the context of future directions.

1.6 References

1. Bos JL, Rehmann H, Wittinghofer A (2007) GEFs and GAPs: critical elements in the control of small G proteins. *Cell* 129:865–77.

2. Bos JL, De Rooij J, Reedquist K a (2001) Rap1 signalling: adhering to new models. *Nature reviews Molecular cell biology* 2:369–77.

3. Del Sol A, Tsai C-J, Ma B, Nussinov R (2009) The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure (London, England : 1993)* 17:1042–50.

4. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins* 57:433–43.

5. Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology* 5:789–96.

6. Changeux J-P, Edelstein SJ (2005) Allosteric mechanisms of signal transduction. *Science (New York, NY)* 308:1424–8.

7. Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. *Nature chemical biology* 4:474–82.

8. Volkman BF, Lipson D, Wemmer DE, Kern D (2001) Two-state allosteric behavior in a single-domain signaling protein. *Science (New York, NY)* 291:2429–33.

9. Das R et al. (2008) Entropy-driven cAMP-dependent allosteric control of inhibitory interactions in exchange proteins directly activated by cAMP. *The Journal of biological chemistry* 283:19691–703.

10. Das R et al. (2007) cAMP activation of PKA defines an ancient signaling mechanism. *Proceedings of the National Academy of Sciences of the United States of America* 104:93–8.

11. Das R et al. (2009) Dynamically driven ligand selectivity in cyclic nucleotide binding domains. *The Journal of biological chemistry* 284:23682–96.

12. Selvaratnam R, Chowdhury S, VanSchouwen B, Melacini G (2011) Mapping allostery through the covariance analysis of NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America* 108:6133–8.

13. Tsai C-J, Del Sol A, Nussinov R (2009) Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Molecular bioSystems* 5:207–16.

14. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein science : a publication of the Protein Society* 9:10–9.

15. VanSchouwen B, Selvaratnam R, Fogolari F, Melacini G (2011) Role of dynamics in the autoinhibition and activation of the exchange protein directly activated by cyclic AMP (EPAC). *The Journal of biological chemistry* 286:42655–69.

16. Christopoulos A (2002) Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nature reviews Drug discovery* 1:198–210.

17. Hardy J a, Wells J a (2004) Searching for new allosteric sites in enzymes. *Current opinion in structural biology* 14:706–15.

18. Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. *Nature* 450:964–72.

19. Kern D, Zuiderweg ER (2003) The role of dynamics in allosteric regulation. *Current Opinion in Structural Biology* 13:748–757.

20. Mittermaier A, Kay LE (2006) New tools provide new insights in NMR studies of protein dynamics. *Science (New York, NY)* 312:224–8.

21. Popovych N, Sun S, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. *Nature structural & molecular biology* 13:831–8.

22. Hamm HE (1998) The many faces of G protein signaling. *The Journal of biological chemistry* 273:669–72.

23. Rehmann H, Wittinghofer A, Bos JL (2007) Capturing cyclic nucleotides in action: snapshots from crystallographic studies. *Nature reviews Molecular cell biology* 8:63–73.

24. De Rooij J et al. (1998) Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature* 396:474–7.

25. Kawasaki H (1998) A Family of cAMP-Binding Proteins That Directly Activate Rap1. *Science* 282:2275–2279.

26. Grandoch M, Roscioni SS, Schmidt M (2010) The role of Epac proteins, novel cAMP mediators, in the regulation of immune, lung and neuronal function. *British journal of pharmacology* 159:265–84.

27. Shirshev S V (2011) Role of Epac proteins in mechanisms of cAMP-dependent immunoregulation. *Biochemistry Biokhimiia* 76:981–98.

28. De Rooij J et al. (2000) Mechanism of regulation of the Epac family of cAMP-dependent RapGEFs. *The Journal of biological chemistry* 275:20829–36.

29. Yang S et al. (2013) ROLE OF GUANINE-NUCLEOTIDE EXCHANGE FACTOR EPAC IN RENAL PHYSIOLOGY AND PATHOPHYSIOLOGY. *American journal of physiology Renal physiology*.

30. Rehmann H, Das J, Knipscheer P, Wittinghofer A, Bos JL (2006) Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state. *Nature* 439:625–8.

31. Poppe H et al. (2008) Cyclic nucleotide analogs as probes of signaling pathways. *Nature methods* 5:277–8.

32. Su Y et al. (1995) Regulatory subunit of protein kinase A: structure of deletion mutant with cAMP binding domains. *Science (New York, NY)* 269:807–13.

33. Kornev AP, Taylor SS, Ten Eyck LF (2008) A generalized allosteric mechanism for cis-regulated cyclic nucleotide binding domains. *PLoS computational biology* 4:e1000056.

34. Kannan N et al. (2007) Evolution of allostery in the cyclic nucleotide binding module. *Genome biology* 8:R264.

35. Rehmann H et al. (2008) Structure of Epac2 in complex with a cyclic AMP analogue and RAP1B. *Nature* 455:124–7.

36. Rehmann H et al. (2003) Structure and regulation of the cAMP-binding domains of Epac2. *Nature structural biology* 10:26–32.

37. Rehmann H, Rueppel A, Bos JL, Wittinghofer A (2003) Communication between the regulatory and the catalytic region of the cAMP-responsive guanine nucleotide exchange factor Epac. *The Journal of biological chemistry* 278:23508–14.

38. Mazhab-Jafari M (2007) Understanding cAMP-dependent allostery by NMR spectroscopy: comparative analysis of the EPAC1 cAMP-binding domain in its apo and cAMP-bound states. *Journal of the …* 129:14482–92.

39. Zagotta W, Olivier N, Black K, Young E (2003) Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* 425:200–205.

40. Das R, Melacini G (2007) A model for agonism and antagonism in an ancient and ubiquitous cAMP-binding domain. *The Journal of biological chemistry* 282:581–93.

41. Abu-Abed M, Das R, Wang L, Melacini G (2007) Definition of an electrostatic relay switch critical for the cAMP-dependent activation of protein kinase A as revealed by the D170A mutant of RIalpha. *Proteins* 69:112–24.

42. Selvaratnam R, Mazhab-Jafari MT, Das R, Melacini G (2012) The auto-inhibitory role of the EPAC hinge helix as mapped by NMR. *PloS one* 7:e48707.

Chapter 2

Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts

2.1 Author's Preface

The work presented in this chapter has previously been published and is reproduced here with permission from the Proceedings of the National Academy of Sciences USA. Full citation is as follows:

Selvaratnam, R., Chowdhury, S., VanSchouwen, B. & Melacini, G. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc. Natl. Acad. of Sci. USA* 108, 6133–8 (2011).

I conducted most of the experiments necessary for the manuscript. Dr. Somenath Chowdury assigned the backbone amides of wild type EPAC bound to Sp-cAMPS, Rp-cAMPS, and 2'-OMe-cAMP. Bryan VanSchouwen replicated the experiment under different buffer conditions. I co-wrote the manuscript with Drs. Somenath Chowdury and Giussepe Melacini.

2.2 Abstract

Allostery is a fundamental mechanism of regulation in biology. The residues at the end points of long-range allosteric perturbations are commonly identified by the comparative analyses of structures and dynamics in apo- and effector-bound states. However, the networks of interactions mediating the propagation of allosteric signals between the end points often remain elusive. Here we show that the covariance analysis of NMR chemical shift changes caused by a set of covalently modified analogs of the allosteric effector (*i.e.* agonists and antagonists) reveals extended networks of coupled residues. Unexpectedly, such networks reach not only sites subject to effector-dependent structural variations, but also regions that are controlled by dynamically driven allostery. In these regions the allosteric signal is propagated mainly by dynamic rather than structural modulations, which result in subtle, but highly correlated chemical shift variations. The proposed chemical shift covariance analysis (CHESCA) identifies inter-residue correlations based on the combination of agglomerative clustering (AC) and singular value decomposition (SVD). AC results in dendrograms that define functional clusters of coupled residues, while SVD generates score plots that provide a residue-specific dissection of the contributions to binding and allostery. The CHESCA approach was validated by applying it to the cAMP-binding domain of the exchange protein directly activated by cAMP (EPAC) and the CHESCA results are in full agreement with independent mutational data on EPAC activation. Overall, CHESCA is a generally applicable method that utilizes a selected chemical library of effector analogs to quantitatively decode the binding and allosteric information content embedded in chemical shift changes.

2.3 Introduction

Long-range allosteric perturbations are propagated not only by structural changes but also by effector-dependent modulations in dynamics (1-23). The end points of these long-range allosteric signal propagations are effectively characterized by the comparative analysis of the structural and dynamic profiles of apo- and effector-bound states (2,7). However, what remains experimentally challenging is often defining the networks of residues that mediate the cross-talk between distal sites. Such clusters of coupled residues are particularly elusive in allosteric processes with a significant dynamically driven component (11-17), as in this case the allosteric signal propagation relies on subtle, but critical, conformational and side chain packing rearrangements that often fall below the resolution of common X-ray or NMR structure determination methods (2,7,24).

Here we introduce a general experimental method to map allosteric networks based on the covariance analysis of NMR chemical shifts. The chemical shift covariance analysis (or CHESCA) is based on two simple but general notions. The first assumption is that the subtle but functionally relevant structural changes that underlie the allosteric modulations of dynamics are effectively probed by accurately measured NMR chemical shift variations, even when they escape detection through traditional structure determination methods (24-27). The second general assumption of the CHESCA approach is that when a system is subject to a set of perturbations, residues that belong to the same effector-dependent allosteric network exhibit a concerted response to the perturbation set. This collective response is effectively sensed through correlations between the chemical shift variations experienced by different residues as a result of the perturbations. Such perturbations may arise from mutations or from chemical modification of the ligand effector. In either case, the perturbation set includes both active and inactive states, enabling the identification of networks of correlated residues linked to activation.

The CHESCA method is illustrated and validated here through its application to the multi-domain signaling protein EPAC (Fig. 1a) (28). EPAC is a guanine exchange factor controlled by the second messenger cAMP and epitomizes signaling proteins that function as molecular switches and signal transducers (28). In such systems, the ligand-dependent activation is often adequately explained by the coupling of binding and activation equilibria, resulting in the four-state thermodynamic cycle illustrated in Fig. 1b (29). According to this model, in the apo-state the activation equilibrium is shifted toward inactive conformations, while in the effector-bound (holo) state the activation equilibrium is shifted toward the active conformations (29) (Fig. 1b). The structures of EPAC in both apo-and holo-states have been recently solved (30, 31), revealing that the main difference between the inactive and active

conformations is in their overall topology. The inactive state is defined by a 'closed' topology, whereby the regulatory region of EPAC (Fig. 1a) sterically occludes access of substrate Rap proteins to the catalytic site (Fig. 1b) (30). In the absence of cAMP, the steric occlusion in the inactive state is secured by a cluster of salt bridges between the catalytic segment of EPAC and the N-terminal α1,2 helices, which are part of the EPAC cAMP-binding domain (CBD) (Fig. 1b) (30). This inhibitory salt bridge cluster is commonly referred to as the 'ionic latch' (IL) (Fig. 1b).

Upon cAMP binding, no significant changes in local structure are observed for the IL region as assessed based on the root mean square deviation (RMSD) values shown in Fig. 1c (red region) (30,31). However, cAMP enhances the ps-ns and ms-µs dynamics in the IL zone (red dots in Fig. 1c), increasing the entropic penalty for the IL and thus contributing to the weakening of the inhibitory IL interactions between the regulatory and the catalytic regions (15,16). The dynamically-promoted release of the IL induced by cAMP assists the transition to 'open' active structures (Fig. 1b), in which a hinge rotation of a helix C-terminal to the CBD (α6, Fig. 1c) moves the regulatory region away from the catalytic region, providing Rap unhindered access to the catalytic site of EPAC (Fig. 1b) (15, 31,32).

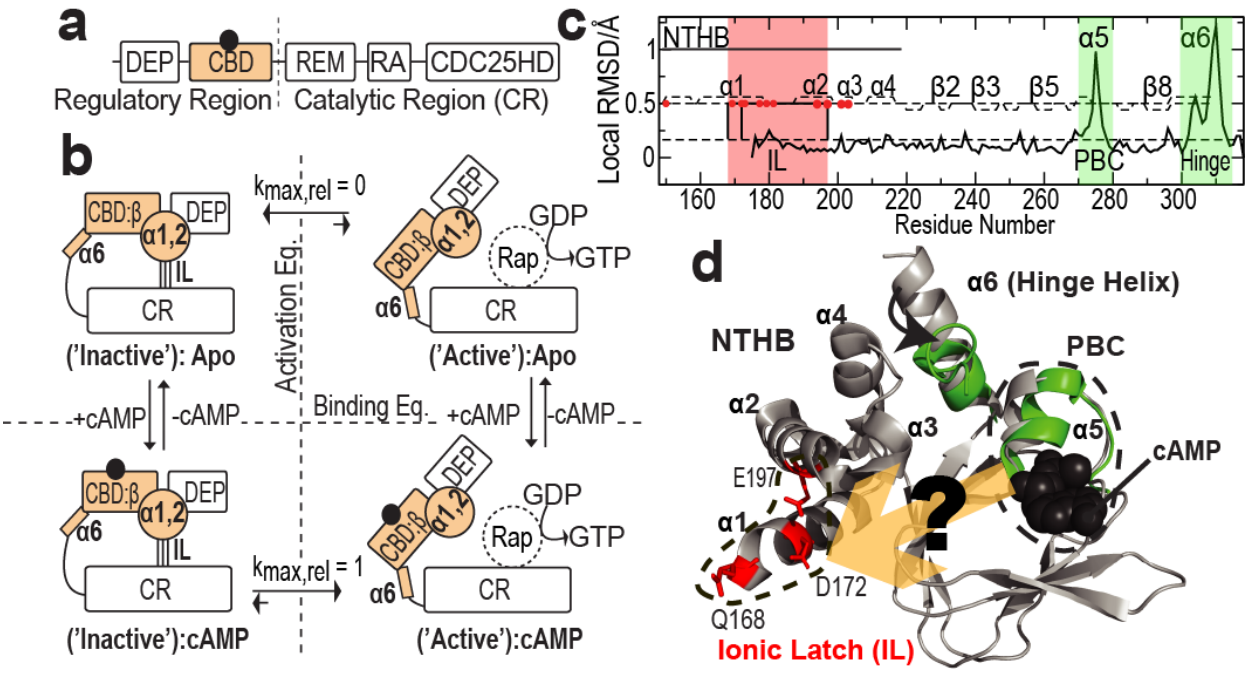

Figure 1: **(a)** Domain and region organization of EPAC1. The cAMP-binding domain (CBD) analyzed by NMR is highlighted in orange; **(b)** Allosteric model of EPAC regulation based on coupled activation and cAMP-binding equilibria. CBD:β denotes the β-subdomain and IL stands for Ionic Latch. **(c)** Local RMSD between the structures of the EPAC CBD in the apo-and holo-states (PDB ID: 2BYV and 3CF6, respectively). The secondary structure is reported with dashed lines. The regions undergoing significant conformational changes (i.e. PBC and hinge helix) are highlighted in green, while the IL zone is marked in red. The IL residues are indicated by vertical bars. Residues in

the N-terminal helical bundle (NTHB) that display enhanced dynamics in the ps-ns and/or ms-µs time-scale upon cAMP binding are indicated by red dots. PBC denotes the cAMP phosphate binding cassette. **(d)** Ribbon diagrams of the superimposed apo (gray) and holo (green) EPAC CBD structures (PDB ID as in panel c). The holo ribbon is shown only for the regions where it differs significantly from the apo-state as judged based on the local RMSD reported in panel (c). cAMP is shown as black spheres. The IL residues are shown in red. The dashed contours define the PBC and IL regions.

The enhancement of IL dynamics linked to cAMP-binding provides an initial explanation for how the inhibitory IL is remotely controlled by cAMP in the framework of dynamically driven allostery, but it is still unclear how the cAMP signal propagates from the cAMP phosphate binding cassette (PBC) and the adjacent α6 hinge helix (green regions in Fig. 1d) to the distal IL region, where dynamics is enhanced (red sites in Fig. 1d). Here we address this question by applying the CHESCA method, which results in a PBC-to-IL signaling pathway independently supported through mutagenesis (33) and conserved co-evolutionary patterns reported for known CBDs (34). The application of the chemical shift covariance approach to EPAC illustrates therefore the effectiveness of the CHESCA principle for defining intra-molecular allosteric networks that control signal transduction.

2.4 Materials and Methods

Sample Preparation: Isotopically labeled samples were prepared as explained in Supplementary Information (page 41).

NMR Spectroscopy: All spectra were acquired at 306 K with a Bruker Avance 700-MHz spectrometer equipped with a 5 mm TCI cryoprobe. Ligand titrations where monitored through [^{15}N-^{1}H] HSQC spectra. The ^{1}H and ^{15}N chemical shifts employed for CHESCA were referenced relative to ^{15}N-Ac-Gly and measured using HSQC spectra acquired at saturating ligand concentrations. Ligand saturation is critical because the four ligands selected as perturbations (Fig. 2a) span a wide range of affinities with K_D values varying up to two orders of magnitude (Table S1). This means that when non-saturating concentrations of ligands are used, the differences between the different ligand-bound states (Fig. 2a) reflect not only the different degrees of activation, as desired for the purpose of mapping allosteric networks, but also the varying extents of binding. To avoid this bias, it is important to ensure that saturating ligand concentrations are reached, as verified by the binding isotherms displaying a clear dose-response pattern (Fig. 2c; Fig. S1). The plateau region of the binding isotherms (Fig. 2c; Fig. S1) defines the ligand concentration range recommended for CHESCA. The ^{1}H and ^{15}N chemical shifts used as input for CHESCA are the average of

the values measured through Gaussian peak fitting for the last three points in the titration (*i.e.* at ligand saturation). The standard deviation for these three measured chemical shift values at saturation is assumed to represent the error in the chemical shifts for the bound states. For the apo-state, the error of the chemical shifts was estimated based on the differences between the ppm values measured for two different apo samples. The measurement of minor chemical shift changes is further detailed in SI Text.

Chemical Shift Analysis: The protocols for the AC and SVD analyses are available as SI Text and Figure S2.

2.5 Results and Discussion

2.5.1 Selection of Perturbations

The first step in the implementation of the chemical shift covariance method is a careful choice of the states and of the associated perturbations used to identify correlations between residues. In general, a primary criterion for the selection of perturbations useful in mapping networks of coupled residues that control biological function, such as enzyme activity, is the inclusion of both active and inactive states that sample a diverse activation spectrum. In the case of the EPAC CBD, such perturbations are provided by the binding of the endogenous activator (*i.e.* cAMP, Fig. 2a) and of three related analogs, the 2'-OMe-cAMP and the diastereoisomeric phosphorothioate cAMP analogs Rp- and Sp-cAMPS (Fig. 2a). The covalent modifications of these three cAMP surrogates perturb key hydrogen-bonds that anchor cAMP to the phosphate binding cassette (Fig. 2b) and result in different degrees of EPAC activation, as indicated by the wide range of relative k_{max} values spanned by the selected cAMP analogs (Fig. 2a) (33). The relative k_{max} value quantifies the guanine exchange activity of EPAC at saturating concentrations of effector ligand and is a useful indicator of the position of the activation equilibrium (Fig. 1b) (33). In the inactive states, such as the apo or the antagonist Rp-cAMPS-bound states, $k_{max,relative} = 0$, corresponding to at least partial guanine nucleotide exchange inhibition, while at saturating concentrations of the cAMP activator $k_{max,relative} = 1$ (Fig. 2a) (33). However, even at saturating concentrations, cAMP does not fully shift the activation equilibrium to the active state and therefore other ligands with enhanced active *vs.* inactive state selectivity cause super-activation ($k_{max,relative} > 1$) by further increasing the population of the active state. This is the case for the super-agonists Sp-cAMPS and 2'-OMe-cAMP that display a $k_{max,relative} > 1$ (Fig. 2a) (33). This wide $k_{max,relative}$ range from 0 to > 1 indicates that the chosen perturbations are able to

modulate the populations of inactive and active conformations, locking the activation equilibrium at different positions and making the chosen ligands an effective tool to map functional allosteric networks. In summary, the five states used for CHESCA are the apo- and the four holo-states (Fig. 2a), all at saturating concentrations (Fig. 2c, Fig. S1). The apo- and Rp-cAMPS-bound states are inactive, while the other three states are active.

2.5.2 The Chemical Shift Covariance Analysis (CHESCA)

Once the ^{1}H-^{15}N HSQC spectra of the five selected states are fully assigned, pair-wise correlations between chemical shift variations experienced by different residues are analyzed for the purpose of identifying networks of coupled residues. A simple proof of the linear correlations between chemical shift variations is obtained when the allosteric activation is adequately described by a single two-state and fast exchanging equilibrium (Fig. 2e; SI Text). In this respect it should be noted that the inactive/active equilibrium can still be and often is in the fast exchange regime even when the apo/holo exchange is slow in the chemical shift time scale. According to the two-state activation model in the fast exchange regime, the chemical shifts measured for residues far removed from the effector binding site are modeled simply as population weighted linear averages of the chemical shifts in the active and inactive states, implying that the chemical shifts observed for different residues sensing the same inactive/active equilibrium are linearly correlated.

The inter-residue linear chemical shift correlation can be proven in more formal terms (SI Text) if we denote as $\delta_{i,Ac}$ (and $\delta_{j,Ac}$), $\delta_{i,In}$ (and $\delta_{j,In}$) and δ_{ik} (and δ_{jk}) the combined chemical shifts of residue *i* (and *j*) in the active, inactive and k^{th} perturbed states, respectively. The combined chemical shift typically refers to weighted contributions from the ppm values of the amide proton and nitrogen (*i.e.* $\delta_{ik} = w^N \delta^N_{ik} + w^H \delta^H_{ik}$, with w^N = 0.2 and w^H = 1) (35). If amino acids *i* and *j* are sufficiently distant from the sites of the perturbations, their chemical shifts sense exclusively the perturbation-linked changes in the active/inactive conformational equilibrium. In this case, δ_{ik} and δ_{jk} are related through the linear equation (SI Text):

$$\delta_{ik} = \delta_{jk}\,\alpha + \beta \qquad (1)$$

where $\alpha = \Delta\delta_i/\Delta\delta_j$, $\Delta\delta_i = (\delta_{i,Ac} - \delta_{i,In})$, $\Delta\delta_j = (\delta_{j,Ac} - \delta_{j,In})$ and $\beta = \delta_{i,In} - \alpha\delta_{j,In}$.

Equation (1) has two key implications. First, it should be noted that equation (1) is valid irrespective of the size of $\Delta\delta_i$ and $\Delta\delta_j$, indicating that, if residues *i* and *j* belong to the same allosteric network, their perturbation-dependent chemical shift variations are expected to be highly linearly correlated regardless of their magnitude. This observation suggests that even minor ppm variations are potentially significant for the definition of allosteric networks, if measured accurately and precisely (SI Text). Second, since equation (1) was derived based on the assumption that the chemical shifts observed in the different perturbed states are a weighted average of the ppm values in the active and inactive conformations (equations SI1, SI2), it is expected that when a network of residues is functionally related to allostery, the chemical shifts of the active states (*e.g.* agonist bound) should be well separated from those of the inactive states (*e.g.* apo and antagonist bound), *i.e.* the active and inactive states should cluster into two distinct groups in the chemical shift correlation plots, as shown for the representative correlations of Fig. 2f-h. Taken together, these two implications of equation (1), *i.e.* the high degree of correlation between the δ_{ik} and δ_{jk} chemical shifts, independently of the magnitudes of $\Delta\delta_i$ and of $\Delta\delta_j$, and the separate clustering of active *vs.* inactive states, define the foundation of the CHESCA approach and provide two simple, but effective, criteria for the identification of allosteric networks. These considerations are not necessarily limited to a two-state model, because in the presence of multi-state equilibria the inter-residue chemical shift correlations are still linear when the relative chemical shift changes experienced by the two correlated residues form isomorphic patterns (SI Text; Fig. 2d).

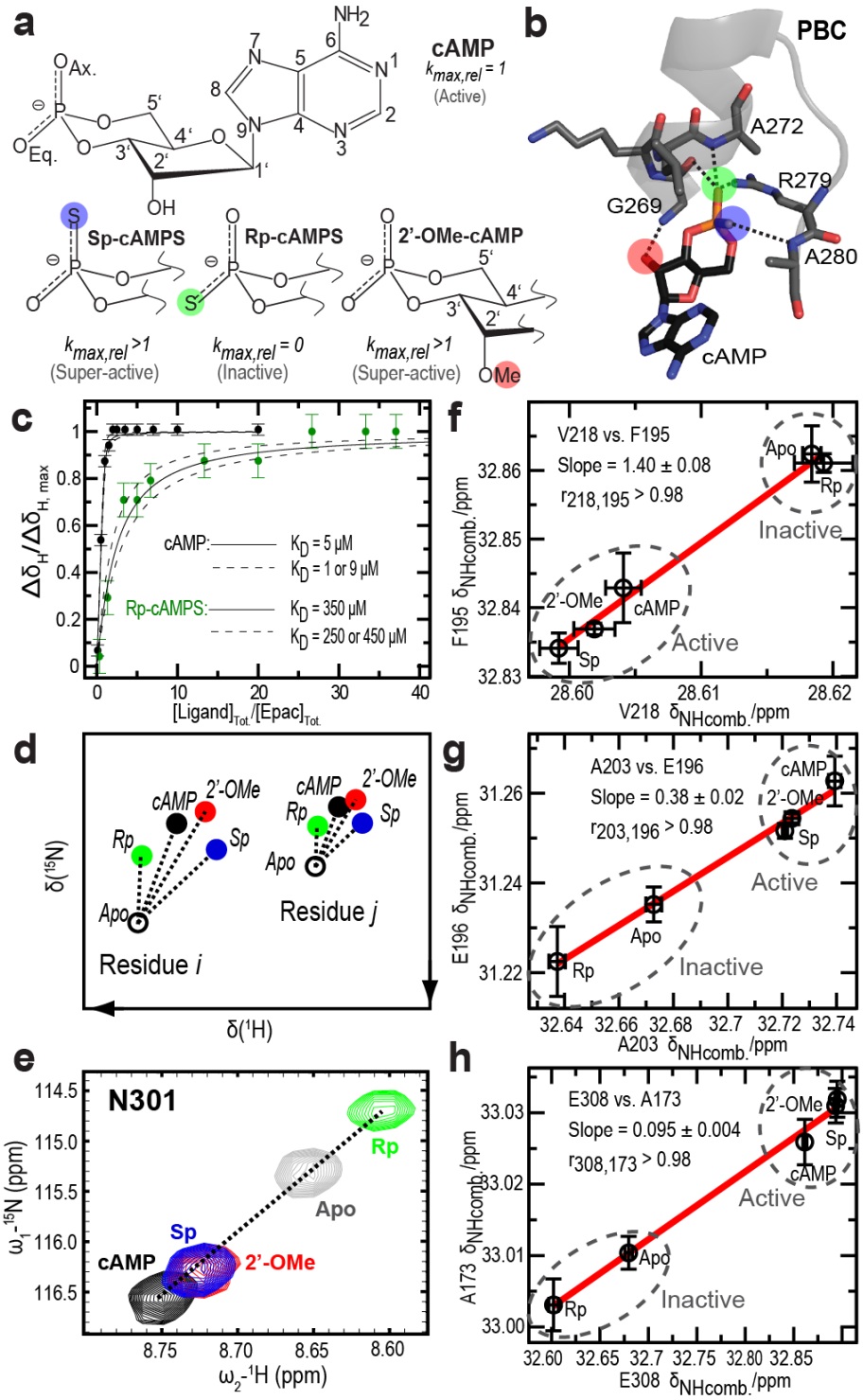

Figure 2: (a) Covalent structure and related modifications of cAMP utilized to modulate the activation equilibrium of EPAC. For each ligand the relative k_{max} value is reported ($k_{max,rel}$). **(b)** Key residues and hydrogen-bonds that anchor cAMP to the PBC. cAMP sites that are modified in the cAMP-analogs used in this work are highlighted with color circles (color coding as in (a)). **(c)** Binding isotherms for high and low affinity ligands measured though relative chemical shift changes of L207. **(d)** Hypothetical scheme showing correlated chemical shift variations for two

different residues *i* and *j*. For residues subject to a two-state active/inactive fast exchange, a linear pattern is often observed **(e)**. **(f-h)** Representative inter-residue combined chemical shift correlations (SI Text).

2.5.3 Identification of Networks

In order to facilitate the systematic implementation of the two equation (1)-based criteria for allosteric networks, the combined $^{1}H_N$ and $^{15}N_H$ chemical shifts of the EPAC CBD construct EPAC1$_h$ (149-318) in the apo- and in the four holo-states (*i.e.* under saturating conditions of Rp-cAMPS, cAMP, Sp-cAMPS and 2'-OMe-cAMP) were compiled into an n_r x 5 data matrix **M**, where n_r is the number of residues for which assignments are available in all five states (Fig. S2). The correlation matrix (**R**) of **M** transpose was then computed, whereby the r_{ij} elements of **R** represent the Pearson's correlation coefficients between residues *i* and *j* (SI Text; equation *SI9*; Fig. S2). The off-diagonal elements of **R** corresponding to $|r_{ij}|$ values ≥ 0.98 are displayed in Fig. 3a and three representative correlations are shown in Fig. 2f-h. While for all three correlations of Fig. 2f-h $r_{ij} > 0.98$, their slopes vary by more than one-order of magnitude (Fig. 2f-h), confirming that the chemical shift variations of two residues can be highly correlated even when their magnitudes are markedly different, as anticipated based on equation (1). This observation explains why in the **R** matrix of Fig. 3a several cross-peaks connect regions undergoing changes in conformation, and subject to major chemical shift variations (green highlights in Fig. 1c and Fig. 3a,f), to regions undergoing changes in local dynamics rather than in local structure, and subject only to minor chemical shift variations such as the IL zone (red highlights in Fig. 1c and Fig. 3a,f). The **R** 'cross-peaks' of Fig. 3a provide therefore an initial appreciation of the potential of chemical shift correlations as a reliable tool to map the interaction networks underlying allosteric processes, including dynamically driven allostery. However, based on simple visual inspection of **R** (Fig. 3a) it is difficult to reliably and systematically identify such allosteric networks. For this purpose it is

necessary to analyze the **R** matrix using agglomerative clustering (AC) methods, which are aimed at effectively identifying groups of coupled residues.

2.5.4 Agglomerative Clustering (AC)

AC is a method commonly used to identify gene networks from microarray data and it can be generally applied to identify clusters based on correlation matrices, such as **R** (Fig. S2) (36). Specifically, AC assigns

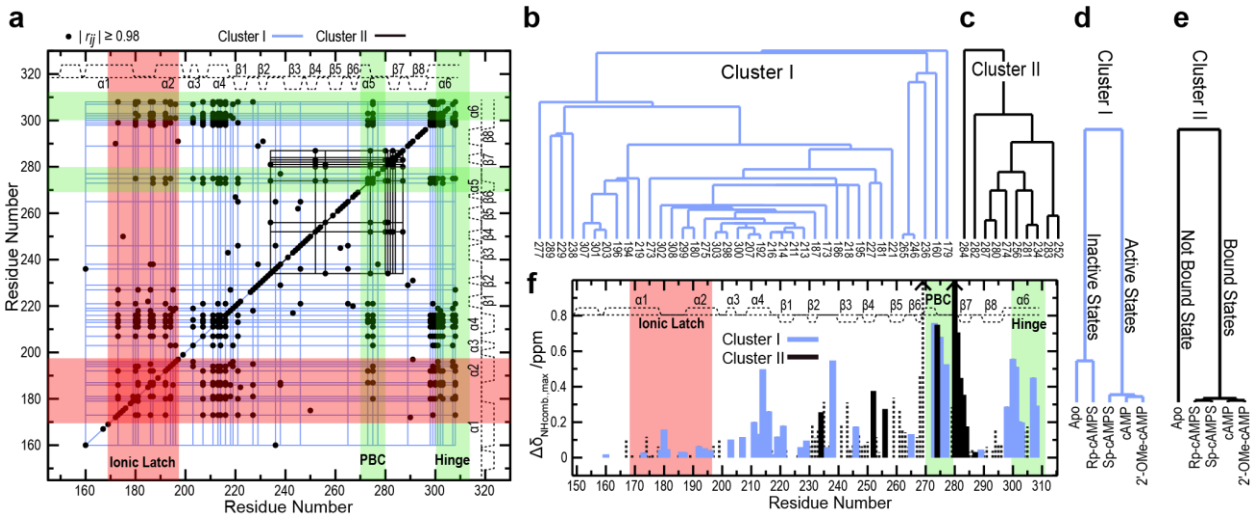

Figure 3: (a) Chemical shift correlation matrix for the EPAC CBD. If the absolute value of the Pearson's correlation coefficient between a pair of residues is ≥ 0.98 the corresponding cross-peak is marked with a dot. The two clusters of residues identified through the agglomerative algorithm illustrated in panels (b, c) are marked with blue and black lines, respectively. The secondary structure is reported as in Figure 1c. **(b, c)** Dendrograms for clusters I and II, respectively. Residues linked through the dendrograms are reported at the bottom of each tree. All nodes shown correspond to correlation coefficients with absolute values ≥ 0.98. **(d)** Dendrogram for the agglomerative clustering of the 5 states (apo and four holo) obtained using the chemical shift sub-matrix containing only the residues of cluster I. **(e)** As panel (d) but for cluster II. **(f)** For each residue the maximum variation in $\delta_{NHcomb.}$ among the five states is reported as absolute value. Residues belonging to clusters I and II are reported as blue and black bars, respectively. Residues that do not belong to clusters I or II are indicated by dotted lines. Residues for which no line is reported are those for which unambiguous assignments in all five states are unavailable. The ppm changes for G269 and A280 are off-scale as indicated by the arrows.

a first intra-cluster link to the residue pair with the highest absolute value of the correlation-coefficient within **R**. A second intra-cluster link is then assigned between the first residue pair and the so-called 'nearest neighbor' among the remaining residues, *i.e.* the residue with the highest absolute value of the correlation coefficient to either of the residues in the original pair. Subsequent links in this hierarchical clustering process are created in a similar way and are graphically represented through dendrograms. Dendrograms are tree-like diagrams in which the linked residues are aligned in a horizontal row, while the vertical axis reports the magnitude of the correlation coefficient at which each successive residue is added

to the previous groups. For instance, Fig. 3b shows the dendrogram for the largest cluster found in **R** with a cut-off of $|r_{ij}| \geq 0.98$ (SI Text). This cluster will be referred to as cluster I and consists of 38 residues (Fig. 3b; Table S4), which are connected by a blue grid in Fig. 3a. The second largest cluster with $|r_{ij}| \geq 0.98$ is denoted as cluster II and contains ten residues (Fig. 3c; black grid in Fig. 3a). All other clusters within the 0.98 $|r_{ij}|$ cut-off have only two or a maximum of three residues and will not be considered further here, as they are not large enough to define a network.

The AC analysis and the related dendrograms illustrated in Fig. 3a-c indicate that the residues in each of the identified networks I and II display a highly concerted response to the perturbations used to build the correlation matrix **R**, *i.e.* they are highly cross-correlated within each cluster and in this respect they fulfill the first criterion for a network, as proposed above based on equation (1) (*i.e.* the high degree of correlation between the δ_{ik} and δ_{jk} chemical shifts). However, the clustering analysis of Fig. 3a-c does not provide any information on the second criterion for an allosteric network mentioned above, *i.e.* the separate clustering of active and inactive states. Namely, the dendrograms of Fig. 3b,c have the merit of identifying the residues belonging to clusters I and II (Fig. S2), but do not provide any information on the function of these networks in terms of their role in allostery or effector binding (*i.e.* functional assignment).

2.5.5 Functional Assignment of Networks

In order to assign a function to clusters I and II, we propose two independent approaches. The first method is still largely based on correlation matrices and subsequent hierarchical clustering, while the second relies on singular value decomposition (SVD). Specifically, the first approach starts with the selection of the sub-matrix of **M** corresponding to cluster I residues (*i.e.* matrix $\mathbf{M}_\mathbf{I}$ in Fig. S2) and the computation of the correlation matrix of $\mathbf{M}_\mathbf{I}$ (*i.e.* $\mathbf{R}_\mathbf{I}$). $\mathbf{R}_\mathbf{I}$ is then analyzed according to the same agglomerative clustering algorithm discussed above for Figures 3b and 3c, but this time to group states based on residues rather than residues based on states (Fig. S2). The resulting dendrogram is shown in Fig. 3d, which clearly indicates that cluster I correctly separates the active from the inactive states, confirming that cluster I functions as an allosteric activation network. For example, Fig. 2f-h displays three representative correlations between cluster I residues with well-separated active and inactive states. Interestingly, when a similar protocol is applied to the cluster II submatrix of **M** (Fig. S2), the separation is

not between active and inactive states but between the apo- and the bound- states (Fig. 3e), strongly suggesting that the function of cluster II, unlike cluster I, is linked more to binding than activation.

A second independent method used to validate the functional assignments of networks I and II to allostery and binding, respectively, is based on SVD (36, 37). Considering that SVD is designed to search for orthogonal directions of maximum variance, an effective assignment of allosteric *vs.* binding functions is obtained when SVD is employed to factorize matrix **M'** rather than **M** (SI Text; Fig. S2). **M'**, unlike **M**, is a matrix that compiles relative rather than absolute compounded chemical shifts. Specifically, **M'** is calculated using the antagonist bound state as reference (*i.e.* Rp-cAMPS). This means that for each residue **M'** reports four δ_{NHcomb}. differences: apo *vs.* Rp-cAMPS, cAMP *vs.* Rp-cAMPS, Sp-cAMPS *vs.* Rp-cAMPS and 2'-OMe-cAMP *vs.* Rp-cAMPS (Fig. S2). The matrix **M'** is then column mean centered and factorized through SVD (Fig. S2). The first two principal components (PCs) identified through SVD account for > 96 % of the total variance (Table S2) and therefore the other PCs can be safely discarded. The resulting loading and score plots for the first two PCs are displayed in Fig. 4.

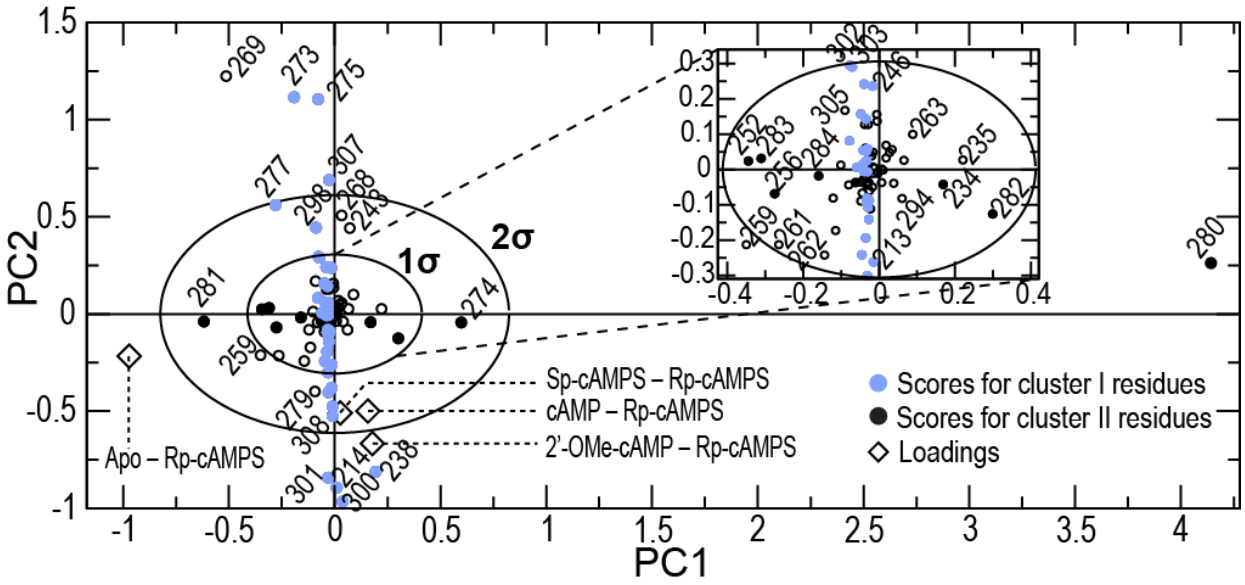

Figure 4: Score and loading plots for the first two principal components derived from the SVD of matrix **M'** (Fig. S2). Circles correspond to scores and diamonds to loadings. Selected scores are labeled with the related residue number (i.e. **M'** row) and each loading is labeled with the respective state difference (i.e. **M'** column). The scores for residues in clusters I and II (Fig. 3) are shown as blue and black filled circles, respectively. Scores for residues that do not belong to cluster I or II are shown as open circles. Ellipsoids at one and two standard deviations for the first two principal components are displayed with solid black lines. The inset shows an expansion of the score plot below one standard deviation.

The loading plot is useful to decode the chemical meaning of each PC, while the residue specific scores quantify the contribution of each residue to each PC. For example, the loading plot (Fig. 4, diamonds)

clearly indicates that the major contribution to the first PC (PC1) is from the apo *vs.* Rp-cAMPS column of **M'**. Since the Rp-cAMPS antagonist binds EPAC, but does not activate it, this means that PC1 mainly quantifies binding as opposed to allosteric contributions. Figure 4 also shows that for PC2, unlike PC1, the most significant contributions arise from the cAMP *vs.* Rp-cAMPS, Sp-cAMPS *vs.* Rp-cAMPS and 2'-OMe-cAMP *vs.* Rp-cAMPS columns of **M'**. Since cAMP, Sp-cAMPS, 2'-OMe-cAMP and Rp-cAMPS all bind EPAC but only the first three ligands are able to activate it, PC2 reflects mostly contributions from allosteric activation rather than binding. In summary, the SVD loadings of Fig. 4 indicate that PC1 and PC2 are mainly associated with binding and allostery, respectively.

Based on the loading analysis, the position of the score for a given residue relative to the PC1 and PC2 coordinates (Fig. 4, circles) reflects mainly the contributions of that specific residue to binding and allostery, respectively. We should then expect that the SVD scores for residues identified by AC as belonging to the allosteric cluster I (Fig. 3b,d) are aligned along PC2, while the SVD scores for residues assigned by AC to the binding-related cluster II (Fig. 3c,e) are distributed along PC1. This prediction is remarkably well confirmed by Fig. 4, which shows that the scores for residues in clusters I and II (blue and black solid circles, correspondingly) tend to be confined along PC2 and PC1, respectively, with only a minimal spread over the rest of the PC1/PC2 plane. The SVD factorization (Fig. 4) therefore corroborates the AC-based functional assignment of these two networks (Fig. 3d,e). Further details on the SVD *vs.* AC comparison are available in the SI Text. Overall, taken together the SVD and AC methods provide a robust protocol (Fig. S2) to evaluate the binding and allosteric relevance of the clusters identified through CHESCA. A further and NMR-independent confirmation of the AC/SVD-based functional assignments of networks I and II to allostery and binding, respectively, is provided by mutational analyses, as explained in SI Text. Once the identification and the functional assignment of networks I and II is validated, further insight into the significance of these clusters is obtained by analyzing how they relate to the residue-specific chemical shift changes (Fig. 3f) and to the known structure of EPAC (Fig. 5).

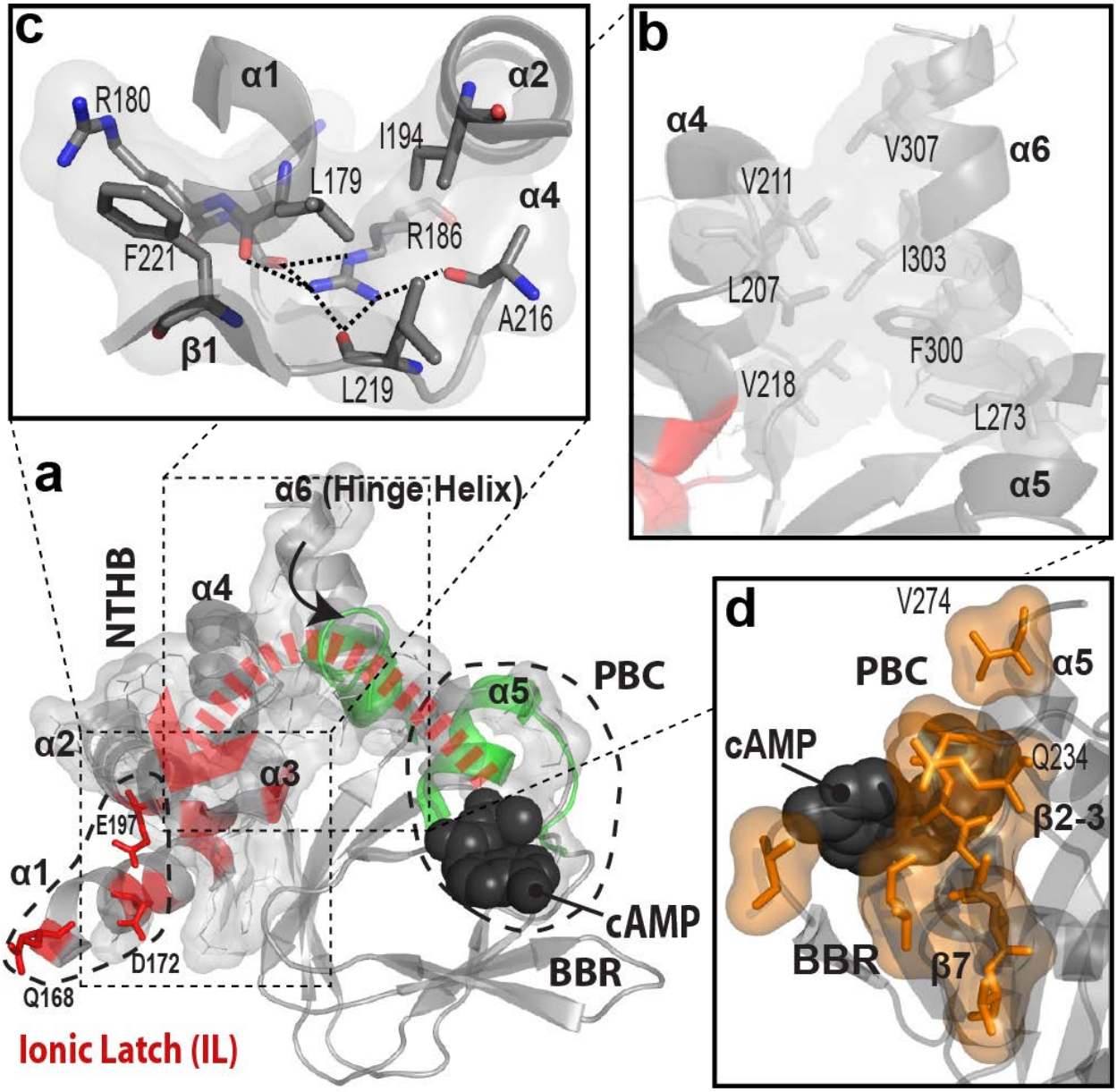

Figure 5: (a) Map of cluster I residues (gray stick and surface representation) onto the ribbon diagram of the EPAC CBD using the same color coding as in Figure 1d and with the NTHB sites displaying increased dynamics upon cAMP-dependent activation marked in red. Cluster I residues that appear as isolated, non-contiguous pairs are shown separately in Fig. S3c,d. The dashed red arrow outlines a possible signal propagation pathway from the PBC to the IL based on cluster I. **(b, c)** Key interactions within cluster I: (b) two spines of hydrophobic residues at the interfaces between α4, α5 and α6; (c) an interaction node centered on R186 and bridging α1, α2, α4 and β1. **(d)** Cluster II (orange surface and sticks).

2.5.6 Significance of Minor Chemical Shift Changes

Figure 3f reports the maximum ppm variations observed for each residue across all the five states of Fig. 2a. One notable result evident from Fig. 3f is that the CHESCA-based allosteric cluster I includes also

several residues subject to minimal ppm changes. For instance, Fig. 3f shows that the allosteric cluster I comprises several residues in the IL-spanning N-terminal helices α1 and α2, which display minor ppm differences. Consistently with the minor chemical shift variations (Fig. 3f), this region of EPAC does not change significantly in local conformation upon cAMP-binding (Fig. 1c), but exhibits enhanced local dynamics upon activation (Fig. 1c). This example illustrates a more general scenario in which minor chemical shift changes are associated with subtle variations in the local environment that fall below the resolution of currently available structure determination methods. Such subtle changes may include minor readjustments in backbone and/or side-chain structure and/or dynamics, *e.g.* in the energy landscape of the inter-converting states within the conformational ensemble. Although these variations often escape direct structural observation, they are sensed through chemical shift differences and underlie the modulations in dynamics that are critical for the allosteric signal propagation (7, 24). It is therefore essential to not only identify such allosterically relevant minor chemical shift changes through AC of the **R** matrix (Fig. 3, Fig. S2), but to also ensure that these minor variations in ppm values are reliably measured (SI Text). Further insight into the interactions mediating these chemical shift correlations is gained by mapping clusters I and II onto the structures of the EPAC CBD (Fig. 5).

2.5.7 Structural Analysis of the Allosteric Network Defined by Cluster I

Figure 5a shows that the majority of the residues in cluster I define a continuous surface that bridges the gap between the region with most significant conformational changes, *i.e.* the PBC and the hinge helix (Fig. 5a, green ribbon), and the region which is not subject to any significant local conformational change but displays enhanced dynamics in the active state, *i.e.* the IL zone in the distal N-terminal helical bundle (NTHB) (Fig. 1c; Fig. 5a, red ribbon). Interestingly, these bridging residues of cluster I are almost entirely confined to the non-contiguous α-subdomain of the EPAC CBD, originating from the α5 helix in the PBC and extending into the C-terminal α6 hinge helix and α4-1 in the NTHB, ultimately reaching the inhibitory IL zone (Fig. 5a, red dashed arrow). Within this extensive allosteric network bridging α5 to α1, two sub-sets of interactions are particularly notable.

First, residues L273 in α5 and F300, I303 and V307 in α6 define a hydrophobic spine (Fig. 5b) that interfaces with another hydrophobic spine in α4 including L207, V211 and V218 (Fig. 5b). These coupled spines within cluster I indicate that the previously identified contacts between L273 and F300 (*i.e.* 'hydrophobic hinge') (31) are part of a significantly more extended allosteric network of interactions that couples α5 in the PBC to both α6 and α4 (Fig. 5b). Second, α4 is also part of another sub-set of cluster I

residues that connect the α4/β1 region to helices α2 and α1 (Fig. 5c). Although composed mostly by hydrophobic amino acids (Fig. 5c), this second sub-cluster I is centered around R186 that serves as an interaction hub stabilized by multiple hydrogen bonds between its guanidinum and adjacent carbonyl oxygen atoms (Fig. 5c). Together the interactions formed by the hydrophobic spines (Fig. 5b) and those that nucleate around R186 (Fig. 5c) provide an effective allosteric network to propagate the cAMP signal from the PBC to the distal IL region. Such cAMP signal propagation relies on subtle structural re-packing and/or re-arrangements that escape detection by conventional structure determination methods, but are effectively sensed by NH chemical shift changes.

The bridging role from the PBC to the IL outlined above for cluster I helps explain how the IL dynamics are enhanced upon cAMP binding and rationalizes how the IL residues mediate in a cAMP-dependent manner several critical inhibitory salt bridges. Furthermore, cluster I also includes residues that don't bridge the PBC to the IL (*i.e.* 9 out of 38 cluster I residues) and reside in the β-subdomain of the EPAC CBD (Fig. 5a; Fig. S3). Interestingly, these cluster I residues are not part of the β-strands but of the inter-strand loops of the typical β-barrel of CBDs (Fig. S3), suggesting that the β-strands serve primarily as a passive scaffold, while the loops play a more active allosteric role. This observation applies both to the isolated inter-strand regions located opposite to the PBC (Fig. S3c,d) and to those adjacent to the PBC (Fig. S3b). The structural analysis of cluster II (Fig. 5d) is available as SI Text.

2.6 Conclusions

We have shown that the covariance analysis of subtle but significant NMR chemical shift changes caused by a set of covalently modified agonists and antagonists reveals extensive allosteric networks that bridge the site of effector binding to distant regions mediating inhibitory interactions, including those that are dynamically-controlled. The chemical shift covariance analysis (CHESCA) relies on agglomerative clustering (AC) and singular value decomposition (SVD) of the chemical shift matrix. The combined analysis of the dendrograms and score plots generated by AC and SVD, respectively, not only identifies networks of coupled residues, but also provides a residue-specific dissection of the relative contributions to binding and allostery. These CHESCA results are independently validated by several known EPAC mutants as well as by previous assessments of allostery based on co-evolutionary patterns (34). Overall, the CHESCA method employs a small but carefully selected chemical library of effector analogs to decode the binding and allosteric information content embedded in chemical shifts. We anticipate CHESCA to be

of general applicability for mapping allosteric networks in proteins amenable to NMR, complementing results from mutant cycles (38) and analyses of functional couplings between co-evolving sites (39).

2.7 Acknowledgements

We thank P. Britz-McKibbin and F. Fogolari for helpful discussions and the Canadian Institute of Health Research (CIHR) and the National Sciences and Engineering Research Council (NSERC) for financial support. We are also indebted to the Heart and Stroke Foundation of Canada (HSFC) for a Maureen Andrew New Investigator to G.M. and to NSERC for a graduate scholarship to B.V.S..

2. 8 References

1. Smock RG, Gierasch LM (2009) Sending signals dynamically. *Science* 324: 198-203.

2. Tsai CJ, del Sol A, Nussinov R (2008) Allostery: Absence of a change in shape does not imply that allostery is not at play. *J Mol Biol* 378: 1-11

3. Kuriyan J, Eisenberg D (2007) The origin of protein interactions and allostery in colocalization. *Nature* 450: 983-990.

4. Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* 4: 474-482.

5. Changeux JP, Edelstein SJ (2005) Allosteric mechanisms of signal transduction. *Science* 308: 1424-1428.

6. Kern D, Zuiderweg ER (2003) The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* 13: 748-757.

7. Daily MD, Gray JJ (2007) Local motions in a benchmark of allosteric proteins. *Proteins* 67: 385-399.

8. Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* 5: 789-796.

9. Velyvis A, Yang YR, Schachman HK, Kay LE (2007) A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. *Proc Natl Acad Sci USA* 104: 8815-8820.

10. Hilser VJ, Garcia-Moreno EB, Oas TG, Kapp G, Whitten ST (2006) A statistical thermodynamic model of the protein ensemble. *Chem Rev* 106: 1545-1558.

11. Popovych N, Sun S, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. *Nat Struct Mol Biol* 13: 831-838.

12. Tzeng SR, Kalodimos CG (2009) Dynamic activation of an allosteric regulatory protein. *Nature* 462: 368-372.

13. Cooper A, Dryden DT (1984) Allostery without conformational change. A plausible model. *Eur Biophys J* 11: 103-109.

14. Fuentes EJ, Der CJ, Lee AL (2004) Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. *J Mol Biol* 335: 1105-1115.

15. Das R, *et al* (2008) Entropy-driven cAMP-dependent allosteric control of inhibitory interactions in exchange proteins directly activated by cAMP. *J Biol Chem* 283: 19691-19703.

16. Das R, *et al* (2009) Dynamically driven ligand selectivity in cyclic nucleotide binding domains. *J Biol Chem* 284: 23682-23696.

17. McNicholl ET, Das R, SilDas S, Taylor SS, Melacini G (2010) Communication between tandem cAMP binding domains in the regulatory subunit of protein kinase A-Ialpha as revealed by domain-silencing mutations. *J Biol Chem* 285: 15523-15537.

18. Lipchock JM & Loria JP (2010) Nanometer propagation of millisecond motions in V-type allostery. *Structure* 18: 1596-1607.

19. Masterson LR, Mascioni A, Traaseth NJ, Taylor SS & Veglia G (2008) Allosteric cooperativity in protein kinase A. *Proc Natl Acad Sci USA* 105: 506-511.

20. Masterson LR, *et al.* (2010) Dynamics connect substrate recognition to catalysis in protein kinase A. *Nat Chem Biol* 6: 821-828.

21. Jarymowycz VA, Stone MJ (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. Chem Rev 106:1624–1671.

22. Ma B & Nussinov R (2010) Enzyme dynamics point to stepwise conformational selection in catalysis. *Curr Opin Chem Biol* 14: 652-659.

23. Cui Q & Karplus M (2008) Allostery and cooperativity revisited. *Protein Sci* 17: 1295-1307.

24. Zhuravleva A, *et al* (2007) Propagation of dynamic changes in barnase upon binding of barstar: An NMR and computational study. *J Mol Biol* 367: 1079-1092.

25. Revington M, Zhang Y, Yip GN, Kurochkin AV, Zuiderweg ER (2005) NMR investigations of allosteric processes in a two-domain thermus thermophilus Hsp70 molecular chaperone. *J Mol Biol* 349: 163-183.

26. Shen Y, *et al* (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105: 4685-4690.

27. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104: 9615-9620.

28. Gloerich M, Bos JL (2010) Epac: Defining a new mechanism for cAMP action. *Annu Rev Pharmacol Toxicol* 50: 355-375.

29. Kraemer A, *et al* (2001) Dynamic interaction of cAMP with the rap guanine-nucleotide exchange factor Epac1. *J Mol Biol* 306: 1167-1177.

30. Rehmann H, Das J, Knipscheer P, Wittinghofer A, Bos JL (2006) Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state. *Nature* 439: 625-628.

31. Rehmann H, *et al* (2008) Structure of Epac2 in complex with a cyclic AMP analogue and RAP1B. *Nature* 455: 124-127.

32. Berman HM, *et al* (2005) The cAMP binding domain: An ancient signaling module. *Proc Natl Acad Sci USA* 102: 45-50.

33. Rehmann H, Schwede F, Doskeland SO, Wittinghofer A, Bos JL (2003) Ligand-mediated activation of the cAMP-responsive guanine nucleotide exchange factor epac. *J Biol Chem* 278: 38548-38556.

34. Kannan N, *et al* (2007) Evolution of allostery in the cyclic nucleotide binding module. *Genome Biol* 8: R264.

35. Aden J, Wolf-Watz M (2007) NMR identification of transient complexes critical to adenylate kinase catalysis. *J Am Chem Soc* 129: 14003-14012.

36. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863-14868.

37. Sakurai K, Goto Y (2007) Principal component analysis of the pH-dependent conformational transitions of bovine beta-lactoglobulin monitored by heteronuclear NMR. *Proc Natl Acad Sci USA* 104: 15346-15351.

38. Horovitz A & Fersht AR (1990) Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. *J Mol Biol* 214: 613-617.

39. Lockless SW & Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286: 295-299.

2.9 Supplementary Information: Supporting Text

2.9.1 Derivation of Equation 1. Linear Inter-Residue Chemical Shift Correlations for a Two-State Activation Equilibrium in the Fast-Exchange Regime

When the active-inactive exchange is fast in the chemical shift time scale, the chemical shifts observed for residues *i* and *j* in each of the perturbation sets (*i.e.* δ_{ik} and δ_{jk}) are a simple linear average of the ppm values for the active and inactive states (1):

$$\delta_{ik} = p_k \delta_{i,Ac} + (1 - p_k) \delta_{i,In.} \qquad (SI1)$$

$$\delta_{jk} = p_k \delta_{j,Ac} + (1 - p_k) \delta_{j,In.} \qquad (SI2)$$

where $\delta_{i,Ac}$ (and $\delta_{j,Ac}$) and $\delta_{i,In}$ (and $\delta_{j,In}$) denote the combined ppm values of residue *i* (and *j*) in the active and inactive states, respectively and p_k indicates the fractional population of active states when the k^{th} perturbation is in effect. Combined chemical shifts are computed as a weighted average of the amide ^{15}N and ^{1}H ppm values: $\delta_{ik} = 0.2\,\delta^{N}_{ik} + \delta^{H}_{ik}$, $\delta_{i,Ac} = 0.2\,\delta^{N}_{i,Ac} + \delta^{H}_{i,Ac}$ and $\delta_{i,In} = 0.2\,\delta^{N}_{i,In} + \delta^{H}_{i,In}$ (2). If we define $\Delta\delta_i$ as $(\delta_{i,Ac} - \delta_{i,In})$ and $\Delta\delta_j$ as $(\delta_{j,Ac} - \delta_{j,In})$, then:

$$\delta_{ik} = p_k \Delta\delta_i + \delta_{i,In.} \qquad (SI3)$$

$$\delta_{jk} = p_k \Delta\delta_j + \delta_{j,In.} \qquad (SI4)$$

and therefore:

$$\delta_{ik} = \delta_{jk}\,\alpha + \beta \qquad (1)$$

where $\alpha = \Delta\delta_i/\Delta\delta_j$, and $\beta = \delta_{i,In} - \alpha\delta_{j,In}$, as mentioned in the text.

Deviations from equation (1) are expected when amino acids *i* or *j* are in the vicinity of the perturbation sites and can therefore potentially sense additional variations in the local environment associated more with the specific chemical nature of the perturbations (*e.g.* ligand-specific neighbor effects at or near the binding site) than with modulations of the activation equilibrium. In other words, in the immediate proximity of the perturbed site the correlations formalized by equation (1) are obscured by local and perturbation-specific effects, resulting in a 'dark-zone'. However, the number of residues within these 'dark-zones' is minimized by choosing perturbations that are confined to a single group of residues, adjacent in both sequence and in space. This approach maximizes the superimposition of the 'dark zones' surrounding each perturbed locus and ensures that the biases on the correlations introduced

by the perturbation-specific effects are not spread across multiple sites of the protein. The cAMP derivatives shown in Fig. 2a meet this 'perturbation confinement' criterion: the covalent modifications of the phosphorothioate cAMP analogs (Rp- and Sp-cAMPS) and of 2'-OMe-cAMP perturb non-covalent interactions that are well localized within the phosphate binding cassette (PBC) (Fig. 2b). Specifically, the methylation of the 2' hydroxyl and the isolobal replacement of the axial exocyclic oxygen with the bulkier and less electronegative sulfur in Sp-cAMPS affect the hydrogen bonds that secure cAMP to the N- and C-termini of the PBC, respectively (Fig. 2a,b), while the substitution of the equatorial exocyclic oxygen in Rp-cAMPS targets the hydrogen bonds in the central region of the PBC (Fig. 2a,b). The high degree of localization of these perturbations maximizes the overlap between 'dark-zones', as shown in Fig. S4a, and therefore maximizes also the number of residue pairs that comply with equation (1).

2.9.2 Linear Inter-Residue Chemical Shift Correlations for a Three-State Activation Equilibrium in the Fast-Exchange Regime

While the main effect of binding cAMP analogs is a shift in the inactive/active equilibrium according to the differential affinity of the ligand for these two states, it is also possible that the cAMP analogs cause minor distortions in the structures that correspond to the inactive and active states. Although minimal and certainly negligible compared to the structural differences between the active and inactive states, these subtle structural variations can cause minor changes in the chemical shifts of the inactive and active states. Such structural distortions caused by the covalent modification of cAMP-analogs are not fully captured by a two-state model, but are effectively modeled in the context of a three-state model. It is therefore important to justify the use of the $|r_{ij}|$ correlation coefficient also in the context of a more general three-state model. Specifically, we have analyzed the conditions under which the pairwise inter-residue correlations remain linear in the presence of three-states. For this purpose, we start again from the simple two-state equation (*SI1*):

$$\delta_{ik} = p_k \delta_{i,Ac} + (1 - p_k) \delta_{i,In.} \qquad (SI1)$$

In the presence of a third state this equation becomes:

$$\delta_{ik} = p'_k \delta_{i,Ac} + p''_k (\delta_{i,Ac} + \varepsilon_i) + (1 - p'_k - p''_k) \delta_{i,In.} \qquad (SI5)$$

where p'_k and p''_k are the populations of the active state and of the new third state, respectively, with residue *i* chemical shifts $\delta_{i,Ac}$ and $(\delta_{i,Ac} + \delta_i)$, correspondingly. The ε_i term denotes the difference between

the chemical shifts of residue *i* in the third and in the active states. If p_k is defined as $p'_k + p''_k$, which represents the total population of active states assuming the third state is active as well, and if $\Delta\delta_i$ is defined as $(\delta_{i,Ac} - \delta_{i,In})$, then equation (*SI5*) becomes:

$$\delta_{ik} = p_k \Delta\delta_i + \delta_{i,In.} + p''_k \varepsilon_i \qquad (SI6)$$

Replacing residue *i* with residue *j,* equation (*SI6*) leads to:

$$\delta_{jk} = p_k \Delta\delta_j + \delta_{j,In.} + p''_k \varepsilon_j \qquad (SI7)$$

Combining equations (*SI6*) and (*SI7*) into a single equation, we obtain:

$$\delta_{ik} = \delta_{jk}\alpha + \beta + p''_k \Delta\delta_i\gamma \qquad (SI8)$$

where $\alpha = \Delta\delta_i/\Delta\delta_j$, $\beta = \delta_{i,In} - \alpha\delta_{j,In}$, as seen before for the two-state model, and $\gamma = (\varepsilon_i/\Delta\delta_i) - (\varepsilon_j/\Delta\delta_j)$. The first two terms in equation (*SI8*) (*i.e.* $\delta_{jk}\alpha + \beta$) are as in the previously derived two-state model equation (*1*) and they describe linear correlations. The third term in equation (*SI8*) (*i.e.* $p''_k \Delta\delta_i \gamma$) is unique to the three-state model. Due to the dependence of this new term on the k^{th} perturbation, it cannot be combined with the β constant and it is a potential source of non-linearity if it is non-negligible relative to the linear terms in α and β. It is therefore important to understand under which circumstances the $p''k\Delta\delta_i\gamma$ non-linear term becomes negligible and equation (*SI8*) is reduced to the linear equation (*1*). This occurs if at least one of the three following conditions is met:

I. $p''_k \approx 0$

II. $\varepsilon_i/\Delta\delta_i \approx 0$ and $\varepsilon_j/\Delta\delta_j \approx 0$ or $\varepsilon_i \approx -\Delta\delta_i$ and $\varepsilon_j \approx -\Delta\delta_j$

III. $\varepsilon_i/\Delta\delta_i \approx \varepsilon_j/\Delta\delta_j$

Condition (I) corresponds to the case in which the population of the third state p''_k is minimal. However, even when the third state is significantly populated, it is still possible that the $p''_k\Delta\delta_i\gamma$ term is negligible provided that the factor $\gamma \approx 0$. This occurs if conditions (II) and/or (III) are met. Considering that ε_i (ε_j) is the difference between the residue *i* (*j*) chemical shifts of the third and of the active states and taking into account that changes in chemical shifts reflect changes in the structure of the states, scenario (II) reflects the case in which the third state is very similar, although not identical, to either one of the other states (*i.e.* the active or the inactive state). For instance, the third state could correspond to a structure that falls within the same broad energy well as the active state. Overall, conditions (I) and (II) apply when the three-state model tends to converge towards a two-state equilibrium.

Scenario (III) corresponds to the case in which the structural changes that occur in going from the

active to the third state cause correlated perturbations in the local environments of both residues *i* and *j.* Such correlated perturbations indicate that, even in the context of a three-state model, values of the $|r_{ij}|$ correlation coefficient close to unity are still evidence that residues *i* and *j* display a collective and concerted response to the applied perturbations, *i.e.* residues are part of the same network of coupled residues. We therefore conclude that the use of $|r_{ij}|$ in the proposed CHESCA protocol is justified also in the context of the general three-state model summarized by equation (*SI8*).

Finally, it should be noted that even when the three-state equation (*SI8*) leads to linear correlations, it is still possible that the presence of a third state causes chemical shift reordering effects similar to those observed in the pairwise correlations of Figure 2f-h. This is most directly appreciated with a simple numerical example. For instance, let's assume that the populations and the chemical shifts are as in Table S5, which mimics a scenario similar to that observed in Figure 2g. Let's also assume that for residue *i:* $\delta_{i,Ac}$ = 32.74 ppm, $(\delta_{i,Ac} + \mathcal{E}_i)$ = 32.71 ppm, $\delta_{i,In}$ = 32.64 ppm and that for residue *j:* $\delta_{j,Ac}$ = 31.26 ppm, $(\delta_{j,Ac} + \mathcal{E}_j)$ = 31.25 ppm, $\delta_{j,In}$ = 31.22 ppm. Based on equation (*SI5*), these values result in the correlation shown in black in the Figure S4b. Figure S4b also includes a correlation shown in red, which corresponds to the simple two-state model, *i.e.* $\mathcal{E}_i = 0$ and $\mathcal{E}_j = 0$ so that $\gamma = 0$ as well in equation (*SI8*).

Figure S4b clearly illustrates that the addition of a third state to the two-state model does not necessarily compromise significantly the linearity, but does cause an appreciable re-ordering of the three active forms (*i.e.* cAMP, Sp and 2'-OMe), as shown by the dashed arrows and as also observed in Figure 2g. The re-ordering effect illustrated in Figure S4b is specific to a given correlation pair because it depends on the value of the γ coefficient and therefore of $\mathcal{E}_i$ and $\mathcal{E}_j$, which vary from residue pair to residue pair, explaining why the re-ordering of the active species is observed in Figure 2g but not in Figure 2f or 2h. Overall, the example above clarifies how it is possible that the presence of a third state can cause a partial reordering of the points that correspond to the active ligands, without significantly compromising the linearity of the correlation and the separation between the active (*i.e.* cAMP, Sp and 2'-OMe) and the inactive species (*i.e.* apo and Rp), which are at the basis of the proposed CHESCA method.

2.9.3 Extension to Side Chains:

Although equations (*SI1-SI8*) have been presented for the combined chemical shift of backbone amide ^{1}H and ^{15}N spins, the CHESCA approach can in principle be applied to any type of chemical shifts, including those of side chain ^{1}H, ^{13}C and ^{15}N spins, provided that accurate and precise ppm values are measured.

One possible strategy to implement this extension of the CHESCA analysis is to compute for each residue a compounded chemical shift that includes the ppm values of the ^{1}H, ^{13}C and ^{15}N spins with weight factors that reflect the standard deviations of the respective chemical shifts in the BioMagResBank database. However, one complication in the extension of the CHESCA approach to side chains is caused by the variability of the types of side chains and the consequent possible biases in the pairwise inter-residue linear correlations. This problem can be addressed by limiting the compounded chemical shift calculations to either backbone and C_β spins only or to methyl groups only. The first approach confines the potential biases in the pairwise correlations only to correlations involving Glycine residues, while the second option is anticipated to be particularly useful for large systems for which methyl TROSY-based NMR spectroscopy is suitable (3).

2.10 Supplementary Information: Materials and Methods

2.10.1 Sample Preparation

The EPAC1$_h$ (149-318) construct was chosen because it spans the EPAC CBD and, as previously shown (4, 5), it provides a reliable representation of the interactions perturbed by the selected cAMP analogs. The EPAC1$_h$ (149-318) construct was expressed as a GST-fusion construct in the *E. coli* strain BL-21 (DE3). Cells were grown in isotopically enriched minimal media (1.00 g/L $^{15}NH_4Cl$, 3.00 g/L ^{13}C-glucose, 6.78g/L Na_2HPO_4, 3.00 g/L KH_2PO_4, 0.50 g/L NaCl), supplemented with trace metals, D-biotin (10 mg/mL), and thiamine-HCl (10 mg/mL). Expression was induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside at an optical density of 0.50-0.60 (λ=600 nm), and further incubated for 16-17hrs at 20°C. Cells were lysed and apo-EPAC1$_h$ (149-318) was purified and concentrated to 0.10 – 0.15 mM using previously described protocols and buffers (15, 16). Additionally, 0.5 mM of singly labeled ^{15}N-acetylglycine (Sigma Aldrich) was added to each sample as a ppm referencing compound. Stock solutions of all ligands (25 mM in cyclic nucleotide) were prepared using the same buffer as for the protein solution. Starting from the apo-samples, stock solutions of the cyclic-nucleotide ligands were titrated into the protein solution to full saturation. For all titrations we employed cyclic-nucleotide ligands (BIOLOG and Sigma) of the highest available purity (*i.e.* > 99% with the exception of 2'-OMe-cAMP with a purity > 98%). As a control for potential dilution effects occurring during the ligand titrations, a blank buffer solution without any ligand was also titrated into the protein sample and it did not result in any measurable chemical shift changes.

2.10.2 NMR Measurements:

Upon addition of each titrant, gradient and sensitivity enhanced [^{15}N-^{1}H] heteronuclear single quantum coherence (HSQC) spectra with 256 (t_1) and 1024 (t_2) complex points and spectral widths of 31.82 and 14.06 ppm for the ^{15}N and ^{1}H dimensions, respectively, were recorded with 8 scans and a recycle delay of 1.00 s. All HSQC spectra were processed with NMRPipe (6) employing linear prediction in the ^{15}N dimension, and a resolution-enhancing 60° shifted sine-squared bell window function for both dimension. All ^{15}N-^{1}H HSQC cross-peaks were analyzed with Sparky using Gaussian line-fitting (7). Assignments were confirmed using triple-resonance experiments (8).

2.10.3 Structural Analyses

All structures were analyzed and displayed using Molmol (9) and/or Pymol (10). The local RMSD values reported in Fig. 2c were computed using the coordinates for the holo- and apo- EPAC structures with PDB codes 3CF6 (11) and 2BYV (12), respectively. For each residue the backbone heavy atoms of the amino acid triplet formed by the selected and the two adjacent residues were superimposed. The pairwise RMSD for the backbone heavy atoms of all three residues was then computed. Since the local RMSD is calculated between residue triplets, an error margin of ± one residue should be allowed in the interpretation of the local RMSD.

2.10.4 Measurement of Minor Chemical Shift Changes

Minor ^{1}H chemical shift differences of the order of ~0.01 ppm correspond to frequency variations of ~7 Hz at a field of 700 MHz and are measurable with high accuracy and precision with currently available commercial spectrometers (13). However, it is critical to avoid systematic errors due to the lack of reliable ppm referencing signals or due to potential non-specific effects associated with the perturbations employed to measure the correlation coefficients (*e.g.* binding of agonists and antagonists). To address these possible concerns, ^{15}N-AcGly was used as an internal referencing standard for both ^{1}H and ^{15}N ppm measurements, rather than using the residual water signal for ^{1}H or indirect methods based on the gyromagnetic ratios for ^{15}N. In addition, as a control for potential non-specific binding effects, the chemical shift based dose-dependent binding isotherms for representative residues subject to changes in resonance frequencies ≤ 10 Hz were examined (Fig. S5). Fig. S5 shows that for both ^{1}H and ^{15}N nuclei

clear plateaus are observed for the binding isotherms of all four ligands, indicating saturability for all bound-states and ruling out any significant non-specific binding effect. Furthermore, the K_D values measured through these binding isotherms (Fig. S5; Fig. S1; Fig. 2c) are in agreement with the previously measured dissociation constants (Table S1). Moreover, the agglomerative clustering (AC) and singular value decomposition (SVD) analyses presented here were conservatively confined to residues for which the frequency spread across the five states (*i.e.* apo and four bound states; Fig. 2a) meets the following threshold requirements: $\Delta\nu_{N,max} > 5$ Hz and/or $\Delta\nu_{H,max} > 10$ Hz (Fig. S2). The ppm changes associated with these frequency variations can be minimal ($\Delta\delta_{NHcomb.} > 0.014$ ppm at a field of 700 MHz) but are still surprisingly well correlated with both other minor and major ppm differences as shown in Fig. 2f-h and Fig. 3a,f. Moreover, the $\Delta\nu_{N,max} > 5$ Hz and/or $\Delta\nu_{H,max} > 10$ Hz selection criteria ensure that for all residues in clusters I and II the maximum variation in the combined ppm across the five states is always greater than twice the error in each of the combined ppm values.

2.10.5 Chemical Shift Analyses:

Fig. S2 provides a full description of the overall protocol employed for the AC and SVD analyses of the measured chemical shifts. The SVD analyses as well as all other matrix manipulations outlined in Fig. S2 were implemented with Octave (http://www.gnu.org/software/octave/), while for the agglomerative clustering analysis the 'Cluster' suite of programs was used (http://bonsai.hgc.jp/~mdehoon/software/cluster/) (14). The dendrogram trees were generated using JavaTreeview (http://jtreeview.sourceforge.net/) (15).

The i^{th} row of matrix **M** shown in Fig. S2 is an array: $\{\delta_{i1}, \delta_{i2}, \ldots, \delta_{in}\}$, which will be referred to as $\underline{\delta_i}$. The cross-correlation between residues *i* and *j* is then quantified by the Pearson's correlation-coefficient (r_{ij} element of the correlation matrix **R**), which measures the covariance of arrays $\underline{\delta_i}$ and $\underline{\delta_j}$ (cov_{ij}) normalized according to the product of their standard deviations ($s_i s_j$):

$$r_{ij} = \frac{cov_{ij}}{s_i s_j} = \frac{\sum_{k=1}^{n}\left(\delta_{ik} - \bar{\delta}_j\right)\left(\delta_{jk} - \bar{\delta}_j\right)}{\sqrt{\sum_{k=1}^{n}\left(\delta_{ik} - \bar{\delta}_i\right)^2 \sum_{k=1}^{n}\left(\delta_{jk} - \bar{\delta}_j\right)^2}} \qquad \text{(SI9)}$$

where $\bar{\delta}_i = \left(\frac{1}{n}\right)\sum_{k=1}^{n}\delta_{ik}$ and $\bar{\delta}_j = \left(\frac{1}{n}\right)\sum_{k=1}^{n}\delta_{jk}$ are the average values of δ_i and δ_j, respectively, and n is the number of states (*i.e.* n = n_{st} =5 in Fig. S2 and Fig. 2). As a result of the normalization of the covariance with respect to the standard deviation product, the correlation coefficient between residues i and j is insensitive to the magnitude of the chemical shift variations occurring within each of the $\underline{\delta_i}$ and $\underline{\delta_j}$ arrays, *i.e.* it is possible that $|r_{ij}|$ approaches near maximum values (~1) even when s_i and/or s_j are minimal. This means that the correlation coefficient defined in equation (*SI9*) is an excellent tool to capture the allosteric relevance not only of large ppm shifts, but also of those subtle yet significant and informative chemical variations mentioned in the main text and anticipated based on equation (1). These minor changes in resonance frequencies would otherwise pass unnoticed as they often fall well below the commonly used threshold values derived from simple descriptive statistics, such as the protein grand average plus one standard deviation (4).

2.10.6 Singular Value Decomposition (SVD)

Fig. S2 also illustrates the SVD analysis protocol. SVD is a widely used method of dimensionality reduction (14, 16, 17, 18) and is essentially equivalent to a coordinate rotation whereby each row of the input matrix is re-computed as a linear combination of basis vectors (also called principal components (PCs)) oriented along orthogonal directions of maximal variance (14). For instance, if an n x 3 matrix is represented by an ellipsoidal distribution of n points in the space defined by the x, y and z Cartesian coordinates, then the SVD is equivalent to a rotation of the Cartesian axes along the axes of the ellipsoid, with the longest and shortest axes being defined as the first and third PCs, respectively. If most of the variance in the three-dimensional distribution of points is accounted for by the first two PCs (*i.e.* if the first two axes of the ellipsoid are significantly larger than the third), the deviations of the points from the plane of the ellipsoid defined by the first two PCs are effectively considered noise and thus discarded, leading to a new matrix with reduced dimensionality (*i.e.* n x 2) but still preserving most of the relevant information of the original n x 3 matrix (14). However, it is important to note that the SVD results heavily depend on data pre-processing.

2.10.7 Matrix Pre-Processing for SVD (*i.e.* the Matrix M' vs. the Row Mean Center Matrix M)

A typical pre-processing prior to SVD is row mean centering and therefore we considered first the SVD factorization of the row mean centered matrix **M**. The SVD of the row mean centered **M** indicates that the first two principal components (*i.e.* PC1 and PC2) account for more than 95 % of the total variance and the PC1 *vs.* PC2 scores and loadings plots are shown in Fig. S6. Fig. S6 shows as black dots the scores for residues in the AC-based cluster II (linked to binding) and as red dots the scores for the residues in the AC-based cluster I (linked to allostery). From Fig. S6 it is clear that the scores of cluster II residues are aligned along PC1, which represents binding as it contains a negative loading contribution from the apo-state and positive loading contributions from the holo (*i.e.* cyclic nucleotide-bound) states. However, the cluster I scores do not align with either PC1 or PC2 (Fig. S6). Specifically, Fig. S6 shows that the scores of cluster I residues tend to align along a line that is not orthogonal to PC1 and points towards the loadings of the two super-active states (*i.e.* Sp-cAMPS and 2'OMe-cAMP). This line is also perpendicular to the line that connects the loading of the apo-state to that of the antagonist Rp-cAMPS-bound state (Fig. S6). While this result is fully consistent with the allosteric function assigned to cluster I residues, the lack of orthogonality between the cluster I and II scores (Fig. S6) complicates the biophysical interpretation of the first two principal components (PC1 and PC2 in Fig. S6). We therefore considered an alternative pre-processing of the $n_r \, x \, n_{st}$ matrix **M** (Fig. S2), in which we used the Rp-cAMPS antagonist-bound state as reference rather than the row-mean. In other words we applied the SVD factorization to the $n_r \, x \, (n_{st} - 1)$ matrix **M'** (Fig. S2).

As explained in the main text, when the SVD is applied to **M'** (as opposed to the row-mean centered **M**) the scores of the two clusters tend to align with the first two principal components (Fig. 4). It should also be noted that if matrix **M'** had been obtained by using the apo-state as the reference state as opposed to the Rp-cAMPS antagonist-bound state, the interpretation of the principal components would still be complicated by a lack of orthogonality similar to that observed in Fig. S6. Furthermore, when matrix **M'** is computed using the Rp-cAMPS antagonist-bound state as reference, each of the four columns of **M'** is amenable to a straightforward biophysical interpretation: the first column refers to the differences between the apo- and the antagonist-bound state, *i.e.* a manifestation of binding, while each of the remaining three columns pertains to the differences between an active-bound state and the inactive-bound state, *i.e.* to activation and allostery rather than to binding. For these reasons, we opted to focus on the SVD analysis of a matrix **M'** in which the Rp-cAMPS antagonist-bound state is used as a reference, rather than on the SVD factorization of the row mean-centered matrix **M** or of other **M'** matrices calculated with other reference states.

2.11 Supplementary Information: Comparison of the SVD vs. AC Chemical Shift Covariance Analyses

One clear advantage of SVD compared to AC is that the relative contributions to binding *vs.* allostery are readily quantified for each residue. For instance, the score plot of Fig. 4 shows that the PBC residues L273 and A277, although part of the allosteric cluster I, display a |PC1| component larger than most of the other residues of network I, suggesting these two sites affect not only allostery but also to some extent ligand binding. This prediction is independently confirmed by the L273W mutation, which affects not only the allosteric activation (Table S3), but also the affinity for cAMP (five-fold increase in K_d) (19). Therefore at least some residues belonging to the allosteric network (*i.e.* cluster I) play a role in binding as well. Conversely, it is also possible that selected residues assigned to the binding network (*i.e.* cluster II) affect allostery. For instance, the PBC residue A280 is characterized by a PC2 component markedly larger than the other residues of cluster II suggesting that site 280 is involved not only in cAMP binding, as shown in Fig 2b, but to some degree in the cAMP-dependent allosteric control as well. This prediction is independently confirmed by the A280T mutation, which causes a 50 % reduction in the maximal activation of EPAC by cAMP (20). These residues with a dual allosteric and binding role are expected to be critical for the coupling between the activation and binding equilibria, which is at the basis of the cAMP-dependent regulation of EPAC (Fig. 1b).

Another distinct advantage of SVD over AC is that the score plot (Fig. 4) includes also those residues that do not belong to any AC-based network due to the ligand neighbor effects mentioned above, which compromise the correlations described by equation (1) and create a 'dark zone' for the AC method in the proximity of the perturbation sites (Fig 2b; Fig. S4a). For example, G269 is directly involved in cAMP binding (Fig. 2b) and falls within am AC-'dark zone' (Fig. S4a). As a result, G269 does not belong to clusters I or II (Fig. 4). However, G269 still displays |PC1| and |PC2| scores comparable to some of the highest values observed for residues in networks I and II. This observation suggests that G269 plays a dual role in both binding and allostery. The role of G269 in binding is confirmed by the hydrogen bond it forms with cAMP (Fig. 2b), while the allosteric function of G269 is fully consistent with the observation that this highly conserved glycine is critical for the cAMP-dependent activation of the homologous CBDs of protein kinase A (PKA) (21, 22).

An additional difference between SVD and AC is that only the former provides direct information on the size of the chemical shift variations associated with each residue in networks I and II. For instance, Fig. S7a shows that for the residues in cluster I, the |PC2| projection of the scores correlate well with the maximum chemical shift variations observed across the five states shown in Fig. 2a. Fig. S7b shows that a similar correlation is obtained between the maximum chemical shift variations observed for residues in clusters II and the |PC1| projection of the respective scores. While these correlations indicate that SVD provides a more direct read out of the magnitudes of chemical shift variations than AC, they also mean that the SVD results are possibly dominated by outliers with large chemical shift variations. For instance, Fig. S7c shows that for the whole ensemble of 100 residues assigned in all five states, 84% of the total ppm variance is dominated by only two residues, G269 and A280. It is therefore critical to verify that the SVD-based conclusions are not biased by these two residues alone. Fig. S8a shows the score and loading plots obtained after eliminating G269 and A280 from the SVD analysis. The main change relative to Fig. 4 is that the first two PCs, which still account for more than 97 % of the total variance (Table S2), have been swapped, with cluster I and II now distributed along PC1 and PC2, respectively (Fig. S8a). This change in Fig. S8a relative to Fig. 4 reflects the convention of defining the first PC as that corresponding to the highest variance, which in the case of Fig. S8a corresponds to the allosteric network. However, the main conclusions based on the SVD analysis of Fig. 4 are still valid and similar to those drawn from Fig. S8a.

A further implication of the correlation between the PC sizes and the magnitudes of the ppm variations (Fig. S7a and S7b) is that SVD provides a reliable separation of binding *vs.* allosteric contributions only for the residues with the largest ppm changes. Residues with smaller ppm variations tend to cluster in the proximity of the origin of the score plot (Fig. 4, inset), where SVD alone does not provide a clear assessment of binding or allosteric effects. This limitation is a direct result of SVD being a variance-based method and in principle it could be circumvented by applying SVD after matrix **M** is not only row mean centered but also row standard deviation scaled (*i.e.* normalized **M**). We therefore applied the SVD factorization to the row mean centered and row standard deviation scaled **M** and the resulting scores and loadings plots are shown in Fig. S8b, in which the residues of clusters I and II are marked in red and black, respectively. Fig. S8b shows that, unlike Fig. 4, due to the row normalization of matrix **M**, the scores for the residues affected by minor chemical shift changes now have been moved away from the origin of the PC1/PC2 plane and are shifted toward the scores of other residues belonging to the same cluster and associated with major chemical shift variations (Fig. S8b). However, the sought-after orthogonality between the scores of clusters I and II, as observed in Fig. 4, is lost in Fig. S8b. Therefore the biophysical interpretation of the first two principal components in terms of binding and allostery is

now not as straightforward as in Fig. 4, in which the two clusters align well with the first two PCs. Similar problems are encountered when the SVD is applied to the row standard deviation scaled matrix **M'**. In addition, the two principal components displayed in Fig. S8b account for only 78.5 % of the total variance (*i.e.* 54.7 % + 23.8 %). For these reasons, we opted for agglomerative-clustering (Fig. S2 and Fig. 3) of the correlation matrix **R**, which includes the normalization, rather than for the SVD of the normalized matrix **M**.

2.12 Supplementary Information: Confirmation of the Functional Assignment of the Chemical Shift Covariance Networks through Mutational and Co-Evolutionary Analyses

Mutations at L273, F300 and E308 are known to affect primarily the degree of cAMP-dependent activation, based on measured $k_{max,relative}$ and AC_{50} values (Table S3) (19, 20). The activation data of Table S3 are in full agreement with the results of CHESCA that assigns L273, F300 and E308 to the allosteric cluster I. Furthermore, the prominent allosteric role of L273 and F300 has been highlighted also by the comparison of the apo- and cAMP-bound structures of the EPAC CBD (11, 12, 19), indicating that the van der Waals contacts between these two hydrophobic residues control the relative orientation of the PBC and the hinge helices (*i.e.* hydrophobic hinge hypothesis for EPAC activation) (S11, S14). CHESCA also assigns G238 of the β2-3 loop to the allosteric cluster I (Fig. 3a,b,f; Fig. 4). This is consistent with recent co-evolutionary analyses of conserved allosteric networks, which identified this Gly residue as an essential element of the cAMP-dependent allostery of CBDs in general (21). In summary, the SVD/AC-based functional assignments proposed here for the two CHESCA clusters are independently corroborated by the comparative analyses of EPAC mutants (19, 20, 24) and sequences (21). Additional details on the validation of the functional assignments obtained through the CHESCA approach are available in the previous SI section about the comparison of the SVD *vs.* AC chemical shift covariance analyses.

2.12 Supplementary Information: Structural Analysis of the Binding Network Defined by Cluster II

As shown in Fig. 5d, cluster II, which based on the AC and SVD analyses is mostly coupled to binding, includes, as expected, residues in the base binding region (BBR) (*i.e.* 252, 254) and in the phosphate binding cassette (PBC) (*i.e.* 274, 280). However, it is interesting to note that several sites of this network extend also beyond the immediate cAMP binding site (Fig. 5d). For instance, this is the case for residues

281-284 in β7 that bridge the PBC to the BBR (Fig. 5d) and for Q234 that connects the N-terminus of the β2-3 loop to the BBR (Fig. 5d). These β7 and β2-3 residues belong to cluster II but are formally outside the PBC and BBR, showing that the loci coupled to binding extend beyond the cAMP binding site identified solely based on the structure of the effector-bound CBD.

2.13 Supplementary References

1. Cavanagh J, Fairbrother WJ, Palmer III AG, Skelton NG (1996) *Protein NMR spectroscopy : principles and practice,* (Academic Press, San Diego), pp 290-298.

2. Aden J, Wolf-Watz M (2007) NMR identification of transient complexes critical to adenylate kinase catalysis. *J Am Chem Soc* 129: 14003-14012.

3. Velyvis A, Schachman HK, Kay LE (2009) Assignment of ile, leu, and val methyl correlations in supra-molecular systems: An application to aspartate transcarbamoylase. *J Am Chem Soc* 131: 16534-16543.

4. Das R*, et al* (2008) Entropy-driven cAMP-dependent allosteric control of inhibitory interactions in exchange proteins directly activated by cAMP. *J Biol Chem* 283: 19691-19703.

5. Das R*, et al* (2009) Dynamically driven ligand selectivity in cyclic nucleotide binding domains. *J Biol Chem* 284: 23682-23696.

6. Delaglio F*, et al* (1995) Nmrpipe - a multidimensional spectral processing system based on unix pipes. *J Biomol NMR* 6: 277-293.

7. Goddard TD, Kneller DG (2006) SPARKY - NMR Assignment and Integration Software, University of California, San Francisco, CA

8. Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* 34: 93-158.

9. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graph* 14: 51-55.

10. The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC

11. Rehmann H, et al (2008) Structure of Epac2 in complex with a cyclic AMP analogue and RAP1B. *Nature* 455: 124-127.

12. Rehmann H, Das J, Knipscheer P, Wittinghofer A, Bos JL (2006) Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state. *Nature* 439: 625-628.

13. Skrynnikov NR, Dahlquist FW, Kay LE (2002) Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. *J Am Chem Soc* 124: 12352-12360.

14. de Hoon MJL, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20: 1453-1454.

15. Saldanha AJ (2004) Java treeview-extensible visualization of microarray data. *Bioinformatics* 20: 3246-3248.

16. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863-14868.

17. Alter O, Brown PO, Botstein D (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci U S A* 97: 10101-10106.

18. Trbovic N, Smirnov S, Zhang F & Bruschweiler R (2004) Covariance NMR spectroscopy by singular value decomposition. *J Magn Reson* 171: 277-283.

19. Rehmann H, et al (2003) Structure and regulation of the cAMP-binding domains of Epac2. *Nat Struct Biol* 10: 26-32.

20. Rehmann H, Rueppel A, Bos JL, Wittinghofer A (2003) Communication between the regulatory and the catalytic region of the cAMP-responsive guanine nucleotide exchange factor epac. *J Biol Chem* 278: 23508-23514.

21. Kannan N, *et al* (2007) Evolution of allostery in the cyclic nucleotide binding module. *Genome Biol* 8: R264.

22. Su Y, *et al* (295) Regulatory subunit of protein kinase A: Structure of deletion mutant with cAMP binding domains. *Science* 269: 807-813.

23. Rehmann H, Wittinghofer A, Bos JL (2007) Capturing cyclic nucleotides in action: Snapshots from crystallographic studies. *Nat Rev Mol Cell Biol* 8: 63-73.

24. Rehmann H, Schwede F, Doskeland SO, Wittinghofer A, Bos JL (2003) Ligand-mediated activation of the cAMP-responsive guanine nucleotide exchange factor epac. *J Biol Chem* 278: 38548-38556.

25. Gavina JM, Mazhab-Jafari MT, Melacini G, Britz-McKibbin P (2009) Label-free assay for thermodynamic analysis of protein-ligand interactions: A multivariate strategy for allosteric ligand screening. *Biochemistry* 48: 223-225.

26. Christensen AE, *et al* (2003) cAMP analog mapping of Epac1 and cAMP kinase. discriminating analogs demonstrate that epac and cAMP kinase act synergistically to promote PC-12 cell neurite extension. *J Biol Chem* 278: 35394-35402.

2.14 Supplementary Figures

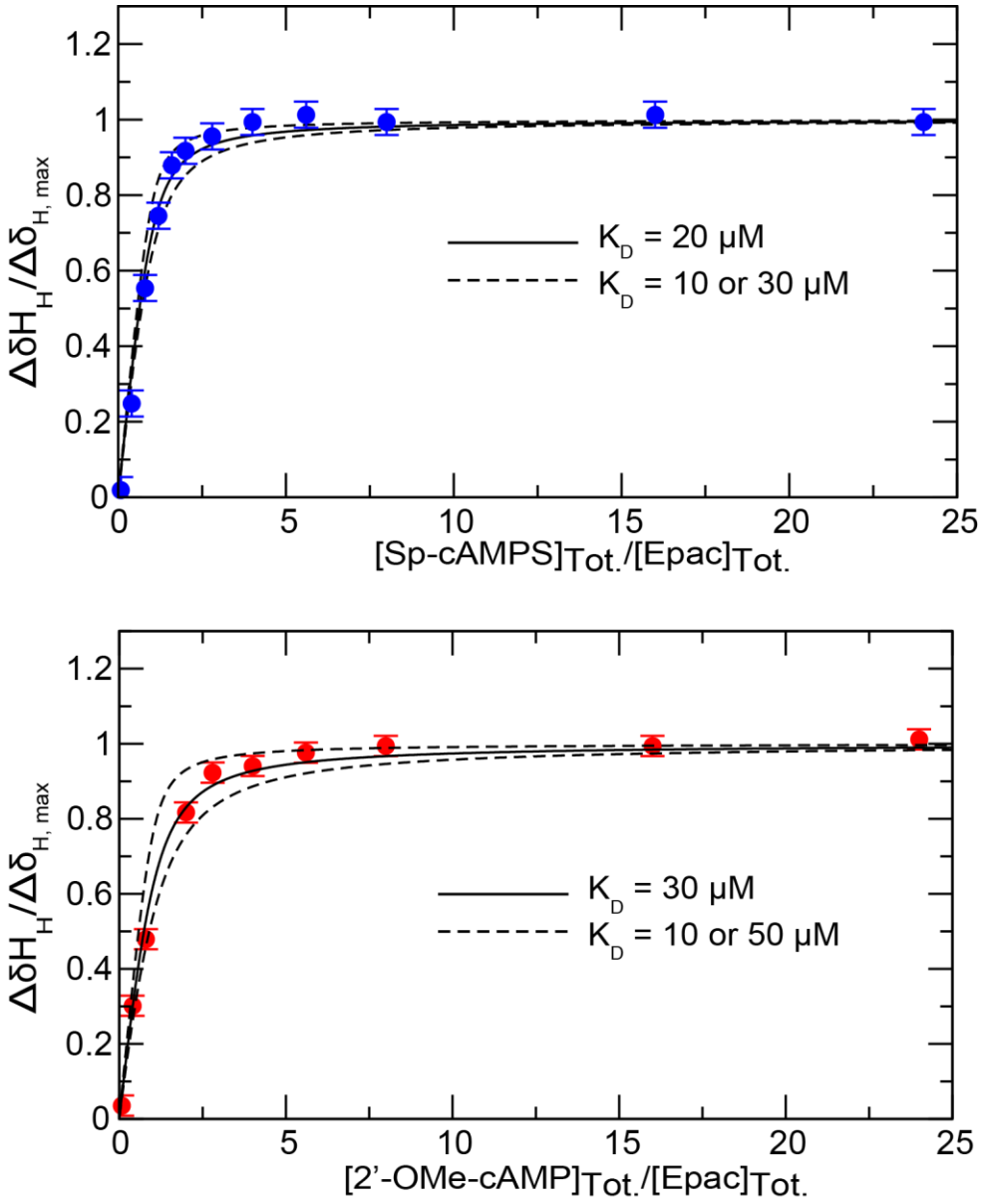

Figure S1: Proof of full saturation based on the binding isotherms for the intermediate affinity ligands Sp-cAMPS and 2'-OMe-cAMP (Table S1), measured though the relative chemical shift changes of L207, which is subject to a fast apo-holo exchange regime.

Flow Chart for the Chemical Shift Covariance Analysis (CHESCA)

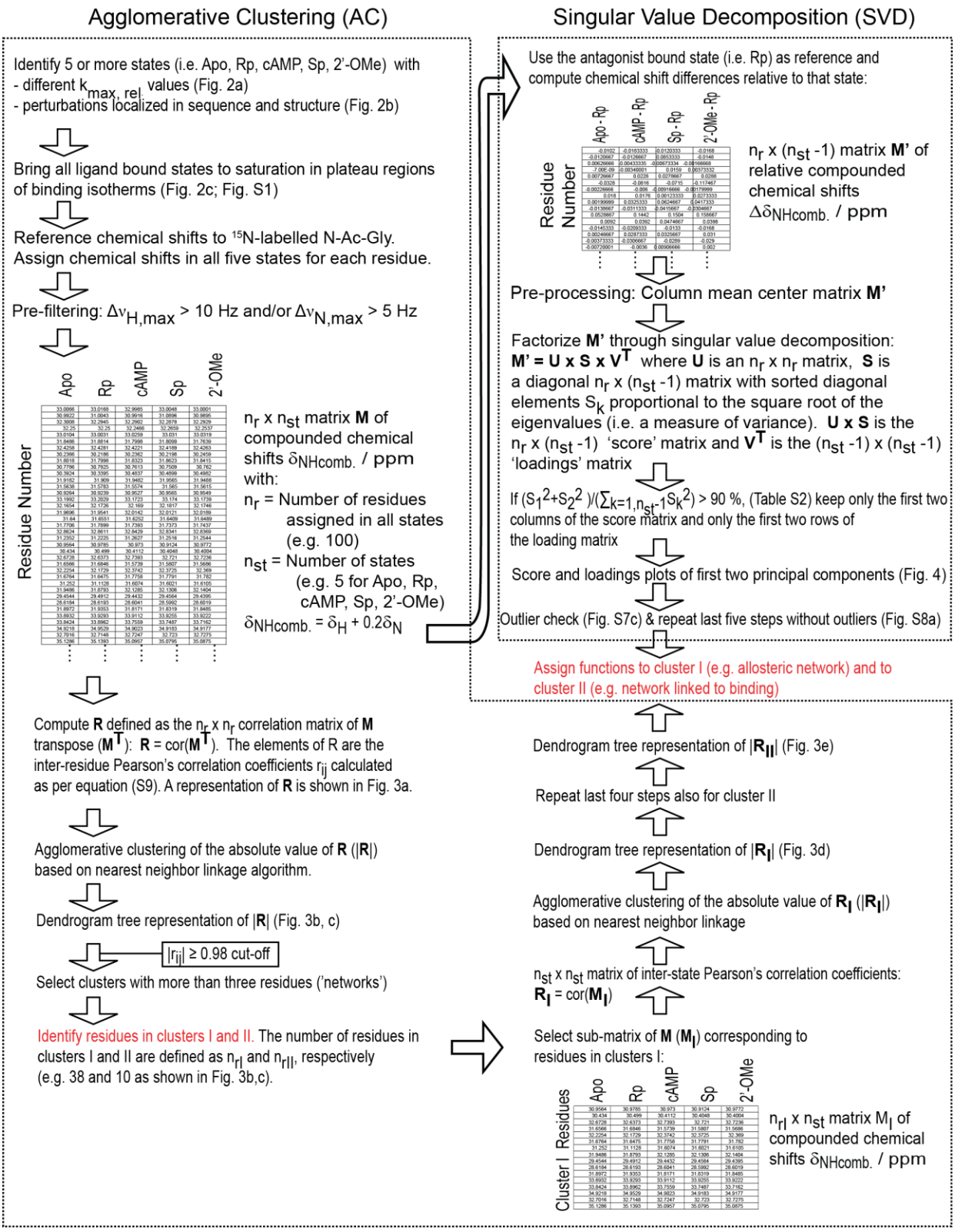

Figure S2: Flow chart (on previous page) summarizing the general CHEmical Shift Covariance Analysis (CHESCA) protocol based on agglomerative clustering of matrix **R** = cor(**M**^{T}) and on singular value factorization. **M** is the chemical shift matrix. The main results of this combined analysis are highlighted in red: (1) the identification of residues in clusters I and II; (2) the functional assignment of clusters I and II. Rp, Sp and 2'-OMe stand for Rp-cAMPS, Sp-cAMPS and 2'-OMe-cAMP analogs (Fig. 2a), respectively. Standard deviations for each principal component (PC) were computed as $S_k/\sqrt{n_r}$ with k = 1 or 2. S_k and n_r are defined in this figure.

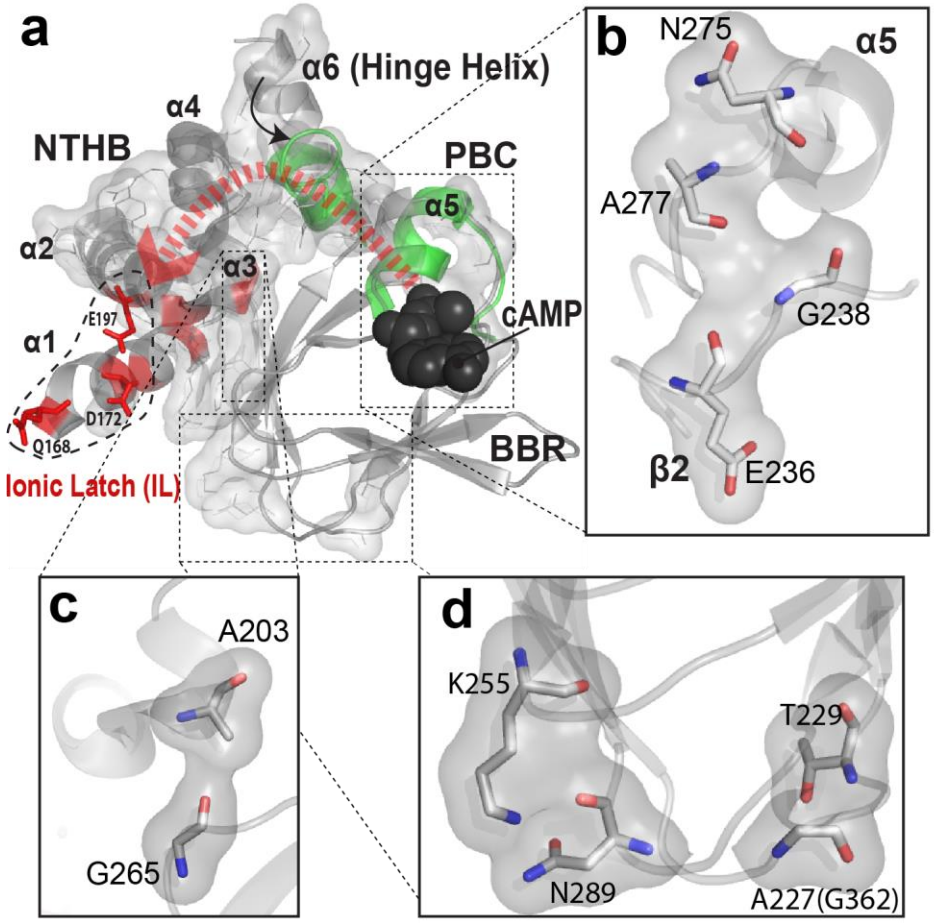

Figure S3: (a) Map of all cluster I residues into the EPAC CBD structure. This figure is similar to Fig. 5a, but it also includes the isolated pairs that are non-contiguous with the rest of cluster I, as shown in expansions **(c)** and **(d)**. Panel **(b)** shows a zoomed in view of the cluster I residues displayed in Fig. 5a and located in the PBC and the β2-β3 loop. Specifically, E236 and G238 in the β2-β3 loop are in the vicinity of cluster I PBC residues N275 and A277 (Fig. S3b), which are subject to significant cAMP-dependent conformational changes (Fig. 1c). EPAC2 residues that differ from the aligned residues in the EPAC1 isoform are reported in parenthesis.

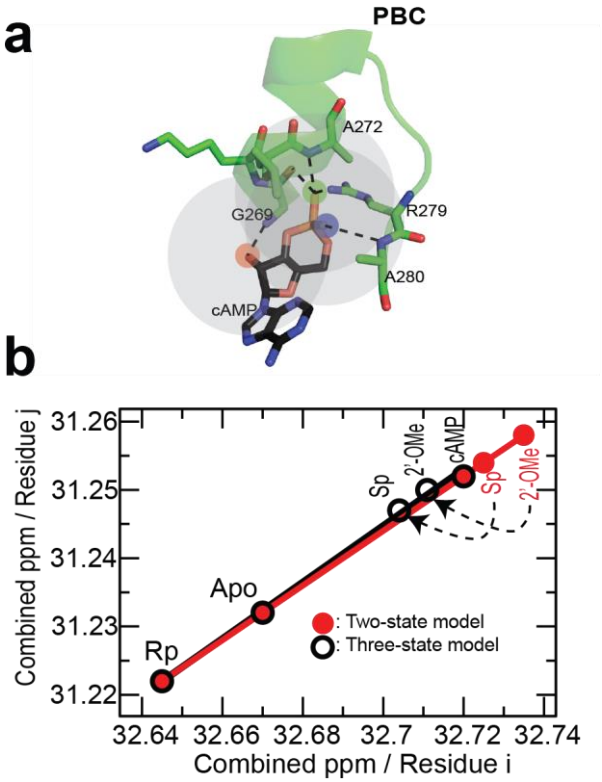

Figure S4: (a) Schematic representation of the 'perturbation-confinement' approach aimed at maximizing the overlap of the 'dark-zones' (large gray circles) surrounding the sites of covalent modification (smaller circles color coded as in Fig. 2a,b). The larger the overlap of the 'dark-zones' surrounding each perturbation locus, the lower the number of spins affected by neighboring effects (*i.e.* the higher the number of spins for which equation (1) applies). **(b)** Effect of a three-state model on the chemical-shift ordering observed in linear pairwise inter-residue correlations. This figure is based on the data of Table S5. See SI Text for further details.

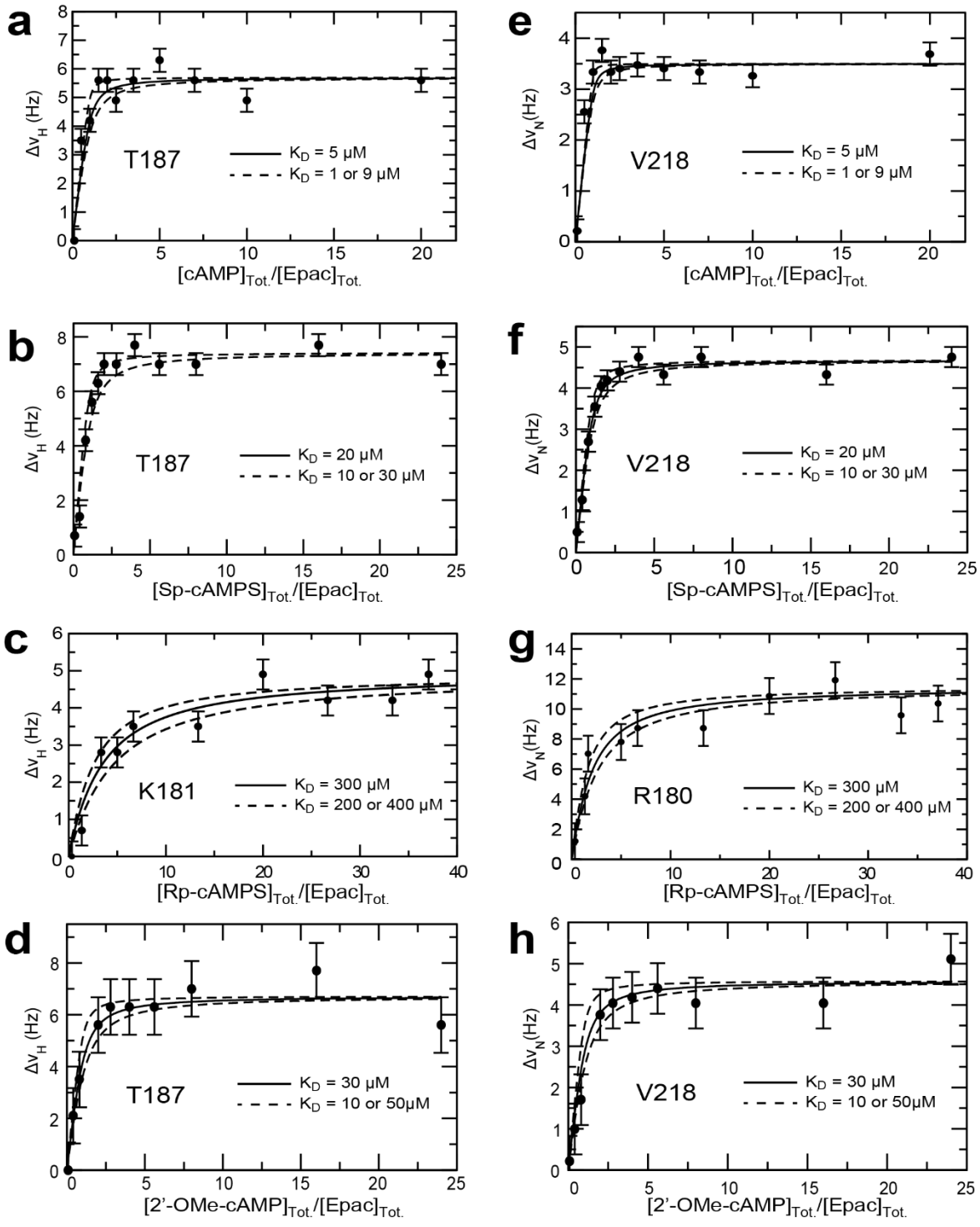

Figure S5: Binding isotherms measured for representative residues with some of the smallest chemical shift changes measured in this study. A clear dose-response pattern is observed with a well-defined plateau region, which is a signature of specific binding. **(a-d)** Representative ^{1}H frequency changes as a function of cAMP, Sp-cAMP, Rp-cAMPS and 2'-OMe-cAMP relative concentrations, respectively. **(e-h)** As panels (a-d) but for representative ^{15}N frequency variations.

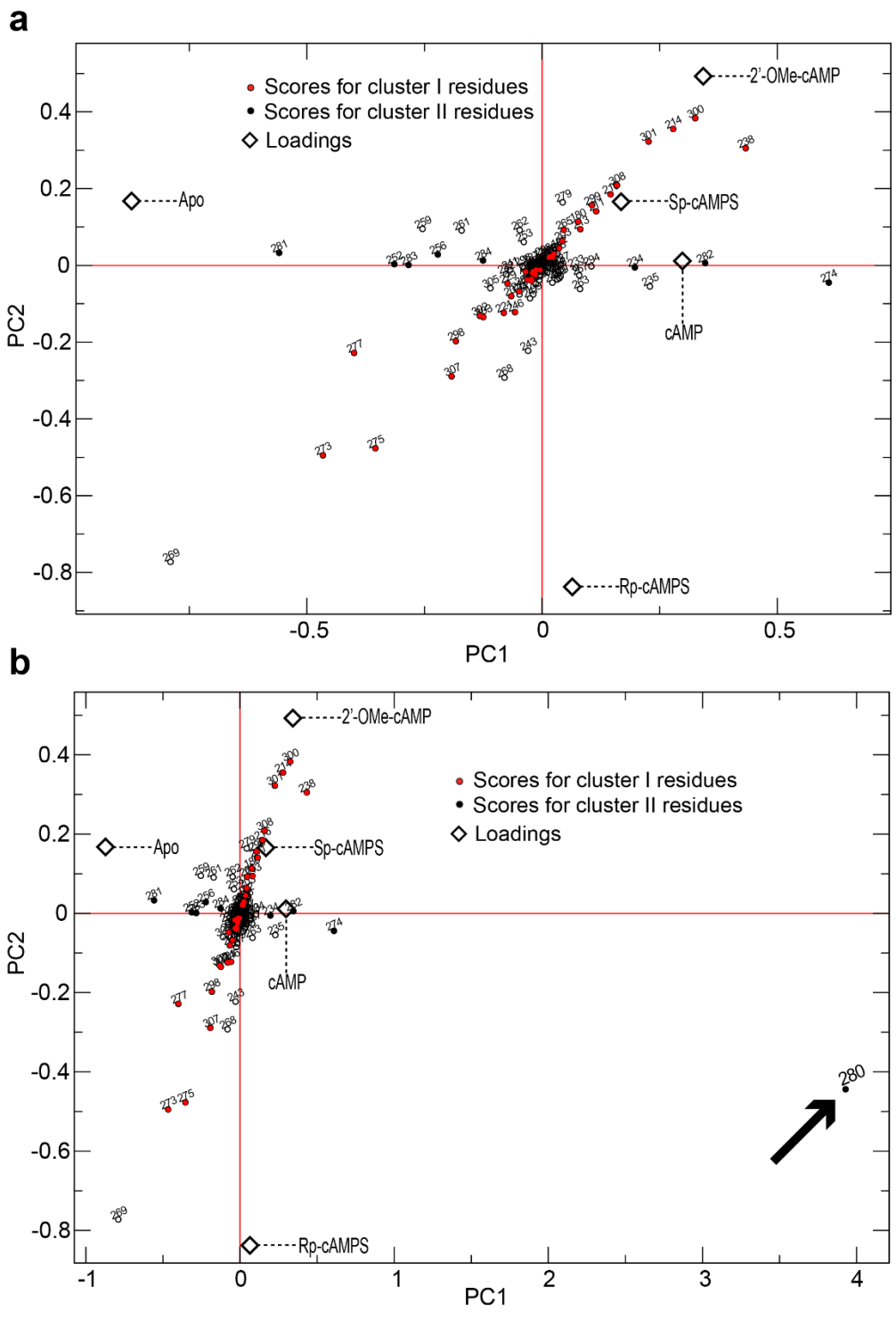

Figure S6: Scores (circles) and loadings (diamonds) plots obtained through SVD of the n_r x n_{st} matrix **M** (Fig. S2) after row mean centering. **(a)** Expansion of region including all residues but the A280 outlier; **(b)** Full plot including also A280 (highlighted by an arrow). Scores and loadings are labeled with the corresponding amino acid number and state, respectively. Cluster I (II) scores are in red (black).

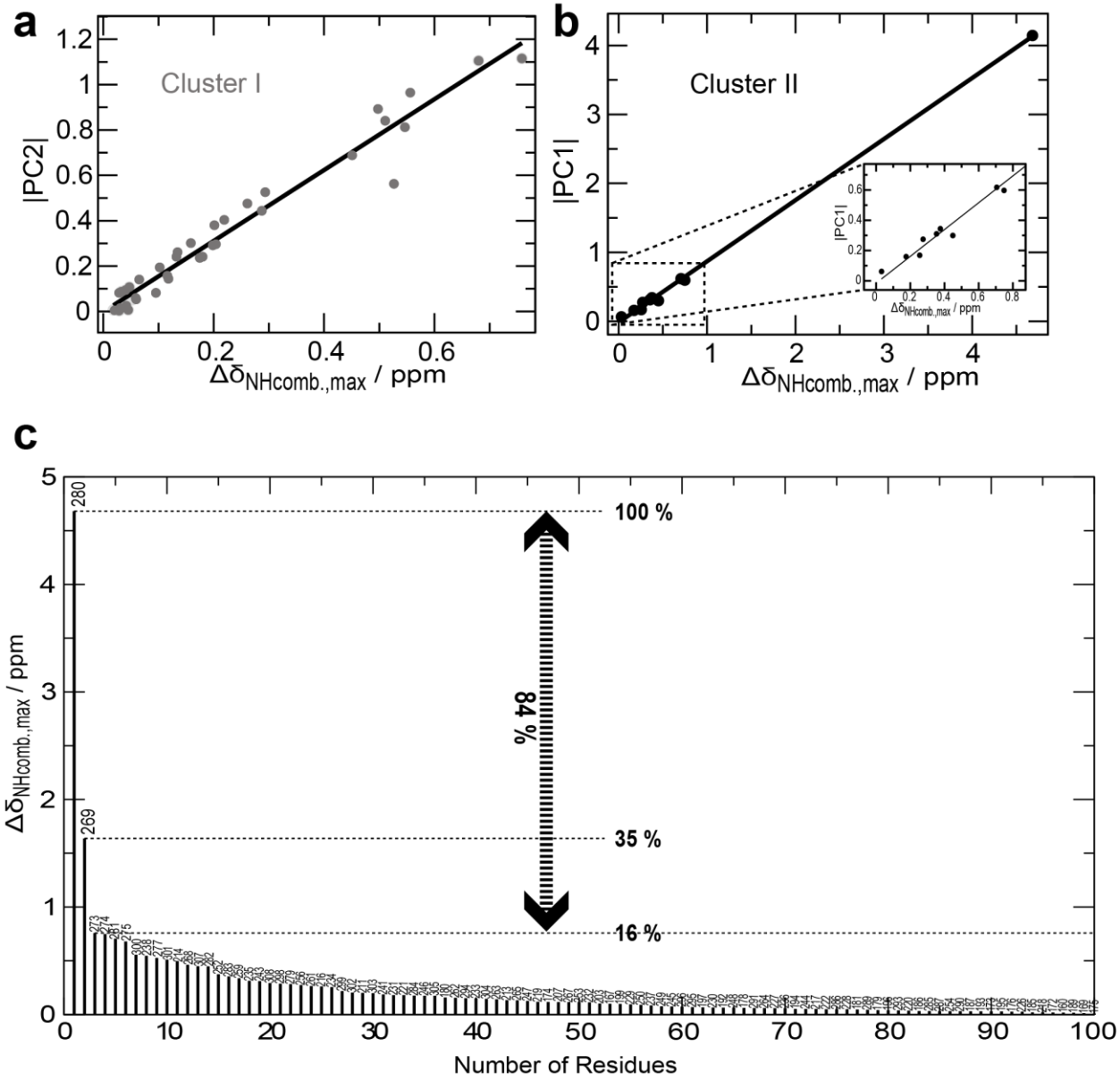

Figure S7: (a, b) Correlations between the absolute values of the second and first principal component scores and the maximum variation in the combined chemical shifts for each residue in clusters I and II, respectively. The slopes in panels (a) and (b) are 1.56 and 0.89, respectively. In both cases the correlation coefficients are ≥ 0.98. The inset in panel (b) shows an expansion of the correlation for cluster II in the 0-1 combined ppm range. **(c)** Sorted distribution of maximum variations in compounded chemical shifts measured for each assigned residue across the five measured states (*i.e.* apo, cAMP, Sp-cAMPS, Rp-cAMPS and 2'-OMe-cAMP saturated states). Residue labels are displayed. The distribution reveals that 84 % of the total variance is accounted for by only 2 % of the residues, *i.e.* two out of a total of 100 residues assigned in all five states. These two residues are G269 and A280, located at the PBC N- and C-termini, respectively. These two residues are therefore likely to dominate and possibly bias the variance probed through singular value decomposition (SVD). To check for this possibility the SVD was then repeated after discarding G269 and A280, as shown in Figure S8a.

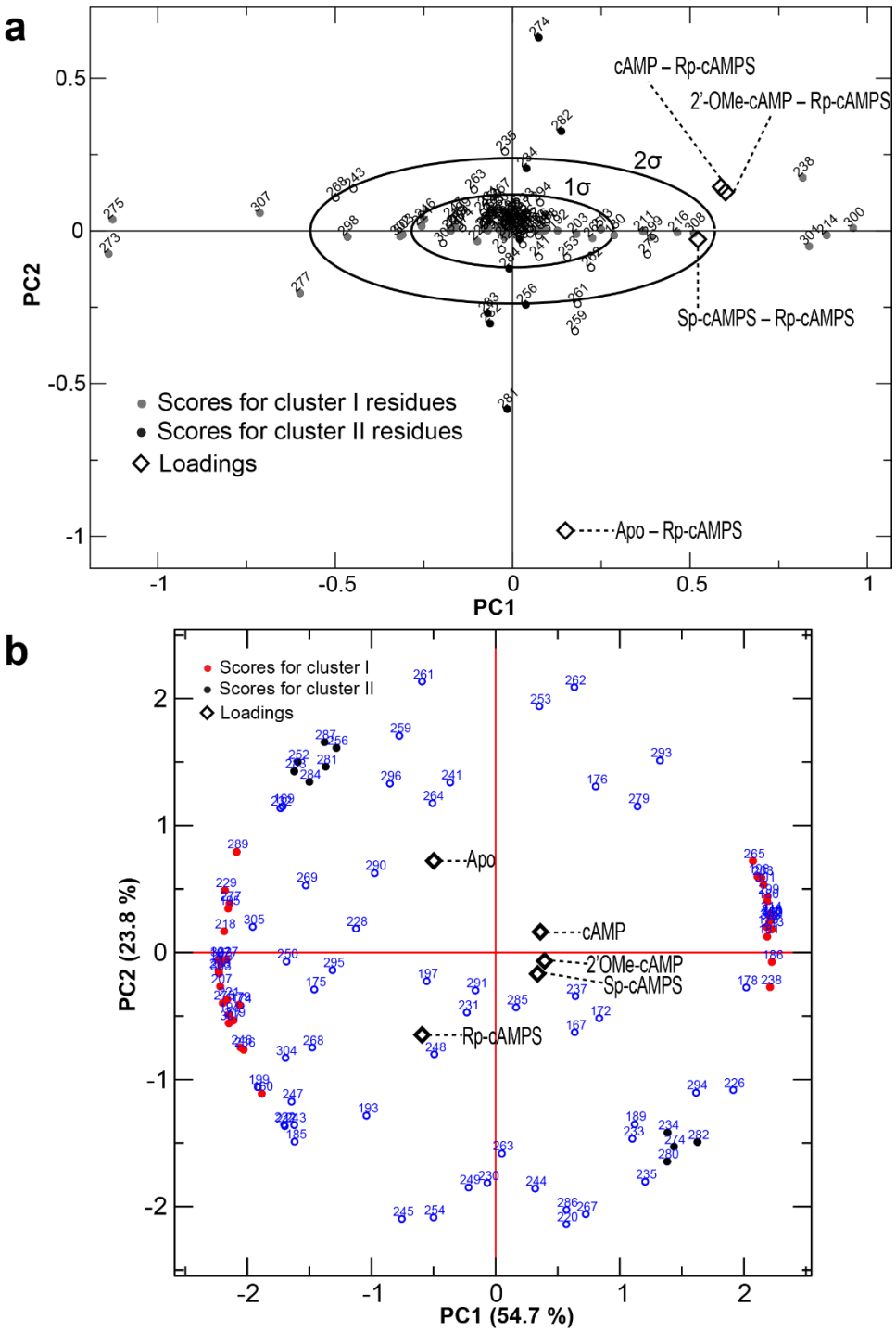

Figure S8: (a) Scores and loadings plots displayed as in Figure 4, but the SVD was computed after eliminating the rows corresponding to G269 and A280 from matrix **M'** in Figure S2. **(b)** Scores (circles) and loadings (diamonds) plots obtained through SVD of the n_r x n_{st} matrix **M** (Fig. S2) after row mean centering and row standard deviation scaling. Scores and loading are labeled similarly to Figure 4. Black and red filled circles represent scores of clusters I and II, respectively. All other scores are shown as open circles. The percentages of the total variance contributed by PC1 and PC2 are reported in parenthesis along the two axes.

2.15 Supplementary Tables

Table S1: EPAC1 Ligand Dissociation Constants

Ligands	***K_D (From Literature)[a]***	***K_D Measured Here[b]***
cAMP	3-5 µM	5 ± 4 µM
Sp-cAMPS	28-34 µM	20 ± 10 µM
2'-OMe-cAMP	> 10 µM	30 ± 20 µM
Rp-cAMPS	305-340 µM	350 ± 100 µM

[a]From SI references (25 and 26); [b]Based on the binding isotherms of Fig. 2c and Fig. S1.

Table S2: Principal Component Breakdown of Total Variance

	Percentage of Total Variance	
Principal Component (PC)	***All Assigned Residues (100)***	***G269 and A280 Outliers Excluded***
PC1	62.3 % (62.3 %)[a]	83.0 % (83.0 %)[a]
PC2	34.4 % (96.7 %)[a]	14.5 % (97.5 %)[a]
PC3	2.7 % (99.4 %)[a]	1.7 % (99.2 %)[a]
PC4	0.6 % (100.0 %)[a]	0.8 % (100.0 %)[a]

[a] The percentage reported in parenthesis reflects the cumulative contributions of the PCs in a given row and above.

Table S3: Effect of Mutants on the cAMP-Dependent Activation of EPAC

Epac1 Construct	***$k_{max,rel}$***	***AC_{50}***
Wt	1[a]	45 µM[a]
L273W	0[b]	N/A
F300A	N/A	20 µM[b]
F300T	N/A	6 µM[b]
E308A	3.0[a]	15 µM[a]

[a]From SI reference (20); [b]From SI reference (19).

Table S4: Residues in Cluster I and II Computed with Different Cut-offs

$\|r_{ij}\| \geq 0.97$		$\|r_{ij}\| \geq 0.98$		$\|r_{ij}\| \geq 0.99$	
Cluster I	*Cluster II*	*Cluster I*	*Cluster II*	*Cluster I*	*Cluster II*
178*	284**	179‡	284**	219	252
185*	282**	160‡	282**	194	283
222*	252	236‡	252	196	256
226*	283	246‡	283	203	234
217*	256	265‡	256	301	281
243*	234	277‡	234	307	274
174*	281	238‡	281	221	280
179	274	229‡	274	181	287
160	280	289‡	280	227	
236	287	219	287	273	
246		194		195	
265		196		218	
277		203		186	
238		301		302	
229		307		173	
289		221		180	
219		181		299	
194		227		187	
196		273		275	
203		195		213	
301		218		211	
307		186		214	
221		302		216	
181		173		192	
227		180		207	
273		299		300	
195		187		298	
218		275		303	
186		213		308	
302		211		219	
173		214			
180		216			
299		192			
187		207			
275		300			
213		298			
211		303			
214		308			
216					
192					
207					
300					
298					
303					
308					

*Residues that are not part of cluster I when $|r_{ij}| \geq 0.98$; ** Residues that are not part of cluster II when $|r_{ij}| \geq 0.99$; ‡Residues that are not part of cluster I when $|r_{ij}| \geq 0.99$.

Table S5: Example of Populations in the Three-State Model

	Populations		
Perturbation (*k*)	**Active State** (p'_k)	**Third State** (p''_k)	**Inactive State** $(1 - p'_k - p''_k)$
Rp	0.05	0.00	0.95
apo	0.30	0.00	0.70
cAMP	0.80	0.00	0.20
Sp	0.15	0.70	0.15
2'-OMe	0.15	0.80	0.05

Chapter 3

The Projection Analysis of NMR Chemical Shifts Reveals Extended EPAC Auto-inhibition Determinants

3.1 Author's Preface

The work presented in this chapter has previously been published and is reproduced here with permission from Elsevier. Full citation is as follows:

Selvaratnam, R., B. VanSchouwen, F. Fogolari, M.T. Mazhab-Jafari, R. Das, et al. 2012. The projection analysis of NMR chemical shifts reveals extended EPAC autoinhibition determinants. Biophysical journal. 102: 630–9.

I conducted most of the experiments necessary for the manuscripts, including the backbone assignment of all EPAC mutants. Bryan VanSchouwen performed and analyzed all MD simulations. Mohammad T. Mazhab-Jafari and Dr. Rahul Das assigned the wild type EPAC and generated the E308A mutation. I co-wrote the manuscript with Bryan VanSchouwen and Dr. Giussepe Melacini.

3.2 Abstract

EPAC is a cAMP dependent guanine nucleotide exchange factor that serves as a prototypical molecular switch for the regulation of essential cellular processes. While EPAC activation by cAMP has been extensively investigated, the mechanism of EPAC auto-inhibition is still not fully understood. The steric clash between the side chains of two conserved residues, L273 and F300 in EPAC1, has been previously shown to oppose the inactive-to-active conformational transition in the absence of cAMP. However, it has also been hypothesized that auto-inhibition is assisted by entropic losses caused by quenching of dynamics that occurs if the inactive-to-active transition takes place in the absence of cAMP. Here, we test this hypothesis through the comparative NMR analysis of several EPAC1 mutants that target different allosteric sites of the cAMP-binding domain (CBD). Using a novel projection analysis of NMR chemical shifts to probe the effect of the mutations on the auto-inhibition equilibrium of the CBD, we find that whenever the apo/active state is stabilized relative to the apo/inactive state, dynamics are consistently quenched in a conserved loop (β2-β3) and helix (α5) of the CBD. Overall, our results point to the presence of conserved and non-degenerate determinants of CBD auto-inhibition that extend beyond the originally proposed L273/F300 residue pair, suggesting that complete activation necessitates the simultaneous suppression of multiple auto-inhibitory mechanisms, which in turn confers added specificity for the cAMP allosteric effector.

3.3 Introduction

The exchange protein directly activated by cAMP (EPAC) is a guanine nucleotide exchange factor (GEF) for the small GTPases, Rap1 and Rap2 (1-5), and is a key regulator of several physiological responses with implications in chronic degenerative inflammatory diseases (6). EPAC consists of an N-terminal regulatory region (RR) and a C-terminal catalytic region (CR). The RR of EPAC1 contains a single cAMP binding domain (CBD), while the CR includes a CDC25-homology domain (CDC25-HD), which binds the Rap substrate (Fig. 1a) (1-5).

The EPAC CBD is an archetype of other eukaryotic regulatory CBDs, such as those of Protein Kinase A (PKA) (7-15) and of the hyperpolarization-activated cyclic nucleotide modulated (HCN) channels (16). The eukaryotic CBDs share a common architecture with a β-subdomain of jelly-roll β-sheets and a non-contiguous α-subdomain of α-helices. The latter includes an N-terminal helical bundle (NTHB), a C-terminal helix (α6), also known as the hinge helix, as well as a short helix (α5) located between β6 and β7 in the phosphate binding cassette (PBC), where cAMP docks (Fig. 1b,d) (8,11,12).

The orientation of α6 with respect to the PBC is controlled by cAMP. In the absence of cAMP (*i.e*. apo-state), the hinge helix is oriented away from the PBC (Fig. 1b-d), but upon cAMP binding the hinge helix rotates towards the PBC (Fig. 1b-d). In the case of EPAC, the hinge helix rotation is also coupled to a partial unfolding of the last two turns of the α6 helix (Fig. 1b,d)(3,17,18). These cAMP-dependent changes in the α-subdomain are critical for the control of auto-inhibitory interactions (3,5,7-14,17-20). Auto-inhibition pertains to the mechanisms that prevent the premature inactive-to-active transition in the absence of the allosteric effector (*i.e.* auto-inhibitory equilibria, Fig. 1a) (4,21). The maintenance of an auto-inhibited state is essential for eliciting cAMP dependence and preventing constitutive activation, which can otherwise have drastic effects on downstream signalling (4,6,21).

The cAMP-dependent hinge rotation has been previously rationalized by the so-called hydrophobic hinge hypothesis, which posits that a highly conserved contact between L273 in the PBC and F300 in α6 would lead to a steric clash if the hinge reorientation towards the β-subdomain occurred in the absence of cAMP (3-5). Such steric hindrance has been suggested to be a key determinant of auto-inhibition and is alleviated by the repositioning of L273 caused by cAMP binding (Fig. 1b,c). Furthermore, it has been hypothesized that an additional determinant of auto-inhibition arises from entropic losses in the PBC α5 and in an adjacent loop (β2-β3) caused by a reduction of dynamics and tighter L273/F300 packing in the apo/active *vs.* apo/inactive states of the CBD (Fig. 1a) (21). However, this hypothesis has not been tested, as currently no experimental data is available for the apo/active intermediate due to its transient nature.

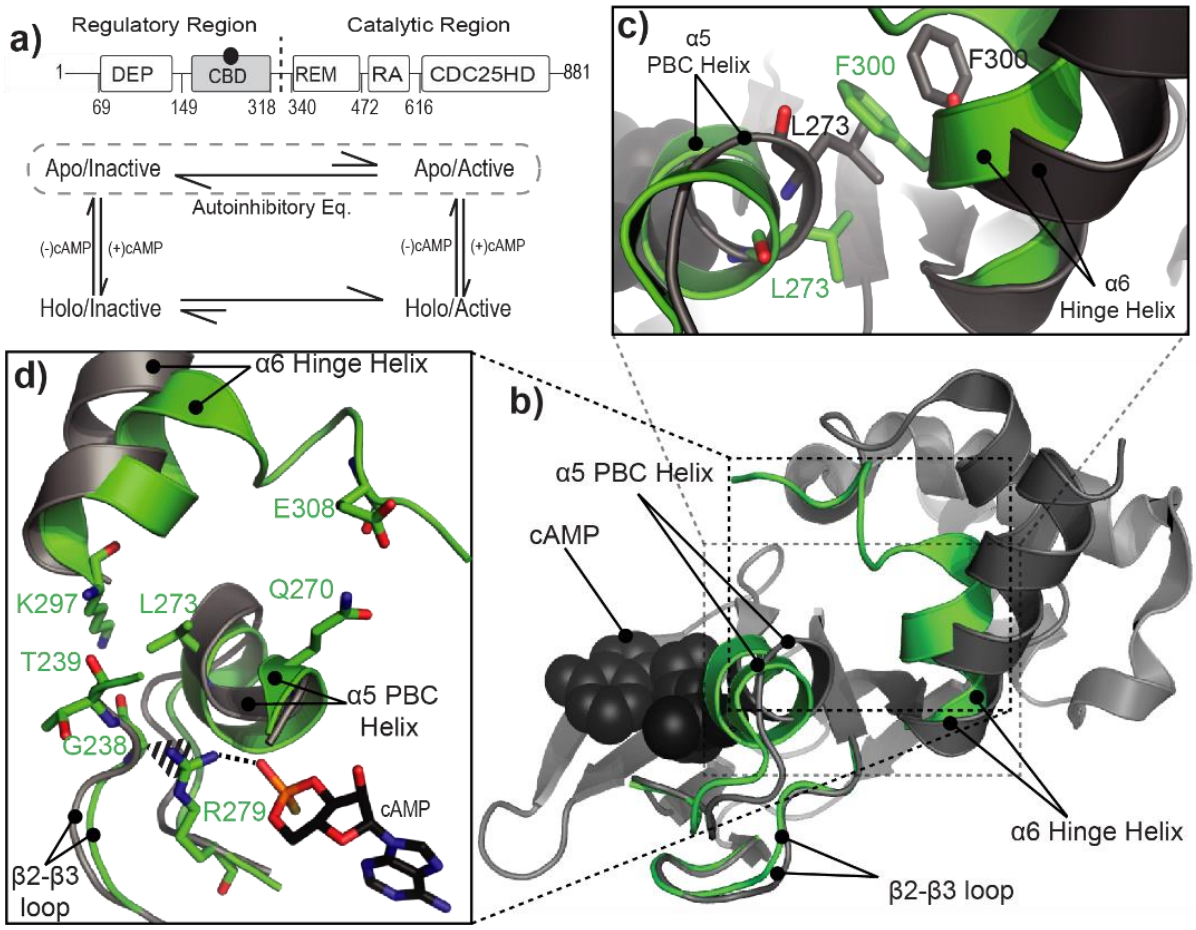

Figure 1: a) Domain organization of EPAC1 and thermodynamic cycle for the coupled cAMP binding and activation equilibria. The catalytic region (CR) consists of the CDC25-homology domain (CDC25-HD), the Ras exchange motif (REM) and the Ras association (RA) domain. The regulatory region (RR) contains the Dishevelled Egl-10 Pleckstrin (DEP) domain and the cyclic-nucleotide Binding Domain (CBD). The CBD under study here is coloured in grey; the black circle represents the cAMP molecule. The vertical black dashed line represents the separation between the RR and CR. The auto-inhibitory equilibrium between the apo/inactive and apo/active states is encircled with a dashed line. **b)** Ribbon diagram of the apo-CBD structure is shown in (grey), with the regions subject to major cAMP-dependent changes superimposed in green with the corresponding labels. cAMP is represented with black spheres. **c)** Expanded view of the L273/F300 pair in the apo (grey) and holo (green) structures. **d)** Contacts mediated by the β2-β3 loop and the PBC. The PBD ID codes for the apo (grey) and holo (green) structures are 2byv and 3cf6, respectively.

Here, we test the hypothesis that the PBC α5 and β2-β3 loop dynamics is quenched in the apo/active conformation relative to the apo/inactive state by investigating three single point mutations that stabilize the transient apo/active state by targeting the alloseric 'hot spots' of the EPAC CBD (*i.e.* G238A in the β2-β3 loop, Q270A in the α5 helix and E308A in the α6 hinge helix; Fig. 1d). The effect of each mutation on the active/inactive equilibrium of the CBD was evaluated by NMR in the apo state, since the auto-inhibited state occurs in the absence of cAMP. Specifically, we used NMR to not only probe at atomic resolution

the changes in dynamics caused by the mutation, but also to gauge how the mutation affects the position of the apo/inactive *vs.* apo/active equilibrium. For this purpose, we propose a method based on ^{1}H and ^{15}N NMR chemical shifts, which are ideally suited to sense changes in the position of the apo/inactive *vs.* apo/active equilibrium of EPAC in the fast exchange regime, minimizing biases due to nearest neighbour effects of the mutants. This approach relies on the chemical shift projection analysis (CHESPA) and was validated using the L273W mutation, which has been shown to stabilize the apo/inactive state (5). We also complemented the NMR analyses with molecular dynamics (MD) simulations and our combined NMR and MD results indicate that whenever a mutant causes a stabilization of the apo/active state, a quench in dynamics is consistently observed for the α5 and β2-β3 loop regions. This result is in agreement with the hypothesis that the auto-inhibited state is stabilized not only by the steric hindrance between L273 and F300, but also by the α5/β2-3 entropic losses occurring when the inactive-to-active transition ensues in the absence of cAMP. Overall, the combined analysis of the EPAC mutants presented here reveals a novel model of EPAC auto-inhibition, in which the stabilization of the inactive *vs*. active structures in the absence of cAMP relies on a complex network of interactions that extends well beyond the previously proposed L273/F300 contact.

3.4 Materials and Methods

Purification of Wt and Mutant EPAC1 Constructs: Purification procedures used for both the Wt and mutant EPAC1 (GenBank No. AAC83381.1) constructs were as previously explained (20). Briefly, cells of *E.coli* strain BL-21 harbouring plasmids encoding EPAC1 constructs were grown in isotopically enriched media with $^{15}NH_4Cl$ and ^{13}C-labelled glucose. Cells were lysed as before and protein was purified using the following chromatographic columns (GE Healthcare): Glutathione Sepharose 4B, HighTrap Q HP and Superdex 75 16/60. Singly labelled ^{15}N-acetylglycine (Sigma Aldrich) (0.5mM) was added to the final concentrated protein sample (0.5mM) as a ppm referencing compound.

NMR Measurements: All spectra were acquired at 306 K with a Bruker Avance 700-MHz spectrometer equipped with a 5 mm TCI cryoprobe. Gradient and sensitivity enhanced [^{15}N-^{1}H] heteronuclear single quantum coherence (HSQC) spectra with 256 (t1) and 1024 (t2) complex points and spectral widths of 31.82 ppm and 14.06 ppm for the ^{15}N and ^{1}H dimensions, respectively, were recorded with 8 scans and a recycle delay of 1.00 s. Carrier frequencies of the proton and nitrogen channels were centered on water and the backbone amide region, respectively. All spectra were processed with NMRPipe (22) employing linear prediction in the ^{15}N dimension, and a resolution-enhancing 60° shifted squared sine bell window

function for both dimensions. All ^{15}N-^{1}H HSQC cross-peaks were analyzed with Sparky using Gaussian line-fitting (23). Assignments were obtained using triple-resonance experiments, as previously explained (17, 18). The measured $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO, ^{1}H, and ^{15}N chemical shifts were utilized to derive order parameters by employing the RCI algorithm with correct referencing (24) which reports on dynamic processes on a broad range of time scales (25). Indeed, the trends captured by the RCI-based S^2 values are in agreement with previously measured ^{15}N relaxation data for the apo-Wt and cAMP-bound Wt states (17). In addition, the RCI based order parameters have the advantage of being measurable also for the critical phosphate binding cassette (PBC) region, which in the apo-state, is subject to significant line-broadening, compromising the signal-to-noise ratio of traditional ^{15}N relaxation experiments.

MD Simulation Protocol: All simulations were executed using the EPAC2 fragment composed of amino acid residues 310-498 (corresponding to 175-363 in $EPAC1_h$). Mouse EPAC2 (GenBank No. AAD09132.1) was used for all MD simulations as this is the EPAC isoform for which experimental structures have been determined by X-ray crystallography in the absence and presence of the agonist Sp-cAMPS (3,4) and contains the functionally-essential CBD, which is highly homologous to the $EPAC1_h$ CBD (~81% sequence similarity and ~69% sequence identity). Initial structures of the apo/inactive and holo/active states were obtained from X-ray crystal structures with PDB IDs 2BYV and 3CF6, respectively (3, 4) by deleting all amino acids other than residues 310-498 (*i.e* 175-363 in $EPAC1_h$) and all water molecules from both structures. In addition, the bound Sp-cAMPS molecule was replaced with the endogenous cAMP and an extra surface-bound Sp-cAMPS molecule was deleted from the structure as it was found to be in a crystal packing contact. For clarity, the corresponding $EPAC1_h$ amino acid numbering is used in the remaining part of the manuscript, unless otherwise specified. For further details, see Supplementary Materials (Page 83).

3.4 Results

NMR chemical shift variations are sensitive reporters of the position of the inactive *vs.* active fast-exchanging equilibrium of signalling proteins, such as EPAC (Figure S1 in the Supporting Material) (20). However, the commonly used compounded ppm changes, computed as $\sqrt{((0.2\Delta\delta_N)^2+(\Delta\delta_H)^2)}$ (9,13,14,17-19) report only on the magnitude of the chemical shift variations caused by a mutation (Fig. 2, dashed circles) and not on the direction in which the given mutation affects the auto-inhibitory equilibrium. For instance, Fig. 3a shows the combined ppm changes of apo-L273W *vs.* apo-Wt $EPAC1_{149-318}$ and indicates that most of the regions subject to the largest ppm shifts caused by the L273W mutation in the apo-state

match quite well the sites displaying the most significant cAMP-dependent ppm changes for the Wt protein (Fig. 3a, grey zones) (17,18,20). However, cAMP is an activator and the L273W mutant is inactive (5). Thus the compounded ppm changes provide an effective quantification of the size of the ppm changes, but they are not sufficient to establish whether a mutation shifts the apo equilibrium towards the inactive or the active state. In order to quantify both the size and the direction of the perturbation caused by a mutation on the auto-inhibitory equilibrium, we introduce the projection analysis of NMR chemical shifts (Fig. 2). We will first illustrate the chemical shift projection analysis (CHESPA) method and then validate and apply it to the EPAC system.

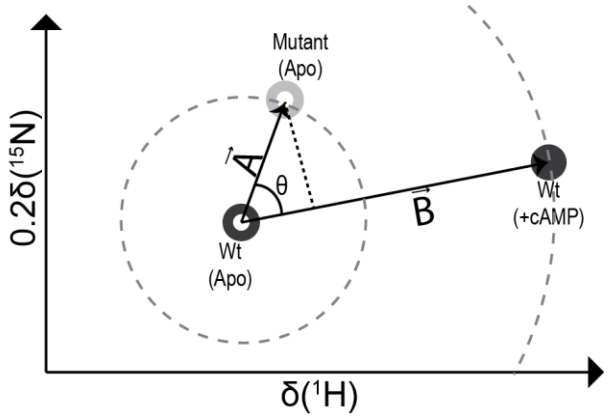

Figure 2: Schematic illustration of the projection analysis of NMR chemical shifts. The open and filled circles represent the apo-and holo-forms respectively, as labelled in the figure. Wt and mutant peaks are shown in dark and light grey, respectively. The compounded chemical shift between the apo-Mutant and the apo-Wt was calculated as the magnitude of vector A (*i.e.* |A|). Similarly the compounded chemical shift between the cAMP-bound Wt and apo-Wt was computed as |B|. The modules of vectors A and B define the radii of the dashed circles centered on the Wt-apo peak.

3.4.1 Chemical Shifts Projection Analysis (CHESPA)

The compounded chemical shift difference between the apo-Wt and the apo-mutant was calculated as the magnitude of vector A in Fig. 2 connecting the apo-Wt and apo-mutant peaks (Fig. 2) and defined in the plane of the ^{1}H and scaled ^{15}N ppm coordinates. The scaling factor of the ^{15}N ppm values is 0.2 (20). Similarly, the compounded chemical shift difference between the apo-Wt and the cAMP-bound Wt (holo-form) is computed as the magnitude of vector B in Fig. 2, which represents the 'activation vector' joining the apo/inactive to the holo/active state. The projection of vector A onto B is a measure of the shift along the 'activation vector' caused by a given mutation. To quantify the extent of activation (or inactivation)

achieved by a mutation in the absence of cAMP, the fractional shift (*X*) is calculated as the ratio of the component of vector A along vector B and the magnitude of vector B (*i.e.* |B|):

$$X = \frac{|\vec{A}|\cos\theta}{|\vec{B}|} \qquad \text{Eq. (1)}$$

where θ is the angle between vectors A and B, as shown in Fig. 2. Considering that:

$$\cos(\theta) = \frac{\vec{A} \bullet \vec{B}}{|\vec{A}||\vec{B}|} \qquad \text{Eq. (2)}$$

Equation (1) is equivalent to:

$$X = \frac{\vec{A} \bullet \vec{B}}{|\vec{B}|^2} \qquad \text{Eq. (3)}$$

The fractional shift (*X*) is a scalar and is complemented by the cos(θ) value, which reports on the relative orientation of vectors A and B. Thus the projection analysis of NMR chemical shifts results in two key residue-specific descriptors of the perturbation caused by the mutation, *i.e.* the fractional shift (*X*) and the cos(θ) (Fig. 2). Only those residues with |B| > 0.05 ppm were considered further in the projection analysis in order to filter out the fractional shifts (*X*) affected by large experimental errors. Specifically, because |B| is used to normalize the fractional shifts in Equation 3, a small value of |B| (*e.g.* |B| ≤ 0.05 ppm) would result in an amplification of the experimental error affecting the measurement of *X*. In other words, the vector B used as reference in the projection analysis must be reliable in order to result in meaningful fractional activation values. Decreasing the |B| cut-off threshold value would introduce more 'noise' in the projection analysis, whereas increasing this threshold value would filter out potentially informative residues as illustrated in the Supporting Material Table S1.

The fractional shift, *X* is positive if the mutation shifts the auto-inhibition equilibrium towards the active state and negative otherwise. The absolute value of *X* approaches 0 if the ppm variations caused by the mutation are negligible compared to those caused by cAMP binding and/or if vectors A and B are orthogonal (Fig. 2). When the mutation and cAMP result in ppm changes of comparable magnitude and direction, |*X*| ~ 1 (Fig. 2). The |cos(θ)| values approach unity (*i.e.* |cos(θ)| ~ 1) when the effect of the mutation is modeled exclusively based on the two-state active/inactive equilibrium, irrespective of the magnitude of the mutation-induced ppm changes, while |cos(θ)| < 1 for residues that are more significantly affected by the mutation through nearest neighbour effects or other structural distortions caused by the mutation and not fully modeled by a shift in the two-state active/inactive equilibrium. Consequently, the presence of more than two states in the fast exchange regime would compromise the linear chemical shift patterns (Fig. S1) that are exploited by the CHESPA method. This lack of linearity

would lead to |cos(θ)| values below unity. Thus, only residues with |cos(θ)| values approaching unity are suitable reporters of the position of the two state auto-inhibitory equilibrium through the fractional shift (*X*). The combined use of cos(θ) and *X* will be illustrated and validated by the application of the CHESPA method to the well-characterized L273W inactive mutant.

3.4.2 The EPAC auto-inhibitory equilibrium is sensitively probed through the projection analysis of NMR chemical shifts: the L273W mutant as a test case

As a first validation of the projection analysis of NMR chemical shifts, we applied it to the apo-L273W *vs.* apo-Wt comparison (Fig. 3b,c, black bars). The L273W mutation has been shown to promote inactivation (5,21) and therefore we expect a shift of the apo-EPAC conformational equilibrium towards the inactive state. This prediction is confirmed by the projection analysis in Fig. 3b,c (black bars), as the fractional shift (*X)* is markedly negative for the residues with |cos(θ)| approaching unity (Fig. 3b,c and Fig. 4a). Furthermore, the projection analysis of L273W bound to cAMP indicates that inactivation is preserved in the holo state (Figure S2 in the Supporting Material), in full agreement with the inability of cAMP to activate the L273W mutant (5). Therefore, in contrast to compounded chemical shift profiles, the chemical shift projection analysis provides an effective tool to establish how a given mutation affects the inactive *vs*. active equilibria of the EPAC CBD. In the following sections, the projection analysis is utilized to assess the effect of point mutations on the auto-inhibitory apo-EPAC CBD equilibria.

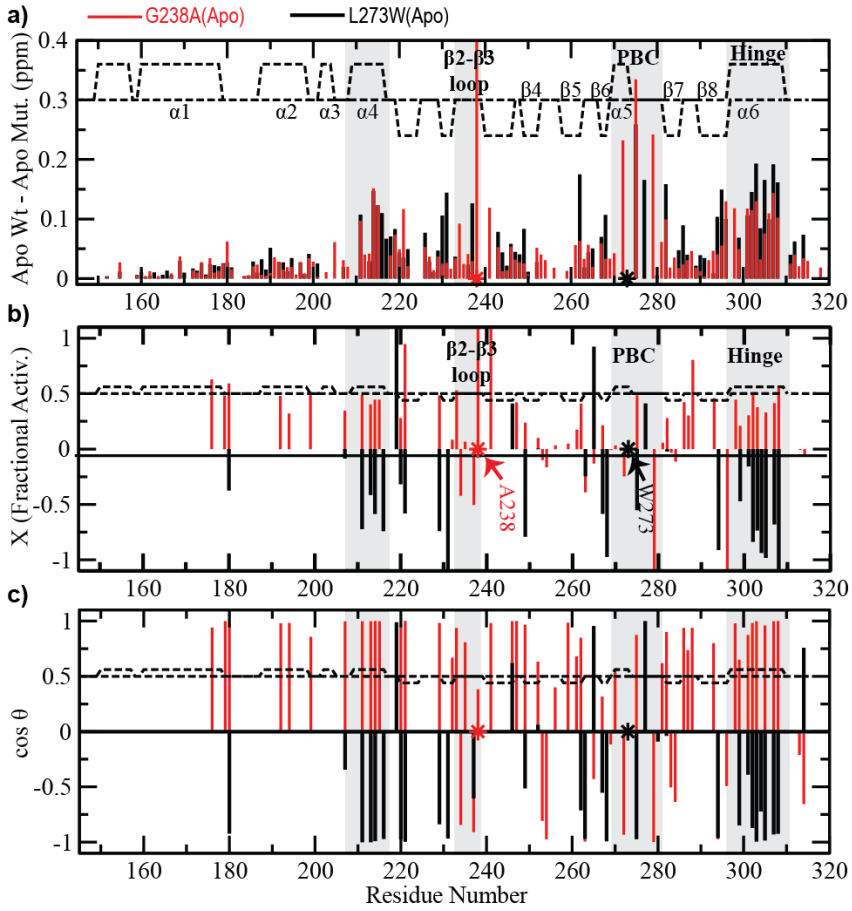

Figure 3: The effects of the L273W and G238A mutations on apo-Wt. **a)** The compounded chemical shift profile of the apo-mutant relative to apo-Wt. The secondary structure of the apo form is shown as dashed lines. The asterisks mark the sites of mutation. Highlighted grey zones denote the sites subject to some of the most significant cAMP-dependent ppm changes for the Wt protein. **b)** Fractional shift toward activation achieved by the mutation in the absence of cAMP for residues with a compounded chemical shift between the apo-Wt and holo-Wt greater than a cut-off value of 0.05 ppm. **c)** Projection angle, which is also an indicator of the direction of chemical shift movement along (positive values) or against (negative values) the activation path (*i.e.* the vector B in Fig. 2).

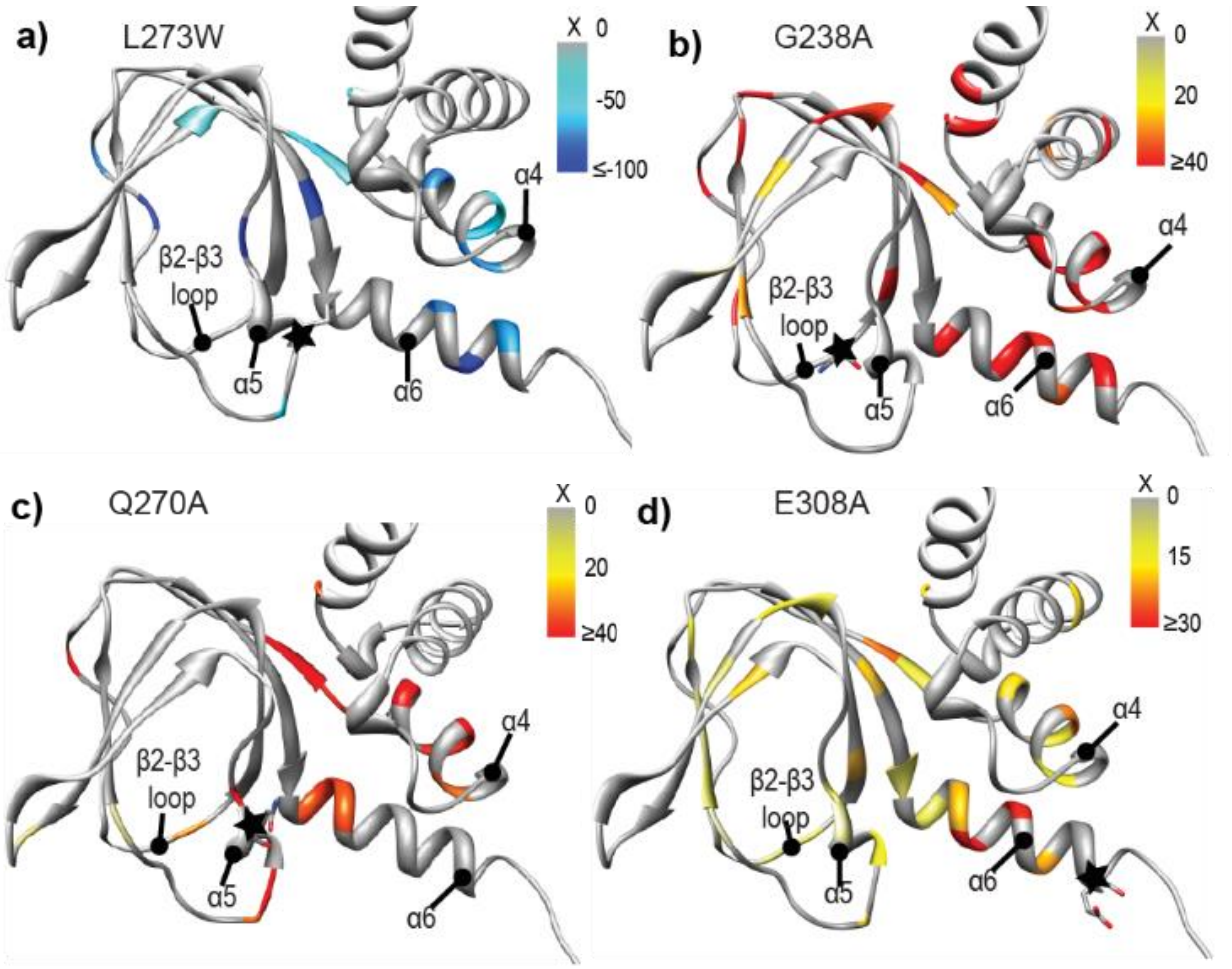

Figure 4: Structural map of the fractional chemical shift changes (X) induced by mutations in the apo-state for residues with |cos(θ)| > 0.9, as identified from Figures 3 and 7. **(a)** L273W, **(b)** G238A, **(c)** Q270A and **(d)** E308A. There are a select few residues for which the fractional chemical shift change is negative (or positive in the case of apo-L273W) when |cos(θ)| > 0.9. However, these residues cluster in the vicinity of the mutation and are excluded as they are likely affected by nearest neighbor effects and not captured by the two-state model as explained in the Results section. The mutation site is indicated by a star symbol and illustrated with sticks.

3.4.2 The projection analysis of the G238A mutant reveals that the highly conserved Gly238 of the β2-β3 loop contributes to auto-inhibition

G238 of EPAC1 was chosen as a mutation site because this residue is more conserved across CBDs than L273/F300 (11,12) and is indirectly involved in bridging contacts between the PBC and the hinge helix (Fig. 1d). In order to probe the contribution of G238 to auto-inhibition, we have purified the G238A $EPAC1_{149\text{-}318}$ mutant and have assigned and analyzed the NMR chemical shifts (Fig. S1). The compounded chemical shifts for the apo-G238A *vs.* apo-Wt comparison are shown in Fig. 3a. Based on Fig. 3a, it is clear that the G238A mutation causes significant long-range chemical shift perturbations that extend to regions

far removed from the site of the mutation (indicated by asterisk). Remarkably, G238A affects all key allosteric loci controlled by cAMP, *i.e.* the PBC, the hinge helix (α6) and the adjacent α4 helix in the NTHB (Fig. 3a), similarly to L273W (Fig. 3a, black bars). However, the projection analysis (Fig. 3b,c) reveals a marked difference between the L273W and G238A mutants in the apo-state. Unlike apo-L273W, for apo-G238A, most of the residues with $|\cos(\theta)| \sim 1$ (Fig. 3c) are characterized by strikingly positive fractional shift (*X*) values (Fig. 4b). For instance, for the hinge helix (α6) as well as the adjacent α4 helix, the majority of fractional shifts are in the ~0.4-0.5 range, clearly indicating that the G238A mutation causes a > 40 % shift of the conformational equilibria in the apo-EPAC1$_h$ CBD towards the active state, *i.e.* in the apo-G238A mutant, the auto-inhibition is partially, but significantly, removed. Therefore the highly conserved G238 of the β2-β3 loop is a key contributor to CBD auto-inhibition, exerting an inhibitory control on the allosteric sites of the CBD, including the adjacent PBC and the distal C-terminal hinge helix region (Fig. 4b). In contrast, the G238A mutation has relatively negligible effects on the holo/active state of the CBD (Fig S2b,c), suggesting that, at the level of the CBD, the G238A mutation results in more pronounced perturbations on the auto-inhibitory equilibria than on activation (Fig 1a). A possible explanation for the effect of G238 on the auto-inhibitory equilibria is through the CH-π interaction between the Cα of G238 and the guanidinium moiety of a highly conserved arginine in the PBC, *i.e.* R279 in EPAC1, which has been reported to co-evolve with G238 (12). In order to probe the effect of the G238A mutation on the CH-π interaction between G238 and the R279 guanidinium (Fig. 1d), we performed MD simulations on the EPAC CBD in three states: the Wt apo, the G238A mutant apo as well as the Wt cAMP-bound forms.

3.4.3 Molecular dynamics simulations indicate that the apo/active mimicking mutation G238A leads to tighter G238/R279 CH-π interactions and quenched dynamics in the PBC/β2-β3 loop regions

The MD simulations were first validated by computing the residue-specific α-helix probabilities (Fig. 5). We have previously shown by NMR spectroscopy that the most significant changes in secondary structure upon cAMP-binding are an increase in the helicity of α5 in the PBC and a concurrent helical unwinding in the C-terminal region of α6 in the hinge region (17,18). These experimental observations are well reproduced by the simulated MD trajectories (Fig. 5a). Furthermore, since the projection analysis of the NMR chemical shifts measured for apo-G238A indicates that this mutation causes a stabilization of the active conformation, we expect that, if the MD simulations are reliable, they should exhibit a stabilization of α5 and a C-terminal destabilization of α6 when Gly 238 is mutated to Ala in the apo-form of EPAC. Figure 5a shows that indeed the MD trajectory computed for the apo-G238A mutant displays the experimentally observed trend for the variation in the helical profile of α5 and α6, indicating that,

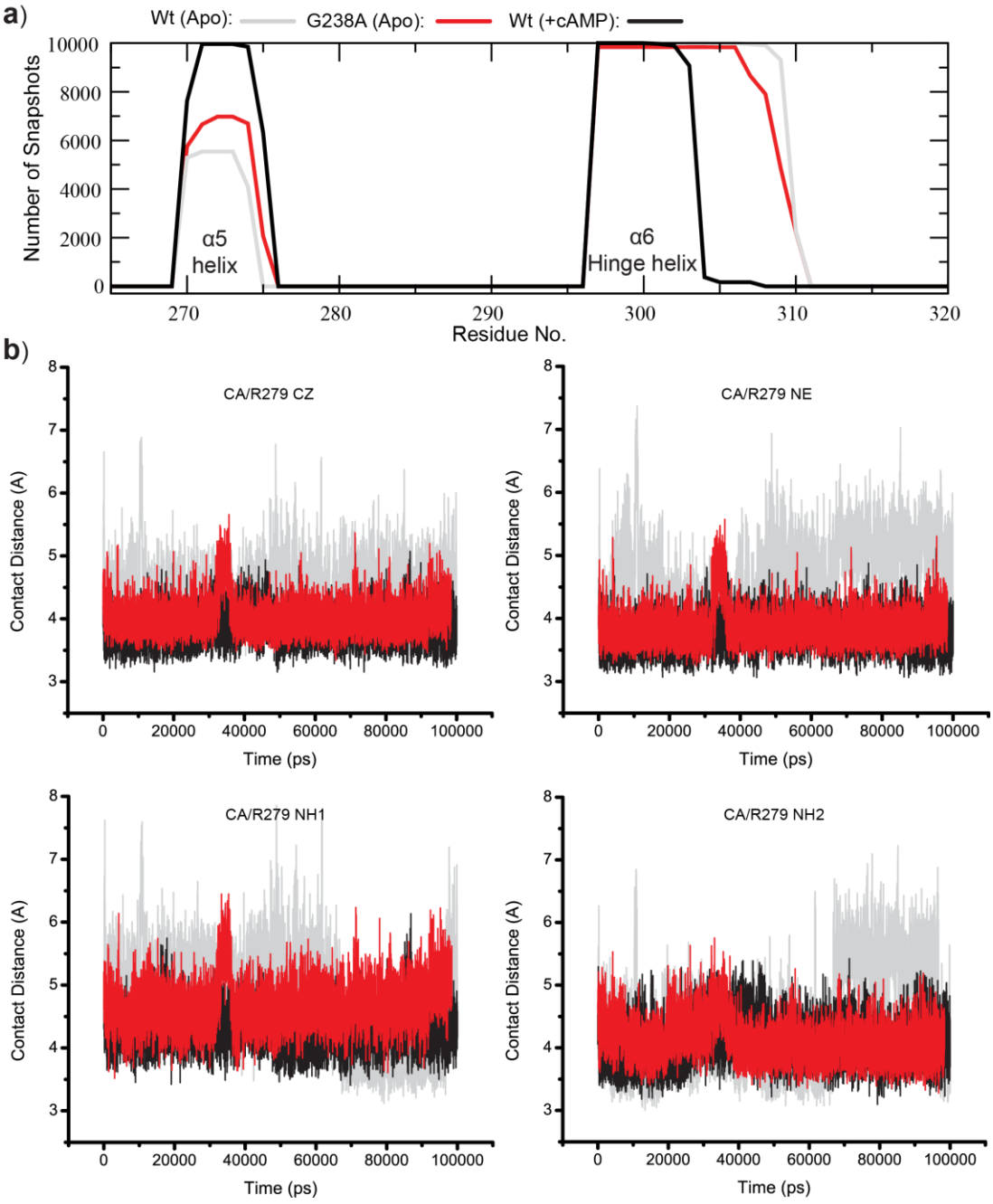

Figure 5: The effect of the G238A mutation as modeled by MD simulations. **(a)** Secondary structure analyses by DSSP over the course of the MD trajectories for the apo-Wt, the holo-wt, and the apo-G238A constructs. The analysis is for the region spanning the α5 and α6 helices, which are known to be cAMP-dependent (17-20). **(b)** Time profile for the contact distances between the Cα of residue 238 and the heavy atoms of the R279 guanidinium group. The color scheme is apo-Wt, holo-Wt, and apo-G238A.

despite the relatively brief 100 ns-time scale of the simulations, the MD trajectories capture at least qualitatively the trends observed experimentally through the projection analysis of the chemical shifts of G238A. The qualitative agreement between the MD simulations and the NMR data suggests that the MD

trajectories are a useful tool to probe how the G238A mutation affects the CH-π interactions between the conserved Gly and Arg residues of the β2-β3 loop and PBC, respectively.

Figure 5b reports the MD-simulated time profiles for the distances that define the CH-π G238/R279 interaction. The G238/R279 distances for the Wt EPAC construct (Fig. 5b, grey and black traces) indicate that cAMP binding results in a tightening of the CH-π interaction between these two residues (17,18). Figure 5b reports also the G238/R279 distances simulated for the apo-G238A mutant. Interestingly, the G238A mutation results in a shortening of the 238/279 distances in the apo-form that is remarkably similar to that seen upon cAMP binding in the Wt construct (Fig. 5b). Thus, both the G238A mutation and cAMP-binding result in a tighter CH-π 238/279 interaction. We therefore expect that, similarly to cAMP binding, the G238A mutation should lead to a quenching of dynamics for the PBC and β2-β3 loop regions. This prediction is confirmed by the Procustean rotation analysis of the MD trajectories (Figure S3 in the Supporting Material), which indicates that the G238A substitution causes a net loss in the relative PBC/β2-β3 loop dynamics of the apo-form. This MD result is in full agreement with the order parameters of G238A (Fig. 6a,b; Fig. S4 in Supporting Material), indicating that this apo-state mutation causes a quenching in the ps-ns dynamics of the PBC and β2-β3 regions.

Overall, the MD results suggest that G238 controls the auto-inhibitory equilibrium of EPAC through its interaction with the guanidinium of R279, which in turn modulates the dynamics and conformational entropy of the PBC and β2-β3 loop regions. Based on this interpretation, G238A stabilizes the active conformation of the apo-state by reducing the entropic loss that occurs in the apo-inactive to apo-active transition (Fig. 1a). This conclusion highlights the importance of the conformational entropy in the PBC and β2-β3 loop for the auto-inhibitory equilibrium of EPAC. However, it should be considered that the G238A mutation not only shifts the auto-inhibitory equilibrium of EPAC, but also causes severe nearest-neighbour perturbations both at the site of the mutation and in the adjacent PBC, as indicated by the large magnitude of the compounded chemical shift changes (Fig. 3a) and by the reduced cos(θ) values observed

for these regions (Fig. 3c). Consequently, we conclude that, while the G238A mutation is an excellent tool to evaluate the auto-inhibitory role of the β2-β3 loop, the auto-inhibitory contribution of entropic losses in the PBC and β2-β3 loop should be further confirmed through other apo/active state-stabilizing mutations that do not directly perturb the β2-β3 loop and are therefore less invasive in this region. The E308A and Q270A mutations are extremely useful in this respect (Fig. 1d).

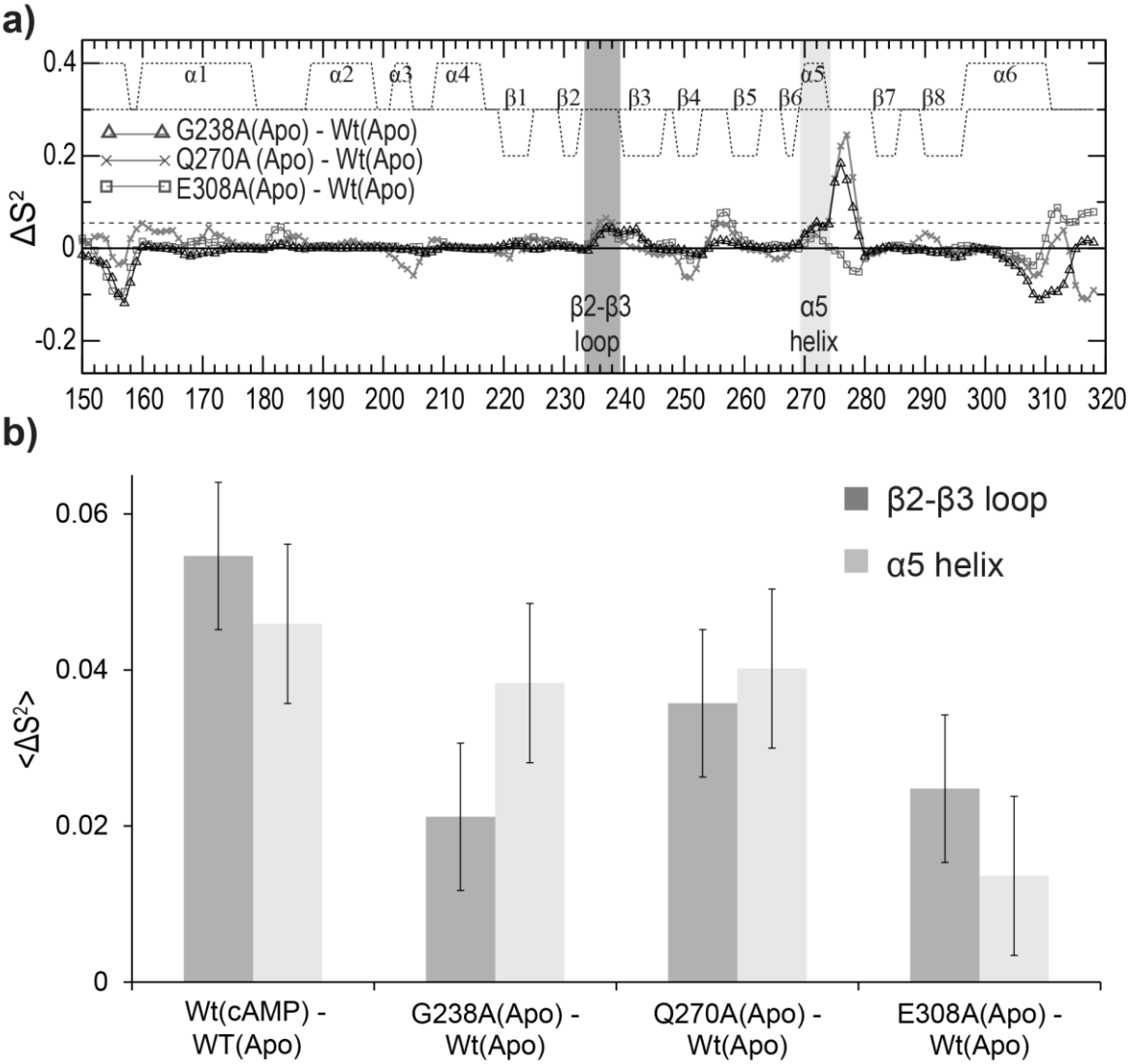

Figure 6: a) Order parameter differences relative to apo-Wt. The dashed horizontal line represents the standard deviation in the S^2 difference between cAMP-bound Wt (18,21) and apo-Wt. **b)** Average differences in S^2 for the regions subject to consistent dynamic quenching relative to apo-Wt when the apo-active state is stabilized in relation to the apo-inactive state.

3.4.4 The E308A and Q270A EPAC1 mutants provide a non-invasive 'glimpse' of the apo/active metastate

The projection analysis of the chemical shifts of E308A and Q270A (Fig. 7b,c) shows that both mutants shift the apo conformational equilibria towards the active state, because the majority of residues with $|\cos(\theta)| \sim 1$ display positive fractional activation values in the ~0.2 – ~0.3 range (Fig. 7b,c and Fig. 4c,d). In this respect, the E308A and Q270A mutants mimic the apo/active state, similarly to G238A, albeit to a smaller extent. However, unlike G238A, the E308A and Q270A mutants do not cause severe nearest-neighbour distortions in the PBC and β2-β3 loop, as indicated by $|\cos(\theta)|$ values close to unity for most of the residues in these regions (Fig. 4c,d). E308A and Q270A are therefore minimally invasive apo/active-stabilizing mutants in the $EPAC1_{149\text{-}318}$ construct and as such they are ideally suited to test the hypothesis that the PBC and β2-β3 loop dynamics are quenched upon transition from the apo/inactive to apo/active state, causing an entropic penalty that prevents premature activation in the absence of cAMP, *i.e.* that promotes auto-inhibition.

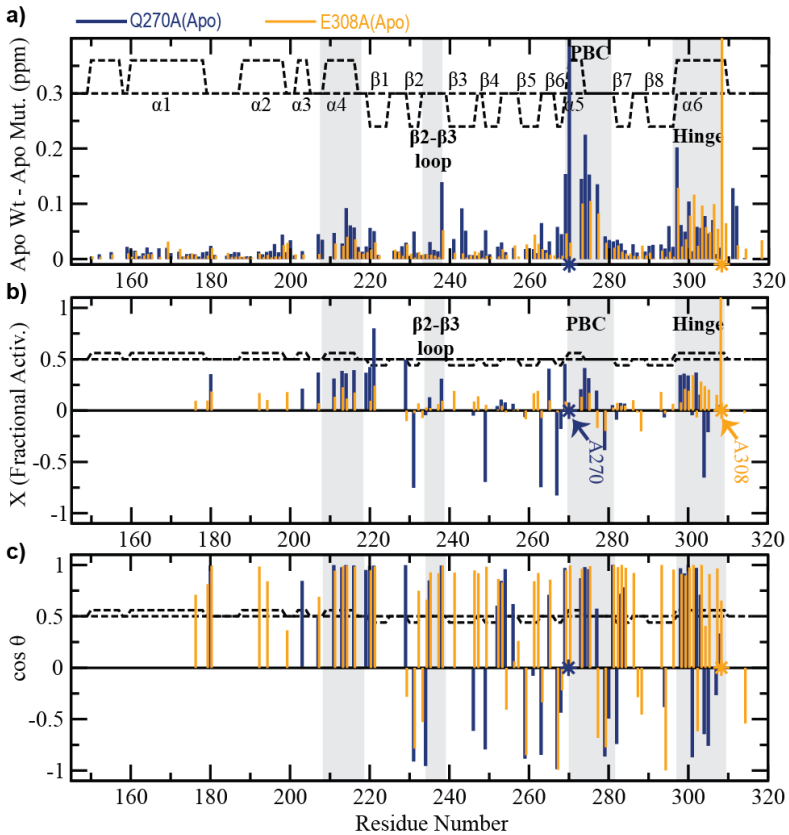

Figure 7: The effects of the E308A and Q270A mutations on apo-Wt (previous page). **a)** The compounded chemical shift profile of the apo-mutant relative to apo-Wt. **b)** Fractional shift toward activation achieved by the mutation in the absence of cAMP, as in Fig. 3b. **c)** Projection angle as in Fig. 3c.

In order to quantify the losses in conformational entropy occurring in the apo/inactive to apo/active transition, we evaluated the order parameters, S^2 (26) for apo-E308A and apo-Q270A (Fig. S4) and compared them to the corresponding apo-Wt values (Fig. 6a,b). A remarkable feature that emerges from Fig. 6 is that for both the β2-β3 loop and the PBC α5 helix, the order parameters measured for the Wt apo are consistently lower than those observed for the mutant apo-samples with increased relative populations of the active conformation (*i.e.* G238A, E308A and Q270A). As shown in Fig. 6b, the average β2-β3 and α5 S^2 enhancement relative to apo-Wt is consistently positive for both apo-mutants, similarly to G238A and to the Wt cAMP-bound (holo) state. We therefore conclude that the β2-β3 and PBC-α5 regions promote auto-inhibition by contributing a conformational entropy loss in the inactive-to-active conformational transition.

3.5 Discussion

3.5.1 The auto-inhibition determinants of EPAC extend beyond the L273/F300 pair of the hydrophobic hinge and include the β2-β3 loop and the α5

The mechanism of auto-inhibition in EPAC has been previously rationalized in terms of the hydrophobic hinge hypothesis (4,5), whereby the reorientation of the hinge helix towards the β-subdomain required for activation is prevented by a steric clash between the L273 in α5 and F300 in α6 (Fig. 1c). Our NMR data supports this mechanism, as the projection analysis of the chemical shifts of L273W shows that this bulky mutant shifts the apo conformational equilibrium towards the inactive state. However, the NMR investigation of the other mutations included in this study clearly reveals that the L273/F300 steric hindrance is not the only determinant of EPAC auto-inhibition. Several other structural elements of the CBD of EPAC contribute to the control of the inactive *vs.* active conformational equilibrium of the CBD in the absence of cAMP. Here we have identified two such elements, *i.e.* the α5 in the PBC and the β2-β3 loop. The β2-β3 loop has been previously shown to play a key role in the cAMP-dependent activation of EPAC (17,18,20,21), but its contribution to auto-inhibition, *i.e.* to the selective stabilization of the apo/inactive *vs.* apo/active state, was not anticipated before. In the following sections we discuss the mechanism utilized by these structural moieties to control the position of the auto-inhibitory equilibrium of EPAC.

3.5.2 Entropic control of auto-inhibition by α5 and the β2-β3 loop

The comparative analysis of the G238A, Q270A and E308A mutants indicates that whenever the apo/active state is stabilized relative to the apo/inactive state (Fig. 1a), the dynamics of the β2-β3 and α5 regions is consistently reduced. This result fully supports the hypothesis that the β2-β3 loop and the α5 in the PBC help control the position of the apo/inactive *vs.* apo/active equilibrium by imposing an unfavourable entropic loss as their ps-ns dynamics are quenched in the inactive-to-active conformational transition. A similar quenching of dynamics was previously reported for the β2-β3 loop and the α5 helix upon cAMP binding to Wt EPAC1 as measured by ^{15}N relaxation measurements (17). However, in the absence of cAMP, this entropic penalty does not appear to be fully compensated by the enthalpic gain resulting from the tighter G238/R279 Cα-π and L273/F300 Van der Waals contacts in the apo/active *vs.* apo/inactive structures. This conclusion highlights how flexible regions in the apo-form can be a critical molecular determinant of auto-inhibition.

3.5.3 The investigated determinants of auto-inhibition in EPAC CBD are non-degenerate and lead to enhanced specificity in the ligand-dependent activation.

The main goal of this work was to identify key determinants of auto-inhibition in the cAMP-binding domain (CBD) of EPAC, which serves also as a model system for other eukaryotic CBDs. The isolated CBD model used here (*i.e.* $EPAC1_{149\text{-}318}$) is ideally suited for this purpose. In addition, we anticipate that, if the NMR data on the CBD are functionally relevant, a shift towards activation in the auto-inhibitory equilibrium of the apo-CBD should result in decreased AC_{50} and increased k_{max} values (27) as measured from bioassays (Fig. S5). This prediction is confirmed not only in the extreme case of L273W as discussed above (5), but also in the case of Q270A and E308A, for which published AC_{50} and k_{max} values are available (Table S2) (27), suggesting that, at least on a qualitative basis, the NMR data on the isolated CBD bears some functional significance. However, a quantitative correlation between the NMR and the bioassay data (*i.e.* AC_{50} and k_{max} values) is not warranted at this stage, because the construct used in our NMR studies is a fragment of the multi-domain EPAC proteins typically required for functional assays. These longer constructs include additional auto-inhibitory elements that involve other domains. For instance, the so-called ionic latch (IL) is a set of salt bridges between the regulatory and catalytic regions of EPAC, which selectively stabilizes the closed inactive topology of EPAC. We therefore expect that the correlation between the NMR results on the isolated CBD and the bioassay data to be only qualitative (Table S2). Despite these limitations, a striking feature of auto-inhibition emerging from this NMR mutational study

is that the suppression of a single auto-inhibition determinant, among those investigated here, is insufficient to elicit a complete shift toward activation, *i.e.* no single mutation out of those tested is able to completely mimic the activation obtained at saturating cAMP concentrations. This observation suggests that it is necessary to simultaneously suppress multiple auto-inhibitory determinants in order to obtain a complete shift toward the active state. In this respect, the different structural elements of the CBD involved in auto-inhibition (*i.e.* α5, β2-β3 loop and α6) are non-degenerate and we propose that such lack of degeneracy in the auto-inhibition mechanism of EPAC confers added specificity towards cAMP. Overall, the requirement that multiple auto-inhibitory determinants must be simultaneously and collectively suppressed builds into the CBD a molecular switch with increased specificity towards the cAMP ligand.

3.5.4 Concluding Remarks

The mutations utilized in this study suggest that the previously proposed hydrophobic hinge hypothesis based on the L273/F300 steric clash is not sufficient alone to account for all the molecular determinants of EPAC auto-inhibition. First, several other highly conserved motifs within the CBD contribute to auto-inhibition, including the β2-β3 loop and the α5 helix. Second, these conserved structural elements fine-tune the auto-inhibitory equilibrium of EPAC for optimal cAMP sensitivity not only through steric and van der Waals interactions, as previously proposed, but also through entropic losses arising from regions that are more dynamic in the apo/inactive state relative to the apo/active state. Third, the multiple auto-inhibitory determinants identified here appear to be non-degenerate with respect to the complete activation of EPAC, suggesting that their collective suppression is required for full EPAC activity. Such non-degeneracy provides an effective means to enhance the specificity of the allosteric effectors that act on the EPAC "molecular switch". Finally, we anticipate that the concepts and methods illustrated here will be useful also for other systems. For instance, the novel chemical shift projection analysis (CHESPA) proposed here will likely be useful in evaluating how mutations modulate the position of conformational equilibria for domains that function as allosteric sensors in the regulation of signalling pathways (28-30).

3.6 References

References (31-32) appear in the Supporting Material

1. de Rooij, J., F.J. Zwartkruis, …, L. Bos 1998. Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature.* 396, 474-477.

2. Kawasaki, H., G.M. Springett, …, A.M. Graybiel 1998. A family of cAMP-binding proteins that directly activate Rap1. *Science.* 282, 2275-2279.

3. Rehmann, H., E. Arias-Palomo, …, J.L. Bos 2008. Structure of Epac2 in complex with a cyclic AMP analogue and RAP1B. *Nature.* 455, 124-127.

4. Rehmann, H., J. Das, …, J.L. Bos 2006. Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state. *Nature.* 439, 625-628.

5. Rehmann, H., B. Prakash, …, A. Wittinghofer 2003. Structure and regulation of the cAMP-binding domains of Epac2. *Nat. Struct. Biol.* 10, 26-32.

6. Grandoch, M., S.S. Roscioni, M. Schmidt 2010. The role of epac proteins, novel cAMP mediators, in the regulation of immune, lung and neuronal function. *Br. J. Pharmacol.* 159, 265-284.

7. Abu-Abed, M., R. Das, …, and G. Melacini 2007. Definition of an electrostatic relay switch critical for the cAMP-dependent activation of protein kinase A as revealed by the D170A mutant of RIalpha. *Proteins.* 69, 112-124.

8. Berman, H. M., L.F. Ten Eyck, …, S.S. Taylor 2005. The cAMP binding domain: An ancient signaling module. *Proc. Natl. Acad. Sci. U. S. A.* 102, 45-50.

9. Das, R. and G. Melacini 2007. A model for agonism and antagonism in an ancient and ubiquitous cAMP-binding domain. *J. Biol. Chem.* 282, 581-593.

10. Johnson, D. A., P. Akamine, …, S.S. Taylor 2001. Dynamics of cAMP-dependent protein kinase. *Chem. Rev.* 101, 2243-2270.

11. Kornev, A. P., S.S. Taylor, and L.F. Ten Eyck 2008. A generalized allosteric mechanism for cis-regulated cyclic nucleotide binding domains. *PLoS Comput. Biol.* 4, e1000056.

12. Kannan, N., J. Wu, …, S.S. Taylor 2007. Evolution of allostery in the cyclic nucleotide binding module. *Genome Biol.* 8, R264.

13. McNicholl, E. T., R. Das, …, G. Melacini 2010. Communication between tandem cAMP binding domains in the regulatory subunit of protein kinase A-ialpha as revealed by domain-silencing mutations. *J. Biol. Chem.* 285, 15523-15537.

14. Das, R., V. Esposito, …, S.S. Taylor and G. Melacini 2007. cAMP activation of PKA defines an ancient signaling mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 104, 93-98.

15. Masterson, L. R., L. Shi, …, and G. Veglia 2011. Dynamically committed, uncommitted, and quenched states encoded in protein kinase A revealed by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 108, 6969-6974.

16. Zagotta, W. N., N.B. Olivier, …, E. Gouaux 2003. Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature.* 425, 200-205.

17. Das, R., S. Chowdhury, …, and G. Melacini 2009. Dynamically driven ligand selectivity in cyclic nucleotide binding domains. *J. Biol. Chem.* 284, 23682-23696.

18. Das, R., M.T. Mazhab-Jafari, …, G. Melacini 2008. Entropy-driven cAMP-dependent allosteric control of inhibitory interactions in exchange proteins directly activated by cAMP. *J. Biol. Chem.* 283, 19691-19703.

19. Mazhab-Jafari, M. T., R. Das, …, G. Melacini 2007. Understanding cAMP-dependent allostery by NMR spectroscopy: Comparative analysis of the EPAC1 cAMP-binding domain in its apo and cAMP-bound states. *J. Am. Chem. Soc.* 129, 14482-14492.

20. Selvaratnam, R., S. Chowdhury, B. …, G. Melacini 2011. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc. Natl. Acad. Sci. U. S. A.* 108, 6133-6138.

21. Vanschouwen, B., R. Selvaratnam, …, G. Melacini 2011. Role of dynamics in the auto-inhibition and activation of the exchange protein directly activated by cyclic AMP (EPAC). *J. Biol. Chem.* [Epub ahead of print] PMID: 21873431.

22. Delaglio, F., S. Grzesiek, …, A. Bax 1995. Nmrpipe - a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR.* 6, 277-293.

23. Goddard, T. D., and Kneller, D. G. 2006. *SPARKY 3*, University of California, San Francisco, CA.

24. Berjanskii, M. V. and D.S. Wishart 2007. The RCI server: Rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Res.* 35, W531-7.

25. Berjanskii, M. and D. Wishart 2005. A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* 127, 14970-14971.

26. Jarymowycz, V. and M. Stone 2006. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* 106, 1624-1671.

27. Rehmann, H., F. Schwede, …, J. Bos 2003. Ligand-mediated activation of the cAMP-responsive guanine nucleotide exchange factor epac. *J. Biol. Chem.* 278, 38548-38556.

28. Ramos, D., T. Ducat, …, N.K. Goto 2006. Conformation of the cell division regulator MinE: Evidence for interactions between the topological specificity and anti-MinCD domains. *Biochemistry.* 45, 4593-4601.

29. Brubaker, W. D., J.A. Freites, …, R.W. Martin 2011. Separating instability from aggregation propensity in gammaS-crystallin variants. *Biophys. J.* 100, 498-506.
30. Popovych, N., S. Sun, …, C.G. Kalodimos 2006. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* 13, 831-838.

31. Phillips, J., R. Braun, …, K. Schulten 2005. Scalable molecular dynamics with NAMD RID A-1249-2010 RID D-1811-2010 RID D-5561-2009. *Journal of Computational Chemistry.* 26, 1781-1802.

32. Oblinsky, D.G., B.M.B. VanSchouwen, …, S.M. Rothstein 2009. Procrustean rotation in concert with principal component analysis of molecular dynamics trajectories: Quantifying global and local differences between conformational samples. *J. Chem. Phys.* 131, 225102(1-8).

3.7 Acknowledgments

We thank M. Akimoto and J. Milojevic for helpful discussions and the Canadian Institute of Health Research (CIHR) and the National Sciences and Engineering Research Council (NSERC) for financial support. We are also indebted to the Heart and Stroke Foundation of Canada (HSFC) for a Maureen Andrew New Investigator to G.M.. We also acknowledge the Shared Hierarchical Academic Research Computing Network (SHARCNET) for use of their high-performance computing resources as well as Dr. Stuart Rothstein research group for use of the Procrustean rotation software. We thank Dr. Holger Rehmann for helpful discussions.

3.8 Supplementary Materials: MD Simulation Protocol

All MD simulations were performed using the NAMD 2.7 software (S1) on the Shared Hierarchical Academic Research Computing Network (SHARCNET). The CHARMm27 force field was used for all simulations and the simulations were set up to mimic experimental conditions utilized in previous solution-state studies of EPAC: a pH of 7.6; explicit water (with periodic boundary conditions) with a 50 mM concentration of NaCl; a constant temperature of 34°C (307 K); and a constant external pressure of 1 atm. Protein structure coordinate and parameter files (with hydrogen atoms) for the EPAC2 structures were constructed using the "Psfgen" module of VMD 1.8.6. Parameters for cAMP were constructed from the parameters for the adenine ribonucleotide ("ADE" in the CHARMm force field) by applying the force field's intrinsic "CY35" patch function. Amino acid substitutions present in the simulated EPAC2 mutants were also introduced using the "Psfgen" module, by applying the module's intrinsic "mutate" routine to

the affected amino acid residues. In order to mimic a pH of 7.6, hydrogen atoms were added such that all His side chains were in their deionized τ-state and the N-/C-termini and all Asp, Glu, Arg and Lys side chains were in their ionized states. The structures were then immersed in a box of TIP3P water molecules using the Solvate module of VMD 1.8.6, ensuring a minimum distance of 12 Å between the protein and the edge of the solvent box. Salt ions (Na^+ and Cl^-) were added to the solvent box using the Autoionize module of VMD 1.8.6, such that the system was neutralized and the effective NaCl concentration in the solvent was 50 mM.

Initial energy minimizations were performed using the conjugate gradient algorithm of NAMD. Minimization was performed for 5000 steps with harmonic position restraints on the protein backbone (force constant of 300.0 kcal/mol·$Å^2$), followed by an additional 2000 steps without restraints. During minimization, a cut-off of 15 Å was utilized for all non-bonded energy calculations. Electrostatic interactions beyond the cut-off distance were computed using the Particle Mesh Ewald (PME) algorithm, with a tolerance of 10^{-6} and a maximum grid spacing of 1.0 Å. Molecular dynamics simulations were then performed under periodic boundary conditions, beginning from the energy-minimized initial structures. A time-step of 1.0 fs was implemented throughout the simulations. All water molecules were constrained to their equilibrium geometries using the SETTLE algorithm and all covalent bonds to hydrogen were constrained using the SHAKE algorithm. A cut-off of 12 Å with PME implementation was utilized for non-bonded energy calculations during the simulations. Short-range non-bonded and long-range electrostatic interactions were evaluated every 2.0 fs and 4.0 fs, respectively, using the RESPA multiple timestep integrator. All minimizations and simulations were executed on a 2.83 GHz octuple-core Xeon cluster, using 56 CPUs per run.

The structures were heated linearly from 0 K to 307 K over 200 ps at constant volume, using the velocity reassignment protocol of NAMD. The heated structures were then simulated at 307 K and constant volume (NVE ensemble) for another 1.0 ns, to allow a period of temperature equilibration prior to introducing pressure regulation. Next, the structures were simulated at a constant temperature and pressure (NPT ensemble) for 1.0 ns, to allow a period of temperature and pressure equilibration prior to the NPT production run. A constant temperature of 307 K was maintained using the Langevin dynamics algorithm, with a Langevin damping coefficient of 1.0 ps^{-1}. A constant pressure of 1 atm (1.01325 bar) was maintained using the Nosé-Hoover Langevin piston method, with a barostat oscillation period of 200.0 fs and a barostat damping time scale of 100.0 fs. Throughout the heating and equilibration runs, weak harmonic position restraints were imposed on the protein backbone (force constant of 5.0

kcal/mol·Å^2), to permit equilibration of the protein side chains and solvent without altering the protein backbone. Finally, production-run simulations were performed at a constant temperature and pressure (NPT ensemble) without restraints. These runs were executed for 110 ns, in order to obtain a 100 ns trajectory for analysis, while allowing for a final unrestrained equilibration period of 10 ns before the 100 ns trajectory. A constant temperature and pressure were maintained using the NPT protocol described above. During the production runs, structures were saved every 10000 timesteps (*i.e.* every 10.0 ps) for subsequent analysis.

Principal Component and Procrustean Rotation Analysis: Assessment of Differences in the Amplitudes of Inter-Residue Distance Fluctuations - To examine patterns in the amplitudes of inter-residue distance fluctuations within the EPAC2 construct, and how these patterns compare between simulated states of the construct, a Procrustean rotation analysis was performed on the MD trajectories. In this method, a principal component analysis is first performed on structures obtained from each simulation to be used for comparison, with inter-α-carbon distances as input variables (S2). The resulting principal component (PC) vectors are computed such that they optimally describe the patterns of fluctuation within the peptide structures, and how the fluctuation is distributed among the inter-α-carbon distances. The analysis is performed on inter-α-carbon distances because such variables have been previously shown to be the most reliable protein backbone structure descriptors for use in PC calculations. Following computation of the PC vectors, a comparison of two simulations is performed by orthogonal Procrustean rotation, followed by computation of factor loading deviations to examine differences in the amplitudes of inter-α-carbon distance fluctuations.

The analysis procedure starts with an (*M x N*) matrix of all *M* inter-α-carbon distances to be examined, for each of the *N* structures obtained from a simulation of interest (S2). The distance fluctuations in this matrix are then centered relative to the respective mean and normalized in accordance with the standard deviation values of each distance:

$$Ndist_{i,k} = \frac{(dist_{i,k} - \langle dist_i \rangle)}{S_{dist(i)}} \qquad \text{(S1)}$$

where $dist_{i,k}$ is the i^{th} inter-α-carbon distance from the k^{th} structure; $Ndist_{i,k}$ is the corresponding normalized inter-α-carbon distance; and $\langle dist_i \rangle$ and $S_{dist(i)}$ are the mean and standard deviation, respectively, of the i^{th} inter-α-carbon distance across all *N* structures. From the normalized distance matrix, an (*M x M*) correlation matrix **R** is computed as follows:

$$R_{i,j} = \frac{1}{N}\left\{\sum_{k=1}^{N}(Ndist_{i,k})(Ndist_{j,k})\right\} \quad \text{(S2)}$$

where $R_{i,j}$ is the computed correlation coefficient for the i^{th} and j^{th} inter-α-carbon distances; and $Ndist_{i,k}$ is the i^{th} normalized inter-α-carbon distance from the k^{th} structure (S2).

The computed matrix **R** is then loaded into the PDSYEVX routine of the ScaLAPACK software package for diagonalization. The (*M x M*) matrix **U**, which contains M normalized eigenvectors, diagonalizes matrix **R** as follows:

$$\mathbf{\Lambda = U^T R U} \quad \text{(S3)}$$

where $\mathbf{U^T}$ is the transpose of matrix **U** and **Λ** is a diagonal matrix that contains the corresponding eigenvalues. The eigenvalues and eigenvectors computed by PDSYEVX are sorted in order of decreasing eigenvalue and the first *P* eigenvalues and eigenvectors are selected for subsequent processing. The *P* selected eigenvectors are then scaled according to the respective eigenvalues to obtain an (*M x P*) matrix of principal component (PC) vectors called **Fac**:

$$Fac_{i,j} = U_{i,j}\sqrt{\lambda_j} \quad \text{(S4)}$$

where $Fac_{i,j}$ is the factor loading for the i^{th} inter-α-carbon distance and j^{th} PC; and $U_{i,j}$ and λ_j are the corresponding eigenvector and eigenvalue (S2).

Once the principal component (PC) vectors have been obtained for two simulations of interest (*e.g.* for Wt-apo and Mutant-apo), a Procrustean rotation is used to rotate the *P* selected PC vectors from one simulation such that they optimally superimpose onto those from the other simulation (S2). An optimal superposition is deemed to have been achieved when the following minimization criterion has been met:

$$g = \sum_{j=1}^{P}\left\{\sum_{i=1}^{M}(Rot_{i,j}{}^2 - Tar_{i,j}{}^2)^2\right\} = \min \quad \text{(S5)}$$

where $Rot_{i,j}{}^2$ is the Procrustean-rotated squared factor loading computed from one simulation for the i^{th} inter-α-carbon distance and j^{th} PC; and $Tar_{i,j}{}^2$ is the corresponding un-rotated squared factor loading computed from the other simulation, to which the first simulation is to be compared. Such rotation ensures that when differences between the simulations are subsequently computed, the two data sets will be positioned in comparable reference frames in multidimensional space.

The analysis was carried out using a Fortran-based software package, in which the PC calculations were performed with SHARCNET's AMD ScaLAPACK library (S2). All analyses were performed on residues 310-462 of the EPAC2 construct, in order to exclude C-terminal conformational fluctuations that could obscure results of interest in the CBD. In addition, a total of 100 PCs were utilized from each simulation, as these PCs captured 80-90 % of the total inter-α-carbon distance variance from the simulations. All PC calculations were run on 16 CPUs of a 2.2 GHz quad-core Opteron computer system at SHARCNET, and the subsequent Procrustean rotations were run as single-CPU routines. The resulting squared factor loadings were used to compute the factor loading deviation for each inter-α-carbon distance as follows:

$$Loading\ deviation(i) = \sum_{j=1}^{100} (Rot_{i,j}{}^{2} - Tar_{i,j}{}^{2}) \qquad \text{(S6)}$$

where $Rot_{i,j}^{2}$ and $Tar_{i,j}^{2}$ are defined as above and the sum of differences is computed across all 100 PCs selected for the analysis. Finally, the factor loading deviations were plotted as two-dimensional graphs, which display differences in the amplitudes of backbone structure fluctuations for all pairs of α-carbon atoms examined – *i.e.* on a per-inter-α-carbon-distance basis.

3.8.1 Supplementary Materials: References

S1) Phillips, J.C., R. Braun, …, and K. Schulten 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26, 1781-1802.

S2) Oblinsky, D.G., B.M.B. VanSchouwen, …, and S.M. Rothstein 2009. Procrustean rotation in concert with principal component analysis of molecular dynamics trajectories: Quantifying global and local differences between conformational samples. J. Chem. Phys. 131, 225102(1-8).

3.9 Supplementary Figures:

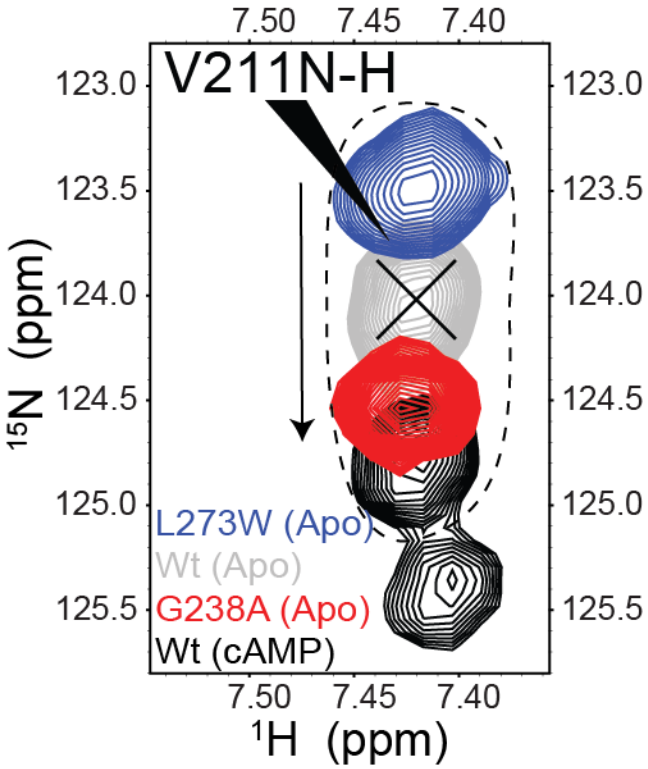

Figure S1: Representative region of the [^{15}N-^{1}H] HSQC spectra of apo-Wt (grey) and cAMP-bound (holo) Wt (black) superimposed with the [^{15}N-^{1}H] HSQC spectra of mutants: apo-L273W (blue) and apo-G238A (red).

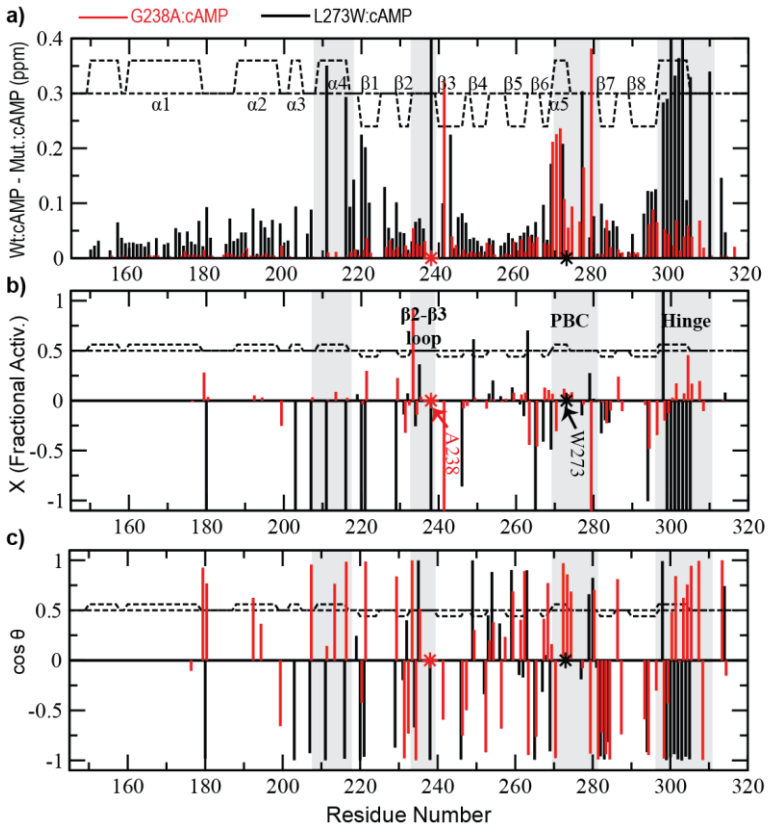

Figure S2: The effects of the L273W (black) and G238A (red) mutations in the presence of cAMP. **a)** The compounded chemical shift profile of the cAMP-bound mutant relative to cAMP-bound Wt. **b)** Fractional shift toward activation/inactivation as achieved by the mutation in the presence of cAMP. **c)** Projection angle as in Fig. 3c.

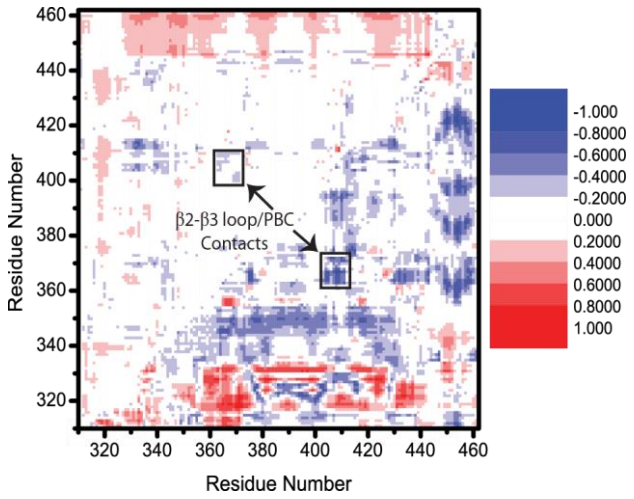

Figure S3: Results of the Procrustean rotation analysis of the MD simulations. The residue numbering refers to EPAC2. The area corresponding to the β2-β3 loop and the PBC is boxed in black lines. Dynamics in these regions are subject to quenching either due to the G238A mutation (upper triangular half) or cAMP (bottom triangular half). Refer to Supplementary Materials for more details.

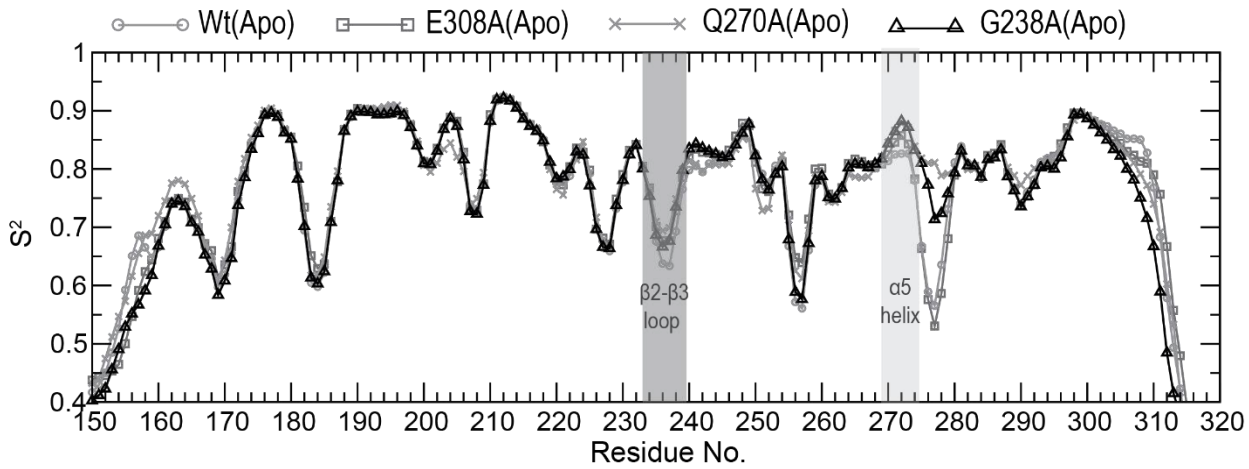

Figure S4: Order parameters (S^2) for the fast (ps-ns) dynamics of the apo-Wt, apo-G238A, apo-Q270A and apo-E308A.

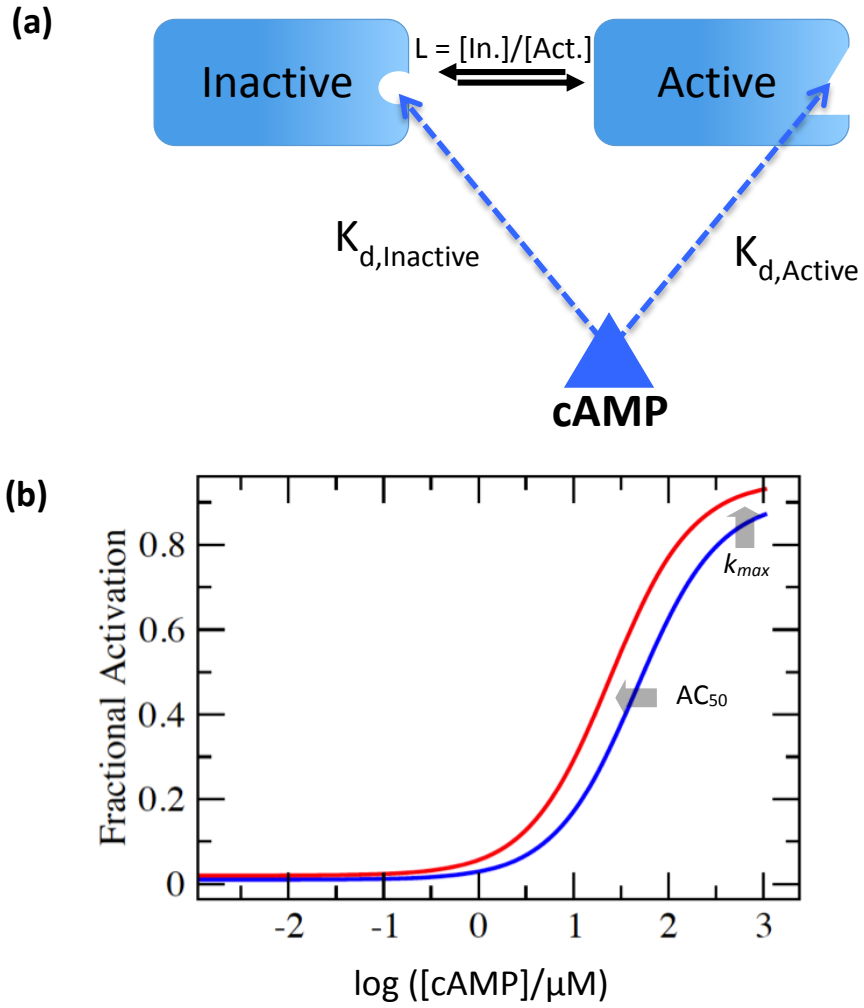

Figure S5: **(a)** Two-state thermodynamic model for the cAMP-dependent activation of EPAC. This model is defined by three key parameters: L = $[Inactive]_{Apo}/[Active]_{Apo}$, which is the equilibrium constant for the apo/inactive *vs.* apo/active equilibrium (*i.e.* auto-inhibitory equilibrium); $K_{d,Inactive}$ and $K_{d,Active}$, which are the dissociation constants for the binding of cAMP to the inactive and active conformations, respectively. In the case of EPAC, the exact values for L, $K_{d,Inactive}$ and $K_{d,Active}$ are not known. However, for illustrative purposes it is useful to note that setting $L = 10^2$, $K_{d,Inactive}$ = 500 µM and $K_{d,Active}$ = 0.5 µM results in an activation profile in the experimentally observed [cAMP] range (panel **(b);** blue curve). The activation profiles were computed based on the following equation: Fractional Activation = $1/(1+L_{app})$ with $L_{apparent} = L(1+([cAMP]/K_{d,Inactive}))/(1+([cAMP]/K_{d,Active}))$, where [cAMP] is the concentration of free cAMP. **(b)** Effect of a reduction of L on the activation profile. Simulated activation profile for wt EPAC (blue) and for a mutant (red) with reduced L value (*e.g.* L = 50 with unchanged $K_{d,Inactive}$ and $K_{d,Active}$). A reduced L models a shift of the apo auto-inhibitory equilibrium towards activation, similarly to what is observed through the NMR projection analysis for the Q270A and the E308A CBD mutants (Figure 7). Overall, a shift of the auto-inhibitory equilibrium towards activation results in decreased AC_{50} and increased k_{max} values (grey arrows), as also observed experimentally for the Q270A and the E308A CBD mutants (Table S2).

3.10 Supplementary Tables

Table S1: Fractional shifts for apo-L273W from projection analysis with varying cut-off

|B| cut-off: 0.025				|B| cut-off: 0.05				|B| cut-off: 0.1			
Residue	*X*, when |cos (θ)| >0.90	Residue	All *X*	Residue	*X*, when |cos (θ)| >0.90	Residue	All *X*	Residue	*X*, when |cos (θ)| >0.90	Residue	All *X*
295	5.6417	295	5.6417	219	1.4105	219	1.4105	277	0.4091	277	0.4091
219	1.4105	219	1.4105	265	0.9237	265	0.9237	220	-0.3176	252	0.0052
265	0.9237	265	0.9237	277	0.4091	277	0.4091	275	-0.5506	280	-0.0002
247	0.8904	247	0.8904	263	-0.2476	246	0.4088	214	-0.5840	282	-0.0222
236	0.4859	248	0.8727	220	-0.3176	252	0.0052	307	-0.6782	301	-0.1568
277	0.4091	236	0.4859	180	-0.3741	280	-0.0002	211	-0.7223	220	-0.3176
194	0.2540	277	0.4091	213	-0.4148	282	-0.0222	303	-0.7346	299	-0.4716
263	-0.2476	246	0.4088	275	-0.5506	207	-0.0853	216	-0.7401	275	-0.5506
220	-0.3176	289	0.2653	221	-0.5807	301	-0.1568	294	-0.9119	214	-0.5840
180	-0.3741	194	0.2540	214	-0.5840	263	-0.2476	268	-0.9738	307	-0.6782
213	-0.4148	250	0.1523	307	-0.6782	220	-0.3176	305	-0.9814	211	-0.7223
275	-0.5506	293	0.0220	211	-0.7223	180	-0.3741	308	-1.1205	303	-0.7346
221	-0.5807	252	0.0052	303	-0.7346	213	-0.4148			216	-0.7401
214	-0.5840	280	-0.0002	229	-0.7391	299	-0.4716			302	-0.8371
307	-0.6782	282	-0.0222	216	-0.7401	275	-0.5506			294	-0.9119
211	-0.7223	207	-0.0853	294	-0.9119	221	-0.5807			268	-0.9738
303	-0.7346	285	-0.1086	268	-0.9738	214	-0.5840			305	-0.9814
229	-0.7391	301	-0.1568	305	-0.9814	267	-0.5852			308	-1.1205
216	-0.7401	287	-0.1773	308	-1.1205	307	-0.6782				
179	-0.8506	263	-0.2476	231	-2.3170	211	-0.7223				
294	-0.9119	178	-0.3099			303	-0.7346				
268	-0.9738	220	-0.3176			229	-0.7391				
305	-0.9814	174	-0.3218			216	-0.7401				
308	-1.1205	180	-0.3741			249	-0.7904				
243	-1.9546	213	-0.4148			302	-0.8371				
231	-2.3173	299	-0.4716			294	-0.9119				
226	-2.6512	233	-0.4774			304	-0.9371				
296	-2.7131	275	-0.5506			268	-0.9738				
230	-2.8985	176	-0.5694			305	-0.9814				
286	-3.0738	221	-0.5807			308	-1.1205				
		214	-0.5840			231	-2.3173				
		267	-0.5852								
		307	-0.6782								
		211	-0.7223								
		303	-0.7346								
		229	-0.7391								
		216	-0.7401								
		192	-0.7444								
		249	-0.7904								
		302	-0.8371								
		179	-0.8506								
		294	-0.9119								
		304	-0.9371								
		268	-0.9738								
		305	-0.9814								
		308	-1.1205								
		237	-1.8461								
		243	-1.9546								
		262	-2.0320								
		231	-2.3173								
		226	-2.6512								
		296	-2.7131								
		230	-2.8985								
		286	-3.0738								

Table S2: Functional Bioassay Data for Q270A and E308A EPAC mutants from ref. (27).

EPAC Construct	**AC_{50} / µM**	***Relative k_{max}***
Wt	50	1
Q270A	40	1.7
E308A	15	3.0

Chapter 4

The Auto-inhibitory Role of the EPAC Hinge Helix as Mapped by NMR

4.1 Author's Preface

The work presented in this chapter has previously been published and is reproduced here under the Creative Commons Attribution License. Full citation is as follows:

Selvaratnam, R., M.T. Mazhab-Jafari, R. Das, and G. Melacini. 2012. The auto-inhibitory role of the EPAC hinge helix as mapped by NMR. PloS one. 7: e48707.

I conducted most of the experiments necessary for the manuscripts. Mohammad T. Mazhab-Jafari and Dr. Rahul Das generated and assigned the de305 mutant. Mohammad T. Mazhab-Jafari performed the saturation transfer difference experiments. I co-wrote the manuscript with Dr. Giussepe Melacini.

4.2 Abstract

The cyclic-AMP binding domain (CBD) is the central regulatory unit of exchange proteins activated by cAMP (EPAC). The CBD maintains EPAC in a state of auto-inhibition in the absence of the allosteric effector, cAMP. When cAMP binds to the CBD such auto-inhibition is released, leading to EPAC activation. It has been shown that a key feature of such cAMP-dependent activation process is the partial destabilization of a structurally conserved hinge helix at the C-terminus of the CBD. However, the role of this helix in auto-inhibition is currently not fully understood. Here we utilize a series of progressive deletion mutants that mimic the hinge helix destabilization caused by cAMP to show that such helix is also a pivotal auto-inhibitory element of apo-EPAC. The effect of the deletion mutations on the auto-inhibitory apo/inactive vs. apo/active equilibrium was evaluated using recently developed NMR chemical shift projection and covariance analysis methods. Our results show that, even in the absence of cAMP, the C-terminal region of the hinge helix is tightly coupled to other conserved allosteric structural elements of the CBD and perturbations that destabilize the hinge helix shift the auto-inhibitory equilibrium toward the apo/active conformations. These findings explain the apparently counterintuitive observation that cAMP binds more tightly to shorter than longer EPAC constructs. These results are relevant for CBDs in general and rationalize why substrates sensitize CBD-containing systems to cAMP. Furthermore, the NMR analyses presented here are expected to be generally useful to quantitatively evaluate how mutations affect conformational equilibria.

4.3 Introduction

The cAMP binding domain (CBD) is an ancient regulatory module found throughout multiple proteins with diverse functions (1-3). For example, in prokaryotes, a CBD is present in the transcription factor, catabolite activator protein (CAP) (4, 5). In eukaryotes, CBDs are found in Protein Kinase A and G (1,2,6-18) , in transport proteins, hyperpolarization activated and cyclic-nucleotide modulated (HCN) channels (19,20), as well as in the guanine nucleotide exchange factors, EPAC (Fig. 1) (3,10,21-28). Although, these aforementioned proteins are functionally diverse, the embedded CBD(s) play a similar allosteric role – regulation by means of auto-inhibition (29,30), *i.e.* the CBDs maintain a state of inactivity in the absence of the endogenous agonist, cyclic-AMP (cAMP) (22,23,25,27,31,32). Binding of cAMP acts by releasing the inhibition exerted by the auto-inhibiting determinants of the CBDs

The CBDs are typically characterized by an eight stranded jelly-roll β-sandwich, flanked by helices at the N- and C-termini as well as a small intervening helix situated between strands β6 and β7 (Fig. 1a,b) (1,2). Recent methods aimed at comparing patterns of amino acid conservations in sequence (1) and in space (2) have identified four conserved structural elements that are universally present in eukaryotic CBDs: the N-terminal helical bundle, the β2-β3 loop, the phosphate binding cassette (PBC) and the hinge helix (2). Previous investigations on the CBD of EPAC1, have established the former three structural elements as crucial determinants underlying auto-inhibition (10,21,27). However, the role of the hinge helix as an auto-inhibitory determinant of the EPAC CBD is currently not fully understood.

The last two turns of the EPAC hinge helix (called α6, Fig. 1a) partially unfold as α6 rotates towards the α5 helix of the PBC upon cAMP binding (Fig. 1b) (21,28,33). This hinge rotation has been rationalized as a consequence of the cAMP-induced repositioning of the PBC L273 residue, which contacts with F300 in the hinge helix. The repositioning of the conserved L273, and consequently F300, retracts the hinge helix toward the PBC helix upon activation (Fig. 1b) (23,25,27,31).

Recent studies mapping the EPAC allosteric network through chemical shift covariance analysis (CHESCA) have revealed that L273 and F300 are part of a larger cluster of allosteric residues, which includes also a hydrophobic spine at the interface between the α4 and α6 helices (Fig. 1d) (26). Such spine spans residues in the C-terminal end of the hinge helix that unwinds upon cAMP binding (*i.e.* 305-310, Fig. 1d).

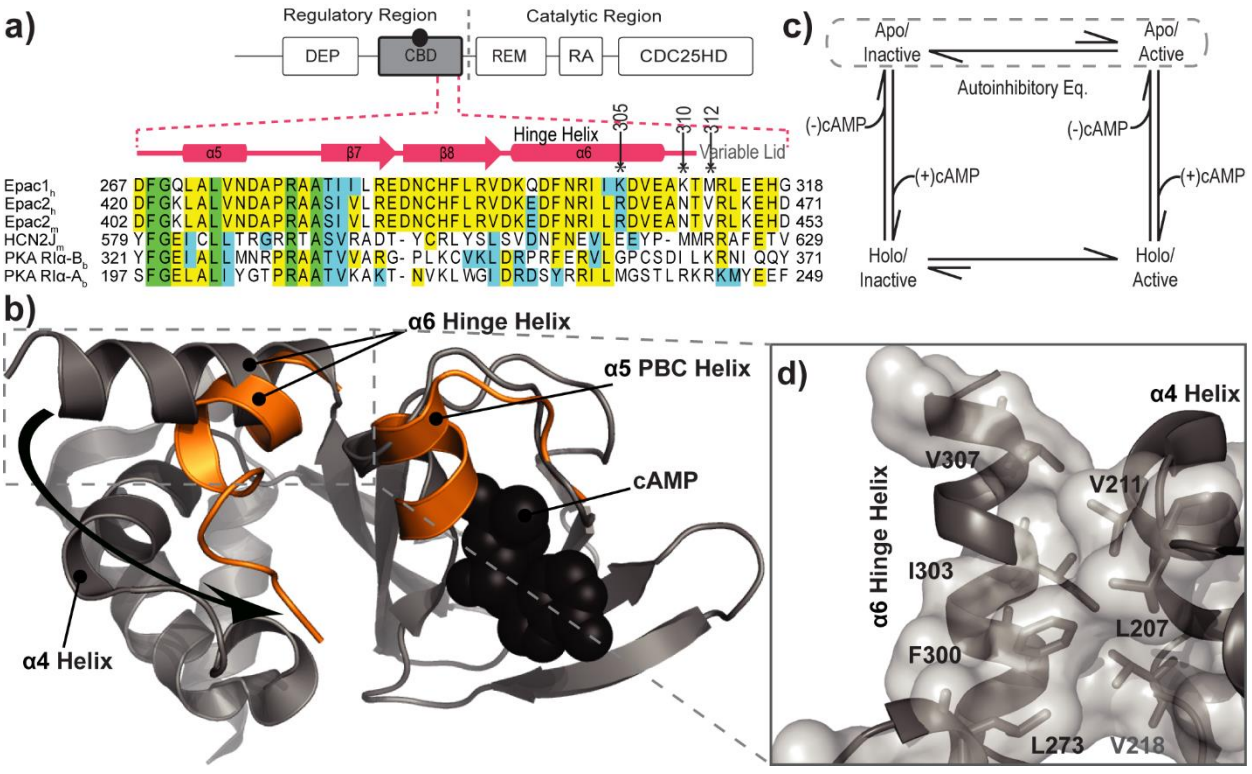

Figure 1: The CBD of EPAC and the domain organization. **a)** The regulatory region consists of the DEP (disheveled Egl-10 pleckstrin) domain and the cAMP binding domain (CBD), colored grey. The catalytic region includes the CDC25 homology domain (CDC25HD), the Ras exchange motif (REM), and Ras association (RA) domain. The dashed red lines illustrate an expanded view of the sequence alignment of CBDs for the regions spanning the PBC α5 helix to the Hinge Helix. The corresponding secondary structure is shown above the sequence. The asterisks mark the site of termination in the deletion mutants. **b)** The structure of the CBD of apo-EPAC is shown in grey, whereas the major changes caused by cAMP (black spheres) binding are shown in orange. The curved black arrow illustrates the transition of the hinge helix from the apo-form (grey; PDB ID: 2BYV) to the holo form (orange; PDB ID:3CF6). **c)** The thermodynamic cycle of cAMP dependent EPAC activation. Dashed lines encircle the equilibrium between the apo/inactive and apo/active states, i.e. the auto-inhibitory equilibrium. **d)** The hydrophobic "spine", a network of residues involving the hydrophobic contacts between the hinge helix and adjacent helices (α4 and α5).

Based on these observations, here we hypothesize that the C-terminal residues of the hinge helix (*i.e.* residues 305-310) are key determinants of EPAC auto-inhibition and that perturbations that destabilize the helix or induce unwinding shift the apo/inactive *vs.* apo/active pre-equilibrium toward the latter state, *i.e.* an active state without cAMP (Fig. 1c). To test this hypothesis, we designed three successive deletion mutations of the 149-318 EPAC1 construct (10,21), which spans the CBD and which from here on forth will be referred to as the Wt-EPAC. Specifically, these mutants are C-terminally truncated at positions 305, 310, and 312 (called de305, de310, and de312, respectively from here on forth) and act as

perturbations that destabilize the hinge helix of apo-EPAC, mimicking the cAMP-induced unwinding (Fig. 1b).

In order to explore how the mutations affect the inactive *vs.* active conformational equilibrium of apo-EPAC, we employed the previously proposed chemical shift projection analysis (CHESPA), which provides residue-specific fractional shift towards activation for each mutant (Fig. 2a) (27). In addition, the allosteric role of the hinge helix was further probed by the chemical shift covariance analysis (CHESCA), for which mutations were utilized as source of perturbations, unlike in previous CHESCA applications where cAMP and analogs were used to perturb the allosteric system (26).

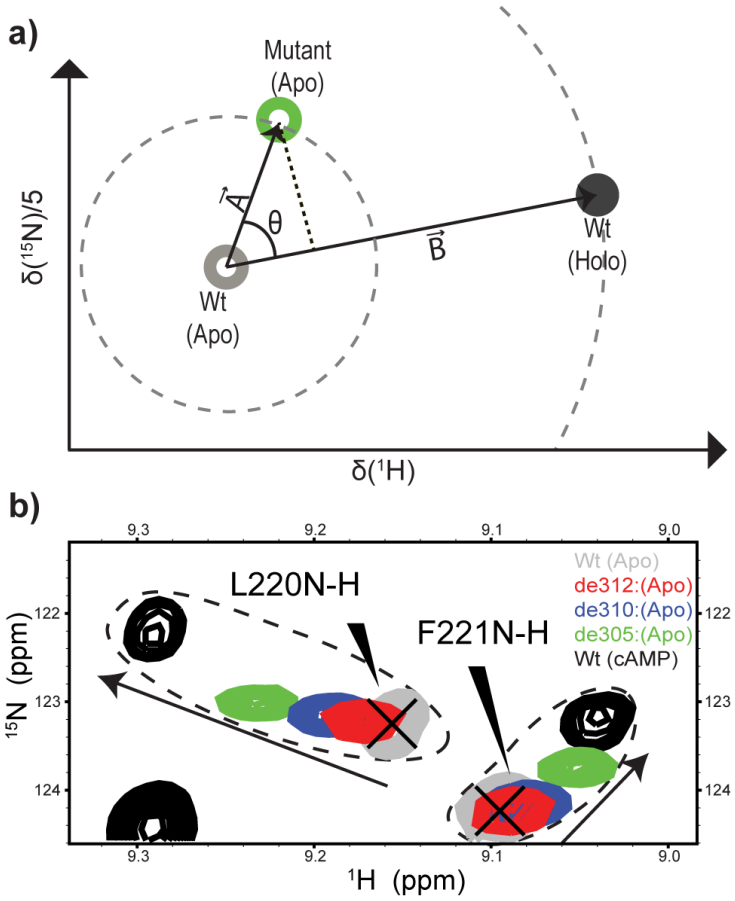

Figure 2: Chemical shift projection analysis (CHESPA) using mutations as perturbations. **a)** Schematic of CHESPA. Open circles indicate HSQC peaks of the apo-forms, whereas the filled circle represents the holo form (cAMP bound) HSQC peak. The green open circle represents the apo-mutant. The compounded chemical shift between the Wt(apo) and Wt(holo) was computed as the magnitude of the vector B, |B|. Similarly the compounded chemical shift between the Wt(apo) and Mutant(apo) was calculated as |A|. The magnitude of vectors A and B define the radii of the dashed circles centered on the Wt(apo) peak **(b)** Representative regions of the [^{15}N-^{1}H] HSQC spectra of Wt(apo) (grey) and cAMP-bound, Wt(holo) (black) overlaid with the [^{15}N-^{1}H] HSQC spectra of apo-Mutants: de312 (red), de310 (blue), de305 (green). Arrows indicate the direction of shift toward activation and dashed contour lines enclose peaks of the same residues.

Our results confirm the hypothesis that the C-terminal residues of the hinge helix (*i.e.* residues 305-310) are a pivotal determinant of EPAC auto-inhibition, showing that the hinge helix is extensively coupled to the other conserved allosteric elements of the CBD, even in the absence of cAMP. These results also lead to the counter-intuitive prediction that deletion of this C-terminal region causes an enhancement in cAMP-affinity, due to an increase in the apo/active relative population. This unexpected prediction was corroborated by the measurement of cAMP-binding isotherms through saturation transfer difference (STD) NMR experiments and the relevance of these results for the substrate-dependent sensitization to cAMP is also discussed (34,35).

4.4 Materials and Methods

Sample preparation: The deletion mutations de312, de310, and de305 were generated by inserting a stop codon at position 313, 311, and 306, respectively, by PCR in the Wt construct ($EPAC1_{149-318}$) and confirmed by DNA sequencing. Wt and all mutant constructs including E308A were purified and labelled according to published methods (26).

NMR Measurements: Spectra were acquired with a Bruker Avance 700-MHz spectrometer equipped with a 5 mm TCI cryoprobe at 306 K. Gradient and sensitivity enhanced [^{1}H-^{15}N] heteronuclear single quantum coherence (HSQC) were recorded for a total of 8 scans per t1 point. The number of digitized complex points were 256 and 1024 for the ^{15}N and ^{1}H dimensions, respectively, with an inter-scan delay of 1 sec. Carrier frequencies of the ^{15}N and ^{1}H channels were centered on water and the backbone amide region, respectively. All spectra were processed using NMRPipe (36) with linear prediction and a resolution-enhancing 60° shifted squared sine bell window function for HSQC spectra. Cross-peaks were assigned and integrated using Gaussian line-fitting in SPARKY (37). Assignments were obtained using triple-resonance experiments (21,38). All samples were referenced using the internal referencing compound ^{15}N-Ac-Glycine.

Chemical Shift Projection Analysis (CHESPA): The projection analysis descriptors, *i.e.* the cos Θ values, the fractional activations X and the compounded chemical shift differences between the apo-Wt and the apo-mutants (Fig. 2a) were computed as previously described (27). In brief, the compounded chemical shift difference between the apo-Wt and the apo-mutants was calculated as the magnitude of vector A in Figure 2a. Similarly, the compounded chemical shift difference between the Wt(apo) and the Wt(holo) was calculated as the magnitude of the activation vector B in Figure 2a. The chemical shift (ppm) of the

^{15}N was downscaled by a factor of 0.2, as indicated in Figure 2a. The cos Θ and fractional activation X were calculated as:

$$\cos\Theta = \frac{\vec{A} \bullet \vec{B}}{|\vec{A}||\vec{B}|} \qquad \text{(Equation 1)}$$

$$X = \frac{\vec{A} \bullet \vec{B}}{|\vec{B}|^2} \qquad \text{(Equation 2)}$$

Chemical Shift Covariance Analysis (CHESCA): The inter-residue correlation matrix (Figure 5) was generated according to published protocols (26). However, in contrast to previous applications of CHESCA, the perturbation set was composed of select mutations that destabilize the C-terminal end of the hinge helix. Such mutations were analyzed in the apo-state, where the extended hinge helix is stable. Thus, the perturbation set used here for CHESCA consisted of: Wt(apo), de305(apo), de310(apo), de312(apo) and E308A(apo).

S*ingular Value Decomposition (SVD) Analysis of Deletion Mutants*: The SVD analysis is based on previously published protocols (26), which were adapted and extended here for the application to deletion mutants. Specifically, a matrix **M** containing the combined chemical shifts for each assigned residue was first generated for five selected states: apo-Wt, cAMP-bound Wt, Sp-cAMPS-bound Wt, Rp-cAMPS-bound Wt and a 5th state that consisted of one of the deletion mutants in the apo-form (*i.e*. de312, de310 or de310) or the apo-L273W. The combined chemical shifts (δ_{NH}) were calculated as $\delta_{NH} = 0.2\delta_N + \delta_H$, where δ_N and δ_H are the individual chemical shift (ppm) values of the backbone ^{15}N and ^{1}H nuclei (26,39). Only residues for which the frequency spread across all five states was greater than 5 and 10 Hz for the individual ^{1}H and ^{15}N nuclei, respectively, were considered. A matrix **M'** was then subsequently generated from **M** in which the Rp-cAMPS-bound Wt state was used as reference for the remaining four states. Specifically, the columns of the **M'** matrix were: Wt(apo)–Wt(Rp-cAMPS), Wt(cAMP)–Wt(Rp-cAMPS), Wt(Sp-cAMPS)–Wt(Rp-cAMPS) and a 4th state with a deletion mutant or L273W in the apo-form measured relative to Wt(Rp-cAMPS) (*i.e.* de312(apo)–Wt(Rp-cAMPS), de310(apo)–Wt(Rp-cAMPS), de305(apo)–Wt(Rp-cAMPS) or L273W(apo)–Wt(Rp-cAMPS)). The matrix **M'** was then column mean centered and factorized through SVD as previously explained (26). The first two principal components (PCs) resulting from the SVD analyses performed here account for > 93% of the total variance (Table 1) and therefore the other PCs were deemed negligible and discarded.

cAMP Binding Measurements: The dissociation constant (K_D) for cAMP from Wt and de305 were measured through the saturation transfer difference (STD) amplification factor (STDaf) (24,40). All STD measurements were carried out with a solution of 25µM of Wt or 15 µM de305 in 20 mM phosphate buffer, pH 7.6, 50 mM NaCl, 99.9 % D_2O and at 25°C. The 1D-STD spectra were acquired at total cAMP concentrations of 25, 50, 75, 100, 150, 200 and 300 µM (24). Separate reference 1D (STR) experiments were also acquired. The STD amplification factor (STDaf) was calculated as the product of the STD/STR ratio (measured for the well resolved cAMP ribose H1' at 6.2 ppm) and of the ratio of the total cAMP and protein concentrations. The STDaf values were then normalized relative the STDaf plateau value reached at high cAMP concentrations ($[cAMP]_{Tot.} \geq 150$ µM). The normalized STDaf values were then analyzed with the binding isotherm equation: Normalized STDaf = $1 - (1 / (1 + ([cAMP]/K_D)))$, where [cAMP] is the concentration of free cAMP (24,40).

Table 1: Total Variance Breakdown in the SVD Analysis of Deletion Mutants and L273W

Mutant	**Principal Components (PCs)**	**Percentage of Total Variance**
$EPAC_{149-312}$ (de312)	PC1	52.4%
	PC2	44.8% (97.2%)*
$EPAC_{149-310}$ (de310)	PC1	59.8%
	PC2	36.3% (96.1%)*
$EPAC_{149-305}$ (de305)	PC1	57.3%
	PC2	35.8% (93.1%)*
$EPAC_{149-318}$ (L273W)	PC1	72.3%
	PC2	23.4% (95.7%)
*The percentages reported in parentheses are the cumulative contribution of PC1 and PC2 for each SVD analysis involving a mutant.		

4.5 Results and Discussion

4.5.1 CHESPA analysis of de305, de310 and de312

To investigate the effects of the C-terminal deletion mutations, we purified and assigned de305, de310 and de312 in the apo-states and compared them to the Wt(apo) and cAMP-bound states (Fig. 2a). We first analyzed the de312 truncation mutant (*i.e.* $EPAC1_{149\text{-}312}$), which leaves the hinge region (residues 296-310) to a large extent intact but removes the C-terminal tail of the Wt construct, $EPAC1_{149\text{-}318}$. The residue profile of the compounded chemical shift differences between Wt(apo) and de312(apo) (Figure 3a, red bars) exhibits local maxima in the regions most affected by cAMP-binding (Fig. 3a, grey regions) (9,21). In addition, the [^{15}N-^{1}H]-HSQC spectral comparison of the de312(apo) mutant relative to the Wt(apo) and cAMP-bound states for well dispersed and isolated peaks (Fig. 2b) reveals a slight but consistent shift for de312 towards the active state. However, in order to systematically assess at residue resolution the effect of the de312 mutation on the apo/inactive *vs.* apo/active auto-inhibitory equilibrium, we took advantage of the recently developed chemical shift projection analysis (CHESPA) (Fig. 2a; Fig. 3b,c, red bars). While the compounded chemical shifts quantify only the size of the perturbation, the fractional activation X obtained from the projection analysis (Fig. 3b) together with the cosine Θ values (Fig. 3c) reflect both the direction and extent of the mutational perturbation toward the apo/active state.

The fractional shifts obtained though the projection analysis reflect four main effects: (a) nearest neighbour effects experienced by residues in close spatial proximity to the site of the mutation; (b) mutation specific perturbations on interaction networks that involve the mutated site; (c) nearest neighbour effects experienced by residues in the binding site for the endogenous allosteric effector, *i.e.* cAMP in our case, as we use the Wt(apo) and Wt cAMP-bound (holo) states to define vector B (Fig. 2a); (d) changes in the inactive *vs.* active two-state equilibrium caused by the mutation (examined here for the apo samples). The projection analysis presented here is aimed at isolating the residues that reflect mainly effect (d). Effect (d) is residue independent, but effects (a-c) lead to residue-dependent variations in the fractional shifts. The effect (d) is best represented by the fractional activation (X) measured for the residue with cosine Θ absolute values ~1 (Figure 3c). In the case of de312(apo), the majority of such residues exhibit positive fractional activation values (Fig. 3b, red bars). These regions are also subject to the largest

chemical shift changes caused by cAMP (Fig. 3, grey zones)(10,21), suggesting de312(apo) mutation shifts the pre-equilibrium toward apo/active conformations.

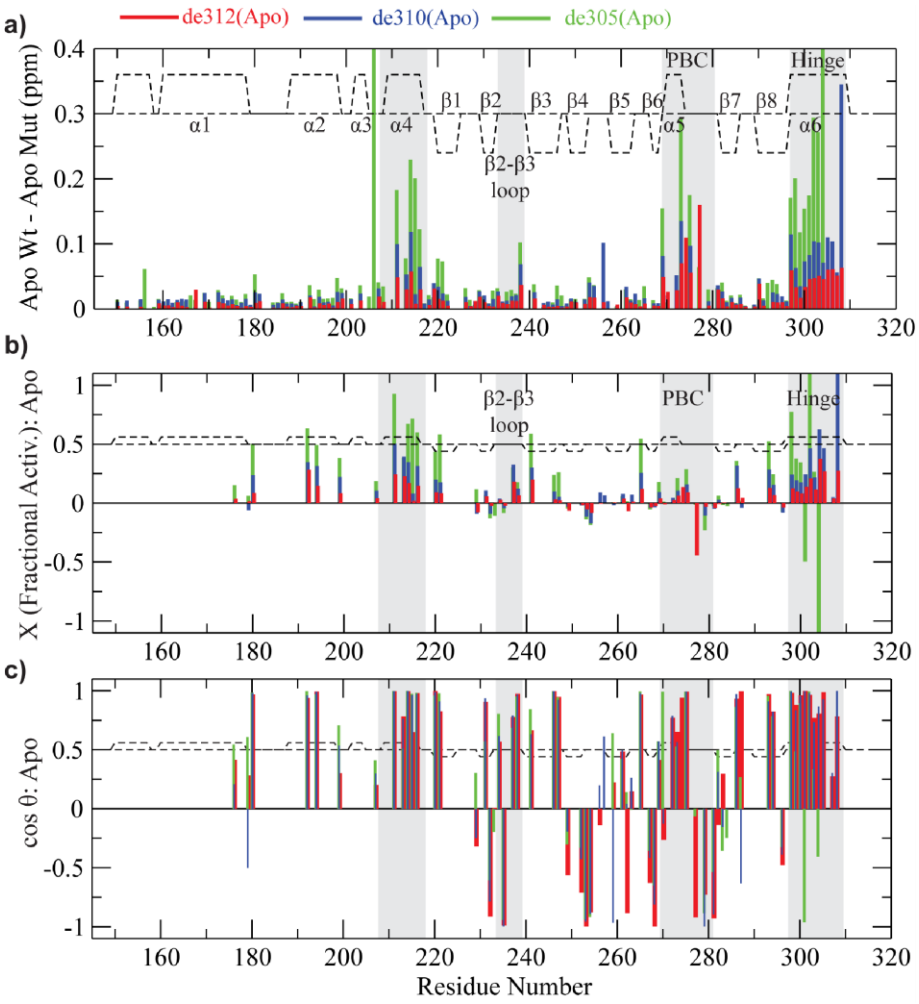

Figure 3: Chemical shift projection analysis to map the effects of the apo truncation mutants de312 (red), de310 (blue) and de305 (green) relative to Wt(apo), The dashed lines represent the secondary structure of the apo-EPAC (PDB ID: 2BYV). The grey highlights are regions subject to some of the most significant cAMP-dependent changes on the Wt(apo). **(a)** The compounded chemical shift profile of the apo-mutants relative to apo-Wt, that is the magnitude of vector A in Figure 2a. **(b)** Fractional shift toward activation achieved by the mutations in the absence of cAMP and with compounded chemical shifts greater than 0.05 ppm between the Wt(apo) and Wt(holo) state. **(c)** Cosine values for the projection angle, as in Fig. 2a, which is also an indicator of the direction of chemical shift movement along the activation path (vector B in Fig. 2a).

The CHESPA analysis of de310(apo) and de305(apo) mutants leads to results similar to those obtained for de312(apo), but with overall larger chemical shift differences and fractional activation values (Figure

3a-c), indicating that these mutations further destabilize the C-terminal hinge helix. The de310(apo) and de305(apo) constructs appear therefore to mimic the apo/active state more closely than de312(apo). However, due to structural distortions introduced by these mutations, the fractional activation values appear to be somewhat residue dependent (Fig. 3b) and based on the projection analysis alone it is not possible to obtain a reliable quantitative estimate of the overall relative shift towards the active state caused by the C-terminal truncation. In order to circumvent this limitation of the projection analysis, we utilized a recently introduced alternative approach based on singular value decomposition (SVD) of NMR chemical shifts (26), which provides an improved isolation of the ppm changes that exclusively reflect variations in the position of the inactive *vs.* active equilibrium.

4.5.2 The SVD analysis of the C-terminal truncation mutant indicates that the hinge helix residues 305-310 contribute to auto-inhibition

In the previously outlined SVD analysis of chemical shifts (26), HSQC spectra for the Wt EPAC1 construct were acquired and assigned in five different states: the Wt(apo) as well as four Wt-bound states, saturated with cAMP, Sp-cAMPS, 2'-OMe-cAMP and Rp-cAMPS. The Sp-cAMPS and 2'-OMe-cAMP analogs are both EPAC activators, while Rp-cAMPS functions as an EPAC antagonist, *i.e.* it binds the EPAC1 CBD without causing activation and is therefore used as a chemical shift reference state in the SVD protocol (26). Here, we use a similar SVD analysis, but we replace the 2'-OMe-cAMP-bound state with one of the mutants under investigation, *e.g.* the de312(apo) mutant. The 2'-OMe-cAMP was selected for this replacement because two other activators are already included in the analysis (*i.e.* cAMP and Sp-cAMPS) and therefore the SVD analysis is meaningful even in the absence of the 2'-OMe-cAMP state. Through this approach, the projection analysis is effectively expanded to include not only the Wt(apo) and cAMP-bound reference states (Fig. 2), but also the Sp-cAMPS- and Rp-cAMPS-bound forms, leading to an improved identification of the chemical shift changes that reflect uniquely variations in the activation equilibrium. For instance, when the 2'-OMe-cAMP-saturated state is replaced with the de312(apo) mutant, the first two principal components (PC) computed through SVD (*i.e.* PC1 and PC2) account for more than 93 % of the total variance (Table 1). PC1 reflects activation whereas PC2 is reflective of binding effects, as illustrated in Figure 4a by the Wt(Sp-cAMPS)–Wt(Rp-cAMPS) and Wt(cAMP)–Wt(Rp-cAMPS) loadings aligned with PC1 and the Wt(apo)–Wt(Rp-cAMPS) loading aligned with PC2. The PC1 component of the difference between the Wt(cAMP)–Wt(Rp-cAMPS) and the Wt(apo)–Wt(Rp-cAMPS) loadings provides therefore a measure of the maximal activation caused by cAMP and is utilized to normalize the PC1

component of the difference between the mutant(apo)–Wt(Rp-cAMPS) and the Wt(apo)–Wt(Rp-cAMPS) loadings (Fig. 4a, red arrows). This ratio of these PC1 components indicates that the de312(apo) deletion mutant causes a 7 % shift towards the apo/active conformers (Fig. 4b). The reliability of this approach was cross-validated by applying the SVD method to L273W (Figure S1 in Supporting Information), which leads to a 47 % shift of the Wt(apo) equilibrium towards the inactive conformers, consistent with previous analyses (27). A similar approach was also used to analyze the other two C-terminal deletion mutants, *i.e.* de310 and de305 (Fig. 4a, blue and green symbols, respectively), which cause further destabilization of the α6 helix. The percentage shifts towards activation caused by the successively truncating mutations de312, de310 and de305 are summarized in Fig. 4b. Figure 4b shows that the de310 and de305 truncations result in a further dramatic increase in the relative population of the apo/active conformers to 27 % and 35 %, respectively. Overall, the SVD analyses of Fig. 4a indicate that, while deletion of the C-terminal tail

in de312 causes only a subtle shift towards activation, perturbations in the C-terminal region of the hinge helix, implemented through the de310 and de305 truncations, lead to a more drastic stabilization of the active conformation in the absence of cAMP. These results are in agreement with the overall findings of CHESPA (Fig. 3b,c) and together consistently point to a significant and previously unanticipated auto-inhibitory role for residues 305-310 of the EPAC hinge helix.

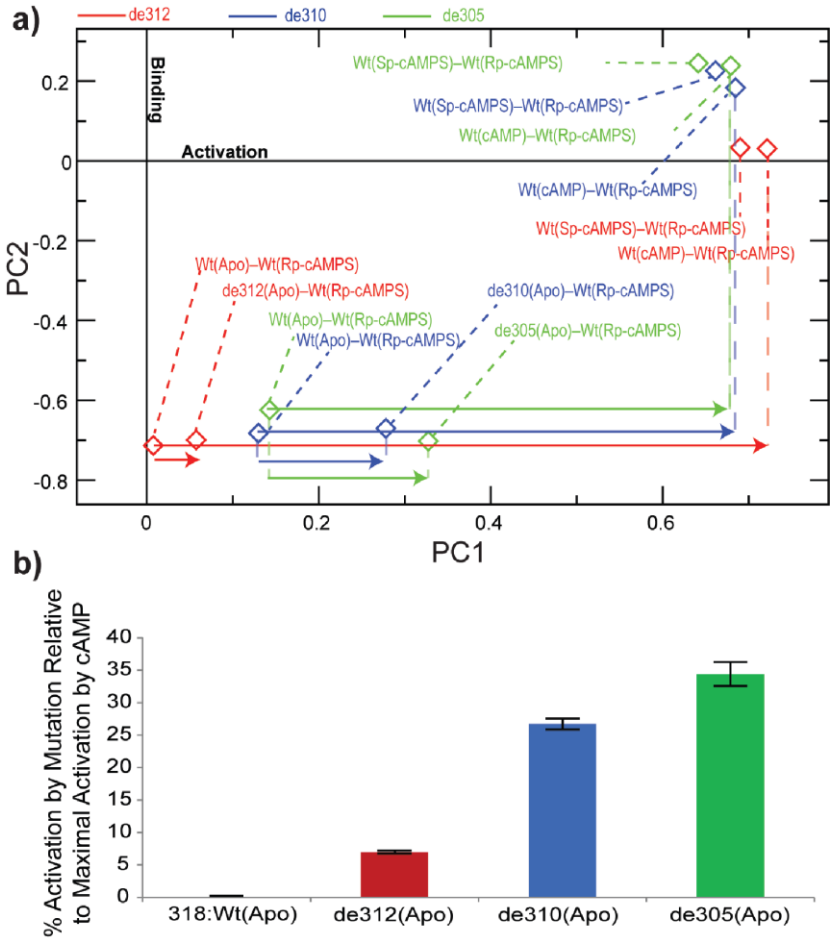

Figure 4: SVD analysis of the chemical shifts measured for the C-terminal truncation mutants de305, de310 and de312. **a)** This panel shows the PC1 vs. PC2 plot with three sets of loadings (diamonds) for each of the C-terminal hinge helix deletion mutants: de312 (red), de310 (blue) and de305 (green). There are four loadings per mutant with each loading corresponding to a state referenced to Rp-cAMPS, as labelled in the figure. The smaller arrows correspond to the separation along PC1 between the Wt(apo) and the mutant(apo) state. The large arrows correspond to the separation along PC1 between the Wt(apo) and the cAMP-bound Wt(holo). **b)** The percentage ratio of the two separations measured in panel (a) (i.e. relative magnitude of the two arrows), provides a quantitative measure of the overall fractional shift toward activation caused by the mutation.

4.5.3 The covariance analysis of chemical shifts reveals that the hinge-helix is coupled to the whole allosteric network of the EPAC CBD

In order to further explore the allosteric network controlled by residue 305-310 of EPAC1 in the absence of cAMP, we implemented the chemical shift covariance analysis (CHESCA) method (26) using as basis set the Wt(apo), de312(apo), de310(apo) and the de305(apo) truncation mutants as well as E308A(apo), which also targets the 305-310 regions. Using these five apo EPAC1 samples, several linear inter-residue chemical shift correlations are observed (Fig. 5a,b), resulting in a residue correlation matrix (Fig. 5c) that reveals the presence of an extensive long-range network of interactions controlled by the 305-310 α6 region. Specifically, the agglomerative cluster analysis (Figure S2 in Supporting Information) of the correlation matrix (blue grid, Fig. 5c) indicates that perturbations on residues 305-310 propagate to all the known allosteric sites of the EPAC1 CBD, from the PBC and the β2-β3 loop to most of the N-terminal helical bundle (red highlights, Fig 5c). Based on these observations, we conclude that the unwinding of residues 305-310 in α6 is coupled to the whole allosteric network controlled by cAMP (Fig. 5c).

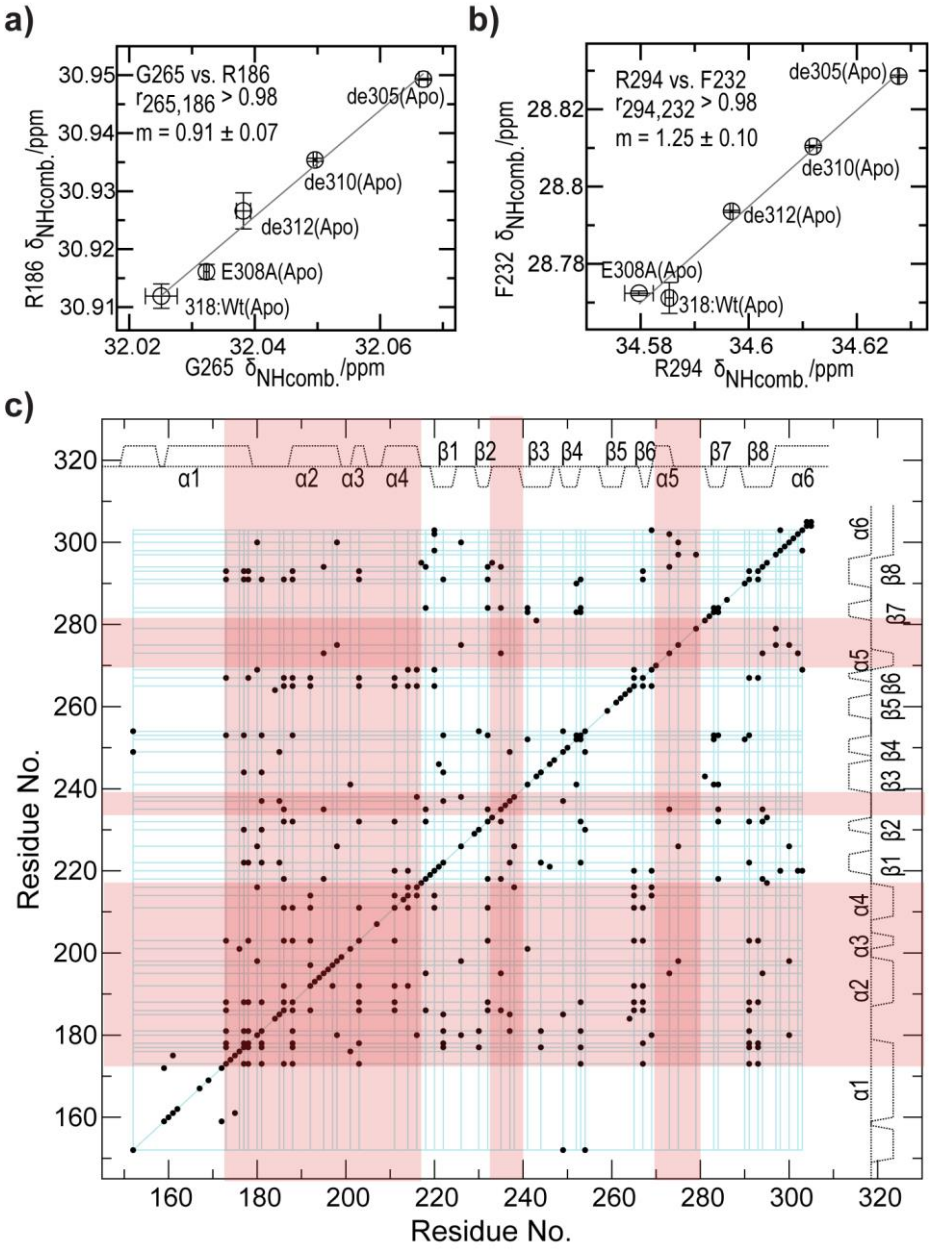

Figure 5: Chemical shift covariance analysis (CHESCA) of the hinge helix mutants. **a)** and **b)** show representative inter-residue chemical shift correlation among the five apo-states (318:Wt, 318:E308A, de312, de310, and de305) and 'm' defines the slope. **c)** The chemical shift correlation matrix. Residue pairs with absolute correlation coefficients ≥ 0.98 are marked with a dot. The blue grid represents the largest agglomerative cluster (Figure S2 in Supporting Information) (26), while regions highlighted in red correspond to key allosteric sites of the CBD other than the hinge helix.

4.5.4 Destabilization of the hinge helix enhances the affinity for cAMP

Considering that the apo/active state binds cAMP more tightly than the apo/inactive state, the coupling between the C-terminal region of α6 revealed by the combined CHESPA and CHESCA methods,

leads to the interesting prediction that de305, the closest mimetic of the apo/active form in our current investigation of the hinge helix (Fig. 4b), should exhibit higher affinity for cAMP than the Wt construct. This counter-intuitive prediction was experimentally confirmed by STD NMR measurements on both the de305 and the Wt construct (Fig. 6). As expected, Figure 6 clearly shows that the de305 mutant binds cAMP more tightly than Wt CBD with the full integral hinge helix. The ~8-fold decrease in K_D observed in going from the Wt to the de305 mutant rationalizes the observation that substrates sensitize CBDs to cAMP (34,35). Substrates promote the open (active) topology of EPAC and consequently the unwinding of the hinge helix in apo-EPAC, which in turn results in higher affinity for cAMP, explaining the lower K_D value measured for the dissociation of cAMP from EPAC in the presence of the Rap substrate (34). Interestingly, such sensitization of the CBD for cAMP in the presence of a substrate has also been observed for the CBD of PKA type I (35), suggesting that the auto-inhibitory role revealed here for the hinge-helix of EPAC may be relevant also for other cAMP-dependent systems.

4.6 Conclusions

The hinge helix is a universally conserved structural element of the CBDs. Here we have shown that in the CBD of EPAC, the C-terminal region of the hinge helix is an important determinant of auto-inhibition and is tightly coupled to the other conserved allosteric CBD elements even prior to cAMP binding. Alleviating the contributions of the hinge helix to auto-inhibition, as engineered for example through mutations, favours the active conformations even in the absence of the cAMP allosteric effector and consequently enhances the affinity of the CBD for cAMP. Overall, these results are relevant for CBDs in general and explain why substrates sensitize CBD-regulated systems to cAMP (34,35). Furthermore, the NMR analyses presented here are expected to be generally useful to quantitatively evaluate how mutations affect conformational equilibria.

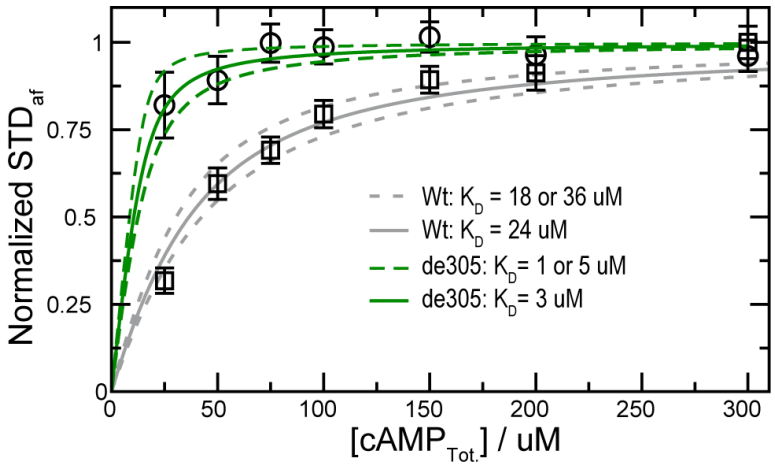

Figure 6: Binding isotherms for the titration of cAMP into an NMR sample with 15 µM de305 (green) and 25 µM Wt (grey) in 20 mM phosphate buffer, pH 7.6, 50 mM NaCl, 99.9% D2O, and at 25 °C. The binding of cAMP to de305 and Wt was monitored through the STD amplification factor (STDaf) normalized to the plateau value and plotted versus the total cAMP concentration. The binding of cAMP to the Wt construct, Epac1$_{149\text{-}318}$ was measured here to ensure an unbiased comparison to de305 since previous measurements (22, 34) were on Epac1$_{149\text{-}317}$ and used different experimental conditions and methods.

4.7 Acknowledgements

We thank M. Akimoto and B.VanSchouwen for helpful discussions and the Canadian Institute of Health Research (CIHR) and the National Sciences and Engineering Research Council (NSERC) for financial support. We are also indebted to the Heart and Stroke Foundation of Canada (HSFC) for a Maureen Andrew New Investigator to G.M.

4.5 References

1. Kannan N, Wu J, Anand GS, Yooseph S, Neuwald AF, et al. (2007) Evolution of allostery in the cyclic nucleotide binding module. *Genome Biol* 8(12): R264. 10.1186/gb-2007-8-12-r264.

2. Kornev AP, Taylor SS, Ten Eyck LF. (2008) A generalized allosteric mechanism for cis-regulated cyclic nucleotide binding domains. *PLoS Comput Biol* 4(4): e1000056.

3. Rehmann H, Wittinghofer A, Bos JL. (2007) Capturing cyclic nucleotides in action: Snapshots from crystallographic studies. *Nat Rev Mol Cell Biol* 8(1): 63-73.

4. McKay DB, Steitz TA. (1981) Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA. *Nature* 290(5809): 744-749.

5. Popovych N, Tzeng SR, Tonelli M, Ebright RH, Kalodimos CG. (2009) Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proc Natl Acad Sci USA*.

6. Esposito V, Sjoberg T, Das R, Brown S, Taylor SS, et al. (2006) NMR assignment of the cAMP-binding domain A of the PKA regulatory subunit RID A-3261-2009. *J Biomol NMR* 36: 64-64. 10.1007/s10858-006-9050-9.

7. Berman HM, Ten Eyck LF, Goodsell DS, Haste NM, Kornev A, et al. (2005) The cAMP binding domain: An ancient signaling module. *Proc Natl Acad Sci USA* 102(1): 45-50. 10.1073/pnas.0408579102.

8. Das R, Melacini G. (2007) A model for agonism and antagonism in an ancient and ubiquitous cAMP-binding domain. *J Biol Chem* 282(1): 581-593.

9. Das R, Esposito V, Abu-Abed M, Anand GS, Taylor SS, et al. (2007) cAMP activation of PKA defines an ancient signaling mechanism. *Proc Natl Acad Sci USA* 104(1): 93-98.

10. Das R, Chowdhury S, Mazhab-Jafari MT, Sildas S, Selvaratnam R, et al. (2009) Dynamically driven ligand selectivity in cyclic nucleotide binding domains. *J Biol Chem* 284(35): 23682-23696. 10.1074/jbc.M109.011700.

11. Masterson LR, Cheng C, Yu T, Tonelli M, Kornev A, et al. (2010) Dynamics connect substrate recognition to catalysis in protein kinase A. *Nat Chem Biol* 6(11): 821-828. 10.1038/nchembio.452.

12. Kim C, Cheng CY, Saldanha SA, Taylor SS. (2007) PKA-I holoenzyme structure reveals a mechanism for cAMP-dependent activation. *Cell* 130(6): 1032-1043. 10.1016/j.cell.2007.07.018.

13. Kim JJ, Casteel DE, Huang G, Kwon TH, Ren RK, et al. (2011) Co-crystal structures of PKG ibeta (92-227) with cGMP and cAMP reveal the molecular details of cyclic-nucleotide binding. *PLoS One* 6(4): e18413. 10.1371/journal.pone.0018413.

14. Masterson LR, Mascioni A, Traaseth NJ, Taylor SS, Veglia G. (2008) Allosteric cooperativity in protein kinase A. *Proc Natl Acad Sci USA* 105(2): 506-511. 10.1073/pnas.0709214104.

15. Das R, Abu-Abed M, Melacini G. (2006) Mapping allostery through equilibrium perturbation NMR spectroscopy. *J Am Chem Soc* 128(26): 8406-8407.

16. Anand GS, Krishnamurthy S, Bishnoi T, Kornev A, Taylor SS, et al. (2010) Cyclic AMP- and (rp)-cAMPS-induced conformational changes in a complex of the catalytic and regulatory (RI{alpha}) subunits of cyclic AMP-dependent protein kinase. *Mol Cell Proteomics* 9(10): 2225-2237. 10.1074/mcp.M900388-MCP200.

17. Badireddy S, Yunfeng G, Ritchie M, Akamine P, Wu J, et al. (2011) Cyclic AMP analog blocks kinase activation by stabilizing inactive conformation: Conformational selection highlights a new

concept in allosteric inhibitor design. *Mol Cell Proteomics* 10(3): M110.004390. 10.1074/mcp.M110.004390.

18. McNicholl ET, Das R, SilDas S, Taylor SS, Melacini G. (2010) Communication between tandem cAMP binding domains in the regulatory subunit of protein kinase A-ialpha as revealed by domain-silencing mutations. *J Biol* Chem 285(20): 15523-15537. 10.1074/jbc.M110.105783.

19. Chow SS, Van Petegem F, Accili EA. (2012) Energetics of cyclic AMP binding to HCN channel C terminus reveal negative cooperativity. *J Biol Chem* 287(1): 600-606. 10.1074/jbc.M111.269563.

20. Zagotta WN, Olivier NB, Black KD, Young EC, Olson R, et al. (2003) Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* 425(6954): 200-205.

21. Das R, Mazhab-Jafari MT, Chowdhury S, SilDas S, Selvaratnam R, et al. (2008) Entropy-driven cAMP-dependent allosteric control of inhibitory interactions in exchange proteins directly activated by cAMP. *J Biol Chem* 283(28): 19691-19703.

22. Rehmann H, Rueppel A, Bos JL, Wittinghofer A. (2003) Communication between the regulatory and the catalytic region of the cAMP-responsive guanine nucleotide exchange factor epac. *J Biol Chem* 278(26): 23508-23514.

23. Rehmann H, Prakash B, Wolf E, Rueppel A, de Rooij J, et al. (2003) Structure and regulation of the cAMP-binding domains of Epac2. *Nat Struct Biol* 10(1): 26-32.

24. Mazhab-Jafari MT, Das R, Fotheringham SA, SilDas S, Chowdhury S, et al. (2007) Understanding cAMP-dependent allostery by NMR spectroscopy: Comparative analysis of the EPAC1 cAMP-binding domain in its apo and cAMP-bound states. *J Am Chem Soc* 129(46): 14482-14492.

25. Rehmann H, Das J, Knipscheer P, Wittinghofer A, Bos JL. (2006) Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state. *Nature* 439(7076): 625-628.

26. Selvaratnam R, Chowdhury S, VanSchouwen B, Melacini G. (2011) Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc Natl Acad Sci USA* 108(15): 6133-6138. 10.1073/pnas.1017311108.

27. Selvaratnam R, Vanschouwen B, Fogolari F, Mazhab-Jafari MT, Das R, et al. (2012) The projection analysis of NMR chemical shifts reveals extended EPAC autoinhibition determinants. *Biophys J* 102(3): 630-639. 10.1016/j.bpj.2011.12.030.

28. Rehmann H, Arias-Palomo E, Hadders MA, Schwede F, Llorca O, et al. (2008) Structure of Epac2 in complex with a cyclic AMP analogue and RAP1B. *Nature* 455(7209): 124-127.

29. Zheng L, Stathopulos PB, Schindl R, Li GY, Romanin C, et al. (2011) Auto-inhibitory role of the EF-SAM domain of STIM proteins in store-operated calcium entry. *Proc Natl Acad Sci USA* 108(4): 1337-1342. 10.1073/pnas.1015125108.

30. Hayashi I, Plevin MJ, Ikura M. (2007) CLIP170 autoinhibition mimics intermolecular interactions with p150Glued or EB1. *Nat Struct Mol Biol* 14(10): 980-981. 10.1038/nsmb1299.

31. VanSchouwen B, Selvaratnam R, Fogolari F, Melacini G. (2011) Role of dynamics in the autoinhibition and activation of the exchange protein directly activated by cyclic AMP (EPAC). *J Biol Chem* 286(49): 42655-42669. 10.1074/jbc.M111.277723.

32. Selvaratnam R, Akimoto M, VanSchouwen B, Melacini G. (2012) cAMP-dependent allostery and dynamics in epac: An NMR view. *Biochem Soc Trans* 40(1): 219-223. 10.1042/BST20110628.

33. Harper SM, Wienk H, Wechselberger RW, Bos JL, Boelens R, et al. (2008) Structural dynamics in the activation of epac. *J Biol Chem* 283(10): 6501-6508. 10.1074/jbc.M707849200.

34. Kraemer A, Rehmann HR, Cool RH, Theiss C, de Rooij J, et al. (2001) Dynamic interaction of cAMP with the rap guanine-nucleotide exchange factor Epac1. *J Mol Biol* 306(5): 1167-1177. 10.1006/jmbi.2001.4444.

35. Viste K, Kopperud RK, Christensen AE, Doskeland SO. (2005) Substrate enhances the sensitivity of type I protein kinase a to cAMP. *J Biol Chem* 280(14): 13279-13284. 10.1074/jbc.M413065200.

36. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, et al. (1995) Nmrpipe - a multidimensional spectral processing system based on unix pipes. *J Biomol NMR* 6(3): 277-293.

37. T. D. Goddard and D. G. Kneller University of California, San Francisco. SPARKY. 3.

38. Sattler M, Schleucher J, Griesinger C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* 34(2): 93-158.

39. Aden J, Wolf-Watz M. (2007) NMR identification of transient complexes critical to adenylate kinase catalysis. *J Am Chem Soc* 129(45): 14003-14012.

40. Stockman B, Dalvit C. (2002) NMR screening techniques in drug discovery and drug design. *Prog Nucl Magn Reson Spectrosc* 41(3-4): 187-231. 10.1016/S0079-6565(02)00049-3.

4.6 Supplementary Figures

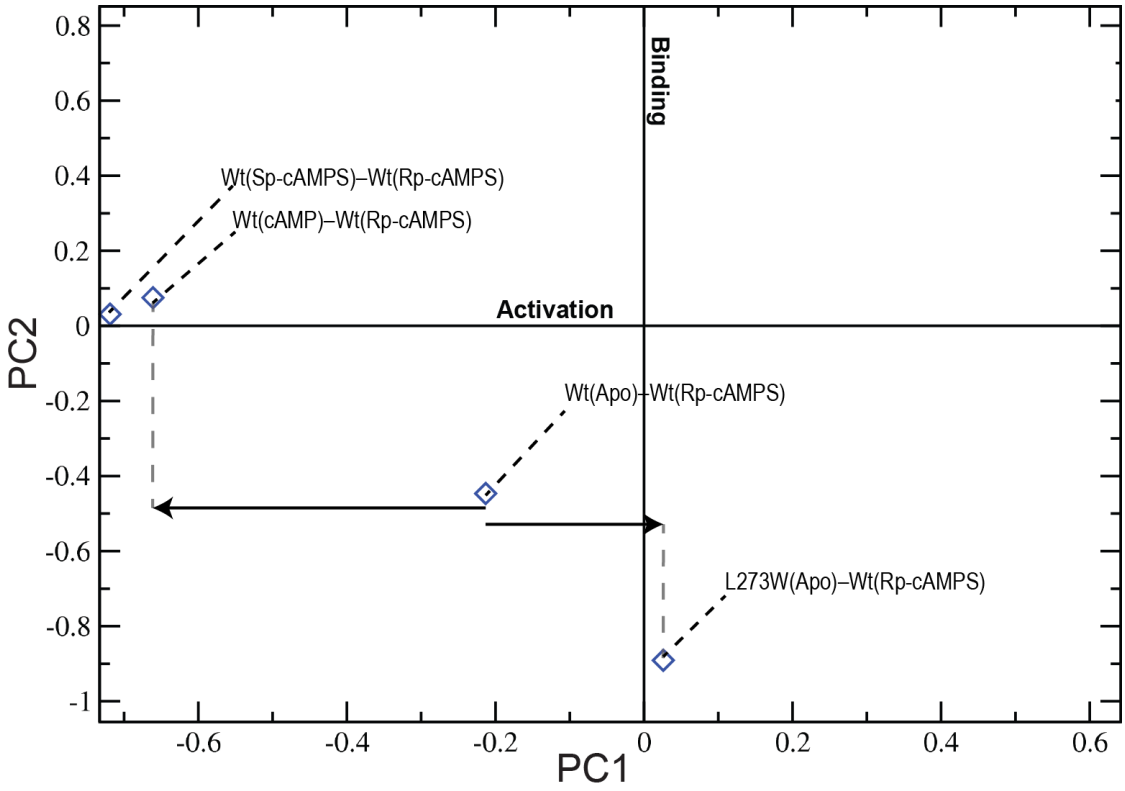

Figure S1: SVD analysis of the chemical shifts measured for the L273W(apo) mutant and other Wt states depicted in the plot relative to the Rp-cAMPS-bound Wt. PC1 and PC2 are as explained in the main text. Blue diamonds are the loadings.

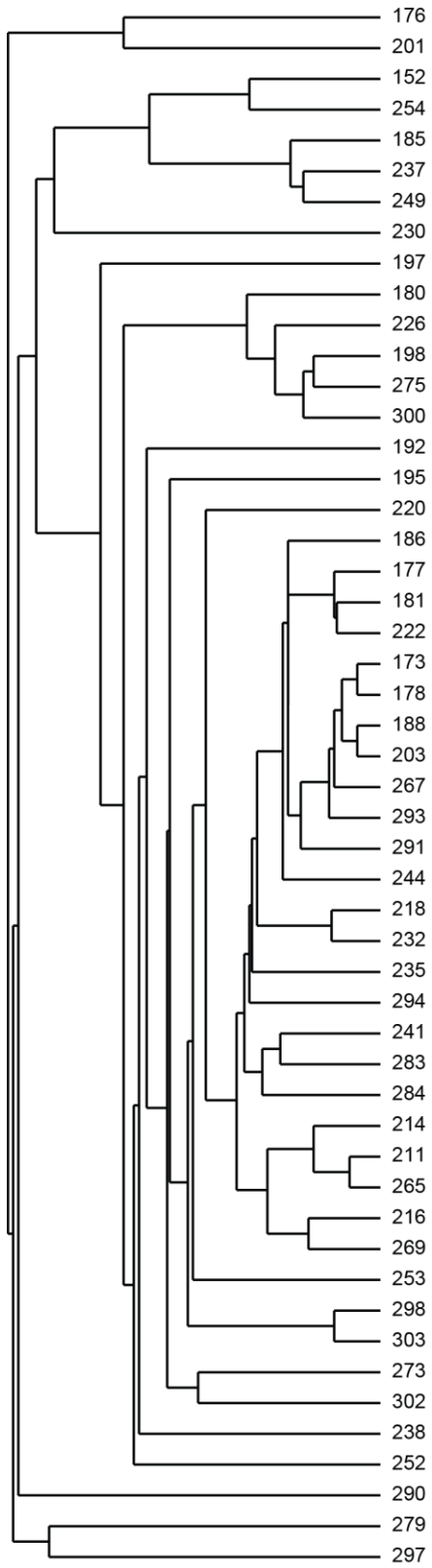

Figure S2: Dendrogram of the largest cluster of residues resulting from the agglomerative cluster analysis of the correlation matrix of Figure 5c. All nodes correspond to Pearson correlation coefficient ≥0.98.

Chapter 5

Conclusions and Perspectives

5.1 Determinants of EPAC Auto-inhibition

The CBD of EPAC is the central regulatory module that couples the cAMP signal to GEF activity on the small GTPases, Rap1 and Rap2. In the absence of cAMP, the CBD facilitates auto-inhibition of GEF activity. Chapters 3 and 4 of this thesis aimed at identifying the key determinants of auto-inhibition within the apo-EPAC CBD. For example, the previous chapter assessed the auto-inhibitory role of the helix at the C-terminus (*i.e.* α6 or hinge helix) of the CBD. The auto-inhibitory portion of this helix had been previously hypothesized to be confined to a phenylalanine residue located in the N-terminal portion of this helix (1–3). However, in Chapter 4 we demonstrate that the auto-inhibitory role of the α6 hinge helix extends beyond F300 and encompasses also residues 305-310, which in apo-EPAC spans the last two turns of the hinge helix and in holo EPAC become disordered and more flexible (4). Mutations that mimic the cAMP-induced destabilization of the 305-310 region cause a right shift in the inactive vs. active conformation equilibrium of the apo-EPAC CBD. This greater sampling of the apo/active state leads to a tighter affinity for cAMP, which is hypothesized to underlie substrate sensitization of EPAC (5).

In addition to the C-terminal end of the hinge helix, we also identified that dynamics of a conserved CBD structural motif, the β2-β3 loop, is an essential determinant of auto-inhibition in apo-EPAC, as predicted by MD simulations (6, 7). Specifically, Chapter 3 showed that mutations that facilitate the inactive to active transition in the apo-CBD consistently quench the dynamics of both the β2-β3 loop and the PBC α5 helix. Moreover, the β2-β3 / α5 dynamic quenching was observed irrespective of the specific location of the active state-promoting mutation (*e.g.* E308A in the hinge helix, Q270A in the PBC and G238A in the β2-β3 loop). These observations suggest that the rigidification of the β2-β3 loop and α5 sites is an intrinsic property of the inactive to active conformational transition and occurs even in the absence of cAMP, suggesting that the entropic losses resulting from the quenching of β2-β3 loop/α5 dynamics stabilize the inactive state of the apo-EPAC CBD. These results provide a validation of the predictions based on MD simulations and corroborate the pivotal role of the β2-β3 loop/α5 sites within the intra-molecular network identified by CHESCA in Chapter Two (7, 8).

Overall, in Chapters Three and Four, we have identified key determinants of auto-inhibition in the CBD of EPAC that were previously unanticipated and, although a various number of mutations were individually assessed through the novel method of CHESPA, it was found that no single mutant could fully suppress the auto-inhibition and confer activation to the same extent as the allosteric effector, cAMP. This observation suggests the presence of multiple auto-inhibitory determinants in structurally conserved features of the CBD that are non-degenerate and act collectively to provide an increase specificity for cAMP (5, 7). This result is in agreement with the recent functional assay for the apo-state of the hinge helix mutant F300G, which results in only ~60% of the GEF activity induced by cAMP in Wt EPAC (3).

5.2 Bridging Sites of EPAC Auto-Inhibition

The auto-inhibitory elements described in the previous section (5.1) and in Chapters Three and Four are subject to both major structural changes and dynamic modulations upon cAMP-binding. On the contrary, another critical auto-inhibitory element of EPAC, *i.e.* the IL in the NTHB, is dynamically modulated by cAMP, but lacks major cAMP-dependent structural changes (Figure 5 of Chapter 1). Specifically, it had previously been shown that cAMP enhances the IL dynamics, thereby imposing an entropic penalty on the formation of the inhibitory IL salt-bridges in the presence of cAMP (4, 9). Since the long-range functional control exerted by cAMP on the IL relies almost exclusively on changes in dynamics, Chapter 2 of this dissertation sought to examine how the cAMP signal originating at the PBC is able to 'diffuse' to distal regions not subject to major structural changes. Through the CHESCA method

outlined in Chapter two, we have identified a spatially contiguous set of residues almost completely confined to the α-subdomain as the carrier of signal transmission between the PBC site, where cAMP docks, and the distal IL site, where the inhibitory salt bridges between the regulatory and catalytic regions of EPAC reside (8). Moreover, all of the auto-inhibitory determinants discussed in the previous section are embedded in this intra-molecular network, further supporting the importance of this signalling pathway.

Previous studies also attempted to address similar questions related to identifying the critical network of residues connecting distant functional sites (10, 11). Two such popular methods include, the double mutant cycles (12) and the statistical coupling analysis (SCA) of multiple sequence alignments (13). The double mutant cycle method relies on the functional comparison between two single residue mutations and a double mutant with the goal of identifying non-additive effects (12, 14). While this method identifies allosteric couplings with a clear functional relevance, the mutant cycle method is typically restricted to only a fraction of the residues in extended allosteric networks. The SCA method makes use of the co-variation of amino acid residues that have co-evolved and therefore may confer allosteric/functional coupling. While the SCA approach has the advantage of not requiring the acquisition of ad hoc experimental data, co-evolution of residues could be due to either structural or functional adaptation. Therefore this can result in residues that are close in space as opposed to those residues that are remote from each other, but energetically coupled. Hence SCA sometimes also misses critical networks of functionally coupled residues as shown for by Chi et al (12). Nonetheless, we presume that both mutant cycles and SCA are likely to be complementary to CHESCA, which connects functionally remote sites such as the IL and the PBC based on an appropriate set of perturbations monitored by NMR.

Overall, the CHESPA and CHESCA methods have provided an unprecedented insight into the allosteric signalling mechanism underlying auto-inhibition and cAMP dependent activation in the CBD of EPAC. The characterization of intra-molecular networks provides new avenues towards designing effective allosteric inhibitors and activators of EPAC. Additionally, these methods, which make use of carefully measured chemical shifts are likely to be broadly applicable with ease once assignment of the perturbed states can be obtained. For example, Axe et al., recently employed both CHESPA and CHESCA relying only on the amide assignment of Alanine residue to probe long range interactions in the α-subunit of tryptophan synthase (15).

5.3. Future Directions

The CHESPA method provides an immediate and systematic readout of the extent of activation or inactivation caused by a given perturbation. This CHESPA feature could be exploited in the context of mutant cycles. Such cycles often make use of functional assays to assess the consequence of the mutation. However, in cases where mutations cannot be assayed as in adaptor proteins or fragments of modular proteins that cannot purified in whole, CHESPA with appropriate inactive- and active-state references could be a suitable substitute in double or triple mutant cycles. This extension of CHESPA to mutant cycles would allow one to probe additive *vs*. non-additive effects of the mutation at residue resolution. Additionally, although we used backbone amides as the probe, the method can easily be extended to methyl side chains (16).

This thesis focused on identifying the intra-molecular networks of the CBD of EPAC. However, future work could also aim to consider how the cAMP induced signalling is propagated to adjacent domains, *i.e.* inter-domain communication. For example, it has been shown that translocation of EPAC to the plasma membrane requires both an open active conformation induced by cAMP and the recruitment of the N-terminally adjacent DEP domain (Figure. 2) (17). At present, it remains elusive as to how cAMP affects the adjacent DEP domain.

Additionally, although the CBD is in the 'central controlling unit' of a number of other proteins such as PKA, PKG and HCN, it is currently unknown to what extent the intra-molecular networks that bridge the effector binding site and the auto-inhibitory loci are preserved in other CBDs. For such comparison, it is critical to generate constructs with similar domain boundaries to avoid contributions from linkers and other non-conserved segments. This comparative CBD analysis would provide an insight into not only the molecular divergence of allostery across structurally convergent CBDs, but would also be of use in the design of system specific agonists and antagonists. The latter is particularly important in the case of HCN which is a direct target of novel inhibitory drugs to cope with coronary heart disease such as angina pectoris (18). However, the mechanism of auto-inhibition in HCN by the CBD has largely been elusive, since the currently available structure of the apo-state of HCN is to a large extent similar to that of the cAMP bound state, although the HCN function is clearly cAMP-dependent. This is a common problem for the apo-state of CBDs(19–21), which is typically highly dynamic due to the exchange between active and inactive states. Upon crystallization often the former conformation is selected, resulting in technically sound, but functionally uninformative apo-CBD structures. However, since the chemical shift analysis methods presented here do not necessarily require structure determination, the application of CHESCA

and CHESPA to HCN would aid in mapping the intra-molecular networks that bridge the sites underlying HCN auto-inhibition, providing an unparalleled allosteric insight useful in designing novel inhibitors of HCN.

5.4 References:

1. Rehmann H et al. (2003) Structure and regulation of the cAMP-binding domains of Epac2. *Nature structural biology* 10:26–32.

2. Tsalkova T, Blumenthal DK, Mei FC, White M a, Cheng X (2009) Mechanism of Epac activation: structural and functional analyses of Epac2 hinge mutants with constitutive and reduced activities. *The Journal of biological chemistry* 284:23644–51.

3. White M a et al. (2012) Structural analyses of a constitutively active mutant of exchange protein directly activated by cAMP. *PloS one* 7:e49932.

4. Das R et al. (2008) Entropy-driven cAMP-dependent allosteric control of inhibitory interactions in exchange proteins directly activated by cAMP. *The Journal of biological chemistry* 283:19691–703.

5. Selvaratnam R, Mazhab-Jafari MT, Das R, Melacini G (2012) The auto-inhibitory role of the EPAC hinge helix as mapped by NMR. *PloS one* 7:e48707.

6. VanSchouwen B, Selvaratnam R, Fogolari F, Melacini G (2011) Role of dynamics in the autoinhibition and activation of the exchange protein directly activated by cyclic AMP (EPAC). *The Journal of biological chemistry* 286:42655–69.

7. Selvaratnam R et al. (2012) The projection analysis of NMR chemical shifts reveals extended EPAC autoinhibition determinants. *Biophysical journal* 102:630–9.

8. Selvaratnam R, Chowdhury S, VanSchouwen B, Melacini G (2011) Mapping allostery through the covariance analysis of NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America* 108:6133–8.

9. Das R et al. (2009) Dynamically driven ligand selectivity in cyclic nucleotide binding domains. *The Journal of biological chemistry* 284:23682–96.

10. Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. *Nature chemical biology* 4:474–82.

11. Rousseau F, Schymkowitz J (2005) A systems biology perspective on protein structural dynamics and signal transduction. *Current opinion in structural biology* 15:23–30.

12. Chi CN et al. (2008) Reassessing a sparse energetic network within a single protein domain. *Proceedings of the National Academy of Sciences of the United States of America* 105:4679–84.

13. Lockless SW (1999) Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science* 286:295–299.

14. Boyer J a, Clay CJ, Luce KS, Edgell MH, Lee AL (2010) Detection of native-state nonadditivity in double mutant cycles via hydrogen exchange. *Journal of the American Chemical Society* 132:8010–9.

15. Axe JM, Boehr DD (2013) Long-Range Interactions in the Alpha Subunit of Tryptophan Synthase Help to Coordinate Ligand Binding, Catalysis, and Substrate Channeling. *Journal of molecular biology*:1–19.

16. Ruschak AM, Kay LE (2012) Proteasome allostery as a population shift between interchanging conformers. *Proceedings of the National Academy of Sciences of the United States of America* 109:E3454–62.

17. Ponsioen B et al. (2009) Direct spatial control of Epac1 by cyclic AMP. *Molecular and cellular biology* 29:2521–31.

18. Postea O, Biel M (2011) Exploring HCN channels as novel drug targets. *Nature reviews Drug discovery* 10:903–14.

19. Kim JJ et al. (2011) Co-crystal structures of PKG Iβ (92-227) with cGMP and cAMP reveal the molecular details of cyclic-nucleotide binding. *PloS one* 6:e18413.

20. Anand GS et al. (2010) Cyclic AMP- and (Rp)-cAMPS-induced conformational changes in a complex of the catalytic and regulatory (RI{alpha}) subunits of cyclic AMP-dependent protein kinase. *Molecular & cellular proteomics : MCP* 9:2225–37.

21. Taraska JW, Puljung MC, Olivier NB, Flynn GE, Zagotta WN (2009) Mapping the structure and conformational movements of proteins with transition metal ion FRET. *Nature methods* 6:532–7.