ALLOSTERIC NETWORK AND QSAR MODEL OF EPAC1

IDENTIFICATION OF CORE ALLOSTERIC NETWORKS AND DEVELOPMENT OF
QSAR MODELS FOR EPAC1

By HEBATALLAH MOHAMED, B.S. Biomedical Sciences

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the
Requirement for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2021) Hamilton, ON (Chemical Biology)

TITLE: Identification of Core Allosteric Networks and Development of QSAR Models for EPAC1

AUTHOR: Hebatallah Mohamed, B.S.Biomedical Sciences (Zewail City of Science and Technology)

SUPERVISOR: Professor Giuseppe Melacini

NUMBER OF PAGES: xii, 101

Lay Abstract

Exchange proteins activated by cyclic AMP (EPAC) are cAMP sensors with several functions in cellular pathways. EPAC has been found to be associated with multiple diseases such as cardiovascular diseases. This study aims to first identify the main residues involved in the regulation of EPAC1 activity and then develop a predictive model that is able to find promising and selective inhibitors for the protein. The two approaches can then be useful in designing effective modulators of EPAC1 for the treatment of cardiovascular diseases such as cardiac hypertrophy.

Abstract

Allosteric regulation is essential to control biological function. In addition, allosteric sites offer a promising venue for selective drug targeting. However, accurate mapping of allosteric sites remains challenging since allostery relies on often subtle, yet functionally relevant, structural, and dynamical changes. In this thesis, a new toolset of NMR-based methodologies known as T-CHESCA and CLASS-CHESCA are proposed to identify key allosteric sites, using isoform 1 of the exchange protein activated by cAMP (EPAC1) as the model system. The T-CHESCA imposes changes on the fast-exchanging active/inactive states of the protein through temperature changes while the CLASS-CHESCA imposes changes through variations in the spin-active nuclei involved in pairwise correlations of residues. The residue ensembles identified by the CHESCA methods were found in previously identified EPAC allosteric sites. EPAC1 has also been identified as a promising drug target for cardiovascular diseases and based on structural analogues of a novel EPAC1-specific inhibitor called I942, the next aim of the work was to generate a quantitative structure activity relationship model (QSAR). The QSAR model was able to predict the affinity of a promising inhibitor with enhanced potency and inhibitory activity compared to I942 which was confirmed through competition assays, ^{15}N-^{1}H HSQC experiments, saturation transfer difference (STD) and chemical shift projection analysis (CHESPA).

Acknowledgment

I would like to express my great appreciation to my graduate thesis supervisor, Dr. Giuseppe Melacini whose immense knowledge and expertise in the field of NMR and allosteric proteins helped me a lot in understanding the foundations of my project. His guidance and mentorship supported me along the way.

I would also like to give thanks to my mentor, Hongzhao Shao who helped me gain the lab skills required to preform the experiments I need for my EPAC-related projects. I am grateful for the great support I received from all the Melacini lab members: Karla Martinez Pomier, Dr. Madoka Akimoto, Dr. Bryan VanSchouwen, Dr. Rashik Ahmed, Dr. Olivia Byun, Dr. Naeimeh Jafari, and Jinfeng Huang.

I also wanted to thank my committee members Dr. Jakob Magolan and Dr. John Whitney for their insightful feedback.

Another appreciation goes to the support from the members of the McMaster NMR facility team, Dr. Bob Berno and Dr. Hilary Jenkins.

Finally, I am grateful for the support I received from my parents and sisters; I could not have done it without them.

Table of Content

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

List of Tables

List of Abbreviations

cAMP	3',5'-Cyclic Adenosine Monophosphate
EPAC	Exchange Protein Activated by cAMP
GEF	Guanine Nucleotide Exchange Factor
GDP	Guanosine Diphosphate
GTP	Guanosine Triphosphate
CNBD	Cyclic Nucleotide Binding Domain
PBC	Phosphate Binding Cassette
NOE	Nuclear Overhauser Effect
BBR	Base Binding Region
QSAR	Quantitative Structure Activity Relationship
SPP	Similarity-Property Principle
RFI	Relative Fluorescence Intensity
8-NBD-cAMP	8-(2-[7-Nitro-4-Benzofurazanyl] Aminoethylthio) Adenosine-3',5'-Cyclic Monophosphate
K_D	Dissociation Constant
NMR	Nuclear Magnetic Resonance
HSQC	Heteronuclear Single Quantum Coherence
CCS	Compounded Chemical Shift
STD	Saturation Transfer Difference
STR	Saturation Transfer Reference
CHESPA	Chemical Shift Projection Analysis
CHESCA	Chemical Shift Covariance Analysis

Declaration of Academic Achievement

For the CHESCA analysis the author Hebatallah Mohamed, Ubaidullah Baryar, Dr. Amir Bashiri, Dr. Rajeevan Selvaratnam and Dr. Bryan VanSchouwen designed, executed, and analyzed all data. Dr. Amir Bashiri and Dr. Rajeevan Selvaratnam prepared samples and acquired all NMR data in 2012.

For the QSAR project the author developed the QSAR model. With the contribution of Dr. Giuseppe Melacini, the author designed the experimental approaches. Furthermore, the author prepared the samples and carried out data acquisition and analysis. Members of the Magolan Laboratory synthesized the chemical compounds investigated.

The content was written by Hebatallah Mohamed and co-edited by Hebatallah Mohamed and Dr. Giuseppe Melacini.

Chapter 1

The Role of Exchange Protein Activated by cAMP (EPAC) in cAMP Signaling and Importance as a Drug Target

1.1 cAMP-mediated Signaling Pathways

Cyclic adenosine 3′,5′-monophosphate (cAMP) is a second messenger that plays significant roles in mediating cellular pathways in response to extracellular signals [1] and the balance between the activities of adenylate cyclases (ACs) and cyclic nucleotide phosphodiesterases (PDEs) regulates the concentration of intracellular cAMP [2]. Most ACs become activated through interactions with the alpha subunit of the G_s protein ($G\alpha_s$) of the G-protein-coupled receptors (GPCRs) [3]. AC inhibition on the other hand, can be due to binding of ligands to GPCRs that are coupled to G_i proteins as opposed to G_s.

cAMP affects three main protein families: 1) Protein kinase A (PKA), 2) Exchange protein activated by cAMP (EPAC) and 3) Cyclic-nucleotide-gated ion channels [4]. These effectors are then involved in several cellular functions such as gene transcription, cell growth, cell adhesion and metabolism [5–7].

Cellular effects are often mediated by the cross-talk between different signalling pathways that involve second messengers [8]. For instance, the cAMP-PKA pathway is associated with the 3′,5′-cyclic guanosine monophosphate (cGMP) as the concentration of one of the nucleotides is affected by the other [9]. An antagonistic relation is observed in certain physiological processes mediated by the secondary messengers. For example, myocardial contraction can be promoted by isoproterenol, [10] which in turn results in an elevation of the cAMP levels, while simultaneously

decreasing the cGMP levels. An opposite effect is observed when administering a PDE inhibitor such as KMUP-1, a xanthine derivative that inhibits PDEs 3,4 and 5 [11]. It showed osteoclastogenic activity mediated by both the cAMP and cGMP pathways and the rise in the levels of the secondary messengers was proposed as a therapeutic avenue for osteoporosis [8,11].

1.2 Cyclic Nucleotide Binding Domains

Increasing levels of cAMP and cGMP levels activate the cAMP and cGMP-mediated protein kinases, respectively, [12] and this is through the binding of the cyclic nucleotides to the cyclic nucleotide binding domains (CNBDs). CNBDs are key regions that act as controlling units, in response to cyclic nucleotides (*e.g.,* cAMP or cGMP), for regulation of multiple cellular pathways in eukaryotes and prokaryotes [13,14]. They are usually coupled to different functional such as those of kinases, ion channels, transcription modulators and guanine nucleotide exchange factors [15,16].

CNBDs maintain an evolutionary conserved beta-subdomain that is observed in CNBD-containing proteins [15]. The beta subdomain includes eight beta strands that form a beta barrel. The β strands 6 and 7 bracket a small helical moiety known as the phosphate binding cassette (PBC), which, facilitates the interactions with the phosphate group of the cyclic nucleotide and is highly conserved among CNBDs [15,17]. Another subdomain of the CNBD is a flexible helical region referred to as the alpha subdomain [13]. The alpha subdomain includes two main non-contiguous helical motifs: 1) An N3A motif at the N-terminus and 2) a B/C helix that is located after the eighth β strand [18]. The beta subdomain is more conserved than the alpha subdomain in structure and sequence across CNBDs compared to the alpha subdomain [13].

The regulatory subunit of PKA (PKA-R), which is regarded as the prime example of protein kinases, consists of two CNBDs as well as a dimerization and a localization domain in the N-terminal region [12]. The two CNBDs of PKA-R, *i.e.*, CNBD A and CNBD B, are located sequentially in tandem after each other, whereby the C helix of CNBD A (C_A) is linked to the A helix of CNBD B (A_B) [19]. Complexes of a monomeric construct of PKA-R with two cAMP molecules show [20,21] that the phosphate group interacts with a conserved arginine residue in the PBC, whereas extended contacts are observed between the nucleotide's base and the N3A and C helix of domain B.

Hyperpolarization-activated cyclic nucleotide–gated (HCN) channels are another type of cyclic nucleotide regulated systems [22], which are activated by the hyperpolarization of cellular membranes as well as the binding of cyclic nucleotides to a CNBD located intracellularly [23]. The effect of cyclic nucleotide binding to HCN channels extends to multiple different functional features of the channels, such as, accelerating the kinetics of channel activation, shifting half-maximal activation voltage ($V_{1/2}$) towards depolarization and, elevating the maximum current [24–26]. The HCN channel consists of four domains, two of which are transmembrane and two intracellular. The transmembrane domains are the voltage sensor domain and the channel pore, and the intracellular domains are the C-linker and the CNBD [23]. The CNBD is located C-terminal to the C-linker and contains structural elements common to other CNBDs. The β-roll or the beta subdomain facilitates the binding of cyclic nucleotides, and the C-helix of the alpha subdomain regulates the effectiveness by which these cyclic nucleotides activate the ion channel [27–32].

In addition to proteins that are cAMP-regulated, cyclic GMP-dependent protein kinases (PKGs) are a class of serine/threonine kinases that are regarded as one of the main

intracellular cGMP receptors. They exist as two types in mammals: PKG I and PKG II [33–35]. Like PKA, PKG also spans two tandem CNBDs: CNBD A and CNBD B which, despite a sequence similarity of 37%, exhibit distinct binding kinetics and affinities for cGMP as well as differences in specificities for cGMP analogs [36,37] Though the overall binding affinity of PKG for cAMP is lower than that of cGMP [38–40], the intracellular cAMP concentrations are significantly higher than that of cGMP [41–43], suggesting the presence of additional factors governing cGMP selectivity. For example, comparative NMR analyses have revealed that cAMP behaves as a partial agonist for PKG [44]. The underlying mechanism relied on the sampling of an additional state other than the conventional two-state conformational model of cGMP activation, and this partially autoinhibited, third state results in partial agonism [44].

Exchange proteins directly activated by cAMP (EPAC) are [45] also a family of cAMP sensors, [46] which act as guanine nucleotide exchange factors (GEF) that activate the small GTP-binding proteins called Rap1 and Rap2 [47]. The CNBDs of the two main isoforms of EPAC: EPAC1-CNBD and EPAC2-CNBD-B, have a high sequence homology between each other and that homology is conserved between humans and other species. The percent identity ranged from 75 to 95%. CNBD-A of EPAC2 on the other hand, did not have conserved residues since EPAC1 lacks that domain [48]. EPAC CNBDs share the common CNBD structural elements. However, unlike other CNBDs, the structure of EPAC2 exhibits some variation in what is known as the 'CNBD lid' region. In HCN and PKA, the lid region is often observed as an α-helix located in the C-terminal side of the β-barrel/role [49]. The lid in EPAC2, however, exists as a two-stranded β-sheet that is part of a 'switchboard' structure composed of five beta strands, and it is positioned away from the cAMP binding region [50].

1.3 EPAC1: Domain Organisation and Mechanism of Action

EPAC1 consists of a regulatory region (RR), which contains the cyclic nucleotide binding domain (CNBD) and a disheveled Egl-10 Plectstrin (DEP) domain, and a catalytic region (CR), containing a RAS-exchange motif (REM), a RAS association (RA) domain, and a CDC25 homology domain (CDC25HD) [51–53] (Figure 1.1). In the absence of cAMP ('apo'), EPAC1 predominantly samples the autoinhibited (inactive) state, whereby the regulatory region blocks substrate access to the CR through salt bridges formed between residues in α helices 1 and 2 of the EPAC1 CNBD and the CDC25HD. These interactions are commonly referred to as the ionic latch (IL). Upon cAMP binding to the EPAC1 CNBD, the relative orientation of the regulatory and catalytic regions shifts to a more open topology in which the catalytic site occlusion is eliminated (Figure 1.2). The CR binds to Rap GTPase to catalyze the exchange of GDP for GTP and the Rap-GTPase–activating proteins (Rap-GAPs) enhance the slow, intrinsic GTP hydrolysis activity of Rap [54,55]. The EPAC1 CNBD, therefore, serves as the central controlling unit for the closed-to-open transition underlying the cAMP-dependent activation of EPAC1.

1.4 cAMP Binding and Allosteric Control of EPAC1

As previously mentioned, the CNBD of EPAC1 consists of a β-barrel which is flanked by an N-terminal helical bundle and a hinge helix located in the C-terminal and is connected to the lid region. The sugar phosphate group of cAMP interacts with the PBC, primarily through hydrogen bonds between the phosphate group and PBC residues such as A272, R279 and G269, as seen in Figure 1.3, where the EPAC1 residues are in brackets next to the analogous residues in EPAC2 [56].

Two main events take place after or during cAMP binding: 1) Rotation of the hinge helix in the CNBD C-terminus region and, 2) Weakening of the IL interactions which in turn, stabilize the open topology of EPAC and facilitate the motion of the hinge that shifts the RR away from the CR [57,58]. The mechanism by which cAMP weakens the IL has been shown, to be mainly governed by dynamics rather than structural changes [59–61]. Indeed, the comparative analysis of NMR relaxation experiments for the EPAC1-CNBD in the presence and absence of cAMP revealed that the active state exhibits an enhancement in dynamics in a large region spanned by the IL [62]. This reflects an entropic penalty imposed on the IL salt bridges, which, in turn, leads to a weakened IL and a destabilization of the EPAC autoinhibited state [38,62]. Further elucidation of the cAMP-mediated control of distal sites of the CNBD was demonstrated through NMR-based chemical shift covariance analyses (CHESCA), which mapped out the allosterically-relevant networks of EPAC [63].

The CHESCA methodology, which assumes that linear inter-residue chemical shift correlations between a pair of residues reflects their concerted response to a library of perturbations [63], has identified multiple allosteric networks. These allosteric clusters are in the α-subdomain of the CNBD and are composed of two main subclusters: 1) Hydrophobic residues positioned around R186, where R186 forms several hydrogen bonds with polar oxygen atoms, and 2) Two spines composed of hydrophobic residues in the α4 helix of the N-terminal helical bundle and the hinge helix (α6) [63]. The CHESCA results coupled with NMR comparative analyses and site-directed mutagenesis confirm that the cAMP allosterically modulates the conformations of the β2-β3 loop as well as the hinge helix [62,64,65] and facilitates the motion of both the PBC and the hinge helix from the 'out' to the 'in' conformations (Figure 1.3) [50,66].

The PBC and the hinge were also confirmed to be allosterically coupled through the interaction of a conserved leucine residues in the PBC, L273, and F300 in the hinge helix. Mutations of either those residues significantly affected the GEF activity, for example, the L273W mutant could not be activated even in the presence of saturating concentrations of cAMP. An opposite effect was observed when F300 was replaced with less bulky residues such as in the F300A and the F300T mutants which reduced the concentration of cAMP required to reach the half maximal activity of the protein [49,72,74,75].

1.5 EPAC: A Drug Target for Cardiovascular Diseases

EPAC has been investigated as a potential target for cardiac diseases due to its involvement in cardiac electric remodeling and cardiac hypertrophy [54]. For instance, an EPAC activator known as 8-CPT was shown to cause ventricular arrhythmogenesis in whole mouse hearts [68]. Other studies reported that the activation of EPAC leads to the lengthening of the action potential through the decrease of potassium current in rat ventricles [69]. The lengthening of action potentials is associated with cardiovascular diseases since it is significantly correlated with arrhythmia [70]. Another study preformed on ventricular myocytes of guinea pig [71] showed that constant stimulation of β1-adrenergic receptors (β1-AR) induces EPAC1 activation, which reduces the concentration of slow delayed rectifier potassium K^+-current (IKs). The IKs are essential regulators of cardiac repolarization and therefore, their decrease enhances the possibility of arrhythmogenesis. EPAC was also shown to raise expression levels of transient receptor potential canonical channels 3 and 4 in rat ventricular cardiomyocytes and the Ca^{2+} influx passing through these channels was proposed as a mechanism by which EPAC activation promotes arrhythmia [72]. Moreover, Okumura

et al [73] reported that knock-out mice that lacked EPAC1 exhibited a reduced chance of atrial fibrillation.

EPAC was also proposed as a contributor to cardiac hypertrophy since its expression, specifically EPAC1, is increased in various animal models exhibiting cardiomyopathy as well as samples of left ventricles obtained from patients with failing hearts [74]. The activation of EPAC or EPAC1 overexpression in rat cardiac myocytes elevated several hallmarks of hypertrophy, such as atrial natriuretic factor expression, protein synthesis and cell-surface area [75]. Furthermore, the previously mentioned sustained β-AR activation, which leads to EPAC1 activation and subsequent upregulation of hypertrophy-related genetic markers, is counter affected by deletion of EPAC1 [76]. EPAC1 knockout mice also display enhanced cardio-protection against stress-induced conditions such as age-related cardiac dysfunction [73]. In addition, EPAC1 knockout mice subjected to transverse aortic constriction or chronically treated with isoprenaline exhibited attenuated cardiac fibrosis [73,76]. Taken together, these observations strongly support the protective effects of EPAC1 deletion under stress-induced conditions.

1.6. EPAC1 Inhibitors

Given the value of human EPAC1 as a drug target, several screening campaigns have aimed at identifying leads for EPAC1 inhibition [49,77–80]. Initial efforts in the search for EPAC1 inhibitors focused on libraries of cAMP analogs, leading to the identification of cAMP antagonists, such as the phosphorothioate Rp-cAMPS and cGMP [49,56]. However, cAMP derivatives suffer from poor selectivity due to cross reactivity with other cyclic nucleotide (cNMP)-dependent systems in humans, such as PKA, PKG and HCN [81]. Furthermore, cAMP analogs are often hydrolyzed by

phosphodiesterases (PDEs), thus limiting their *in vivo* effectiveness. Hence, subsequent screening efforts turned to non-cNMP ligands resulting in the identification of four main core structures for EPAC1 inhibitors, denoted as I942, ESI09, CE3F4 and BAA (Table 1.1; Figure 1.4) [77–80]. All four types of ligands bind directly to the EPAC1 CNBD, without affecting PKA. I942 and ESI09 are competitive inhibitors of cAMP, [77,78] while CE3F4 and TBAA are un- and non-competitive EPAC1 inhibitors, respectively [79,80]. In addition, ESI09 is a pan EPAC inhibitor which inhibits both EPAC1 and EPAC2 [78].

The EPAC1-CNBD:I942 complex has been recently investigated extensively through NMR-based experiments [82] and a binding mode was proposed based on measurements of intermolecular nuclear overhauser effects (NOEs) [82–84]. Such NOEs (Figure 1.5A) are between the PBC and base binding region (BBR) residues and I942 protons in the dimethylbenzene and naphthalene moieties, respectively. Based on those NOEs as well as chemical shift perturbation analyses, I942 was proposed to mimic cAMP (Figure 1.5B), whereby the adenine base of cAMP or the naphthalene group in I942 interact with the BBR, whereas the cAMP's ribose ring or I942's dimethylbenzene group interact with the PBC region [82].

1.7 SAR Studies of I942 Derivatives

A recent study by the Yarwood research group developed structure-activity relationships (SARs) for derivatives of the EPAC1-selective partial agonist, I942 [85]. The study involved obtaining a correlation between structural elements of a library of I942 analogues and their respective relative binding affinities. They showed that a naphthalene ring is essential for EPAC1-binding and that a substituent on position 7 (as per numbering on I942 in Figure 1.4) that is

electron-donating enhances the affinity. The linker region, *i.e.*, the moiety connecting the sulfonamide group to the naphthalene group, was demonstrated to be quite crucial since its modification compromised the binding. Furthermore, substituents on the phenyl moiety that are electron-donating proved to be better than electron-withdrawing groups and, the replacement of the phenyl group with groups of greater π-conjugation also raised the potency levels towards EPAC1.

Evaluation of the effect of promising analogs on the EPAC1-regulated signaling pathway leading to induction of the pro-inflammatory VCAM1 cell adhesion protein, revealed that a compound called '25u' demonstrated greater activation compared to the parent molecule, I942 [85]. Although the I942-based SAR indicated which substituents enhanced or weakened the binding affinity towards the EPAC1-CNBD, the relationships are relatively qualitative in nature. More quantitative approaches, *i.e.*, the application of quantitative structure activity relationships (QSARs), were still lacking for EPAC.

1.8 Quantitative Structure Activity Relationships (QSARs)

Quantitative structure-activity relationship (QSAR) models describe a correlation, of statistical significance, between the target property of small molecules such as, bioactivity, and the molecular descriptors of those molecules, which are calculated from the molecular structure [86,87]. The concept of QSAR modeling was introduced in 1964 by Hansch and Fujita [88]. Since then, it has been extensively applied in the computer-aided drug discovery process [89]. QSAR modeling relies on the idea that compounds that share structural similarity often display comparable biological activity, and this is known as the similarity-property principle (SPP). The SPP proposes

that slight structural modifications of a compound correspond to slight variation in the biological property, such as potency, of that compound, which, in turn, creates the foundation for linear relations that QSAR models generate.

Analogs of a congeneric series share a common 'scaffold' making them chemically similar and it is the gradual structural changes through different substituents on that shared scaffold that result in variation in potency. QSAR models can, therefore, utilize these linear variations for potency predictions of molecules with a shared scaffold and varying substituents. The applicability of these QSAR predictions, however, heavily relies on the SPP and is based the concept of 'SAR continuity' which ensures a linear relation between relatively conserved structural modifications and the corresponding, minor, potency variations [87].

Most of the models are multivariate in nature and offer different means of regression analysis. However, multiple linear regression (MLR) is one of the most widely used methods as it produces a direct, linear correlation between a property of interest (Y) and the molecular descriptors (X_j):

$$Y = b_0 + \sum_{j=1}^{p} b_j \times X_j = b_0 + b_1 X_1 + b_2 X_2 + \cdots b_p X_p \tag{1.1}$$

where Y is a vector of n elements with n representing the number of molecules studied, X is a matrix whose size is $n \times p$ where p is the number of descriptors and the b_j values represent the regression coefficients. The coefficient is a measure of the descriptor's weighted contribution to the QSAR prediction, and the sign (positive or negative) reflects how the descriptor contributes to the target property [90].

1.9 QSAR of Protein Kinase A (PKA)

The application of QSAR models to proteins in the cyclic adenosine monophosphate (cAMP)-dependent signaling pathway goes back to the 1990s [91],even before the first structure of PKA-R1α was resolved [92], and though many advancements in QSAR methods have been developed, only a limited number of QSARs currently exist for the different proteins involved in the cAMP-mediated signaling cascade. One of the earliest papers that discusses QSARs for cAMP-dependent kinases was a study conducted by Mureşan et al. [91] who examined 27 cAMP analogues that contained substitutions in positions 1, 2, 6 and 8 as well as diastereoisomeric phosphorothioate cAMP analogs where the sulfur atom was either in the axial (Sp-cAMPS) or equatorial (Rp-cAMPS) position (Figure 1.6).

They implemented a QSAR method known as the minimal steric difference method (MTD) to map out four different binding sites for the regulatory subunits of protein kinases I and II: AI, BI, AII and BII. MTD measures the steric misfit between the studied compound and the cavity of the receptor [93]. The QSAR coefficients ranged between 0.836 and 0.948 and provided insight on the steric characteristics of each type of receptor site. The AI and BI receptor sites were found to contain a negatively charged moiety able to interact with cAMP derivatives with modifications at position 6. Both BI and BII were determined to be hydrophobic in nature and derivatives with thiophosphoric acid groups were reported to have reduced affinities for the four receptor sites.

The same research group further explored cAMP derivatives in a separate study but with a greater focus on derivatives containing bulky groups on positions 2, 6 and 8 [94]. In addition to MTD, which is primarily a measure of steric contribution, they included parameters such as the predicted base moiety hydrophobicities as well as the charge of substituents on position 6. QSAR

correlations that included multiple parameters rather than the MTD parameter alone yielded higher correlation values [95].

1.10 Thesis Outline

1.10.1 Thesis Objective

The goal of this thesis is to explore the EPAC1 system through two main approaches. The first is to establish an NMR-based toolset that identifies the core allosteric network using EPAC1 as the model system, while the second approach aims at developing a quantitative structure-activity relationship based on a series of EPAC1-selctive modulators.

1.10.2 Chapter Outlines

Chapter 2 will discuss two new proposed methods of CHESCA methodologies to narrow down the allosteric maps of the EPAC1-CNBD to the core allosteric residues. The first method is the temperature CHESCA (T-CHESCA), which is based on the ^{1}H-^{15}N-HSQC readout of EPAC1 in the apo and four other ligand-bound states, acquired at different temperatures, and the second method is the CLASS-CHESCA, which is be based on the ^{1}H-^{15}N-HSQC readout of the five different states of EPAC1-CNBD acquired at a single temperature (*e.g.,* 306K). T-CHESCA selects allosterically coupled residues whose chemical shifts remain strongly correlated across different temperatures, while the CLASS-CHESCA identifies tightly allosteric couplings based on the number of correlations, above a certain threshold, observed between different nuclei within a given residue pair. Networks identified from these two CHESCA methods are then compared with the network identified from complete linkage clustering obtained from the conventional CHESCA analysis [63].

Chapter 3 discusses a QSAR model developed for a library of EPAC1-selective sulfonamide modulators [85]. The goal of the model is to be able to predict the affinities of *de-novo* compounds that still share a similar skeleton as that of I942 [77]. The model is first trained and validated using structures with known affinities [85] and is then utilized as a tool to predict affinities for a set of compounds that are 'unknown' to the model. Predictions are then experimentally validated through a competition assay and compared to I942. Further insight into the possible binding mode and mechanism of action of the most promising compound is obtained through HSQC, STD, and chemical shift projection analyses (CHESPA) [65].

Chapter 4 is a summary of the main findings of the two studies conducted, *i.e.*, the CHESCA toolset, and the QSAR model, and provides a further outlook on future directions. The common ensemble of residues shared between the two, newly proposed CHESCA methods and the previously established allosteric network seem to be critical sites in the allosteric regulation of EPAC1 and can therefore, be prioritized for the design of allosteric drugs. The QSAR model is also promising in terms of predictive power based on the experimentally-measured affinities that confirmed the QSAR predictions.

1.11 References

(1)   Sutherland, E. W.; Rall, T. W. FRACTIONATION AND CHARACTERIZATION OF A CYCLIC ADENINE RIBONUCLEOTIDE FORMED BY TISSUE PARTICLES*. **1958**.

(2)   Stanley McKnight, G. Cyclic AMP Second Messenger Systems. *Current Opinion in Cell Biology* **1991**, *3* (2), 213–217.

(3)   Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J. Seven-Transmembrane Receptors. *Nature Reviews Molecular Cell Biology 2002 3:9* **2002**, *3* (9), 639–650.

(4)   Sassone-Corsi, P. The Cyclic AMP Pathway. *Cold Spring Harbor Perspectives in Biology* **2012**, *4* (12), a011148–a011148.

(5)   Jackson, E. K.; Dubey, R. K. Role of the Extracellular CAMP-Adenosine Pathway in Renal Physiology. *American journal of physiology. Renal physiology* **2001**, *281* (4).

(6)   Seino, S.; Shibasaki, T. PKA-Dependent and PKA-Independent Pathways for CAMP-Regulated Exocytosis. *Physiological reviews* **2005**, *85* (4), 1303–1342.

(7)   Richards, J. S. New Signaling Pathways for Hormones and Cyclic Adenosine 3',5'-Monophosphate Action in Endocrine Cells. *Molecular endocrinology (Baltimore, Md.)* **2001**, *15* (2), 209–218.

(8)   Yan, K.; Gao, L. N.; Cui, Y. L.; Zhang, Y.; Zhou, X. The Cyclic AMP Signaling Pathway: Exploring Targets for Successful Drug Discovery (Review). *Molecular Medicine Reports* **2016**, *13* (5), 3715–3723.

(9)   Denninger, J. W.; Marletta, M. A. Guanylate Cyclase and the .NO/CGMP Signaling Pathway. *Biochimica et biophysica acta* **1999**, *1411* (2–3), 334–350.

(10)   Edgar, V. A.; Cremaschi, G. A.; Sterin-Borda, L.; Genaro, A. M. Altered Expression of Autonomic Neurotransmitter Receptors and Proliferative Responses in Lymphocytes from a Chronic Mild Stress Model of Depression: Effects of Fluoxetine. *Brain, Behavior, and Immunity* **2002**, *16* (4), 333–350.

(11)   Liou, S. F.; Hsu, J. H.; Lin, I. L.; Ho, M. L.; Hsu, P. C.; Chen, L. W.; Chen, I. J.; Yeh, J. L. KMUP-1 Suppresses RANKL-Induced Osteoclastogenesis and Prevents Ovariectomy-Induced Bone Loss: Roles of MAPKs, Akt, NF-KB and Calcium/Calcineurin/NFATc1 Pathways. *PLOS ONE* **2013**, *8* (7), e69468.

(12)   Mohanty, S.; Kennedy, E. J.; Herberg, F. W.; Hui, R.; Taylor, S. S.; Langsley, G.; Kannan, N. Structural and Evolutionary Divergence of Cyclic Nucleotide Binding Domains in Eukaryotic Pathogens: Implications for Drug Design. *Biochimica et biophysica acta* **2015**, *1854* (10 0 0), 1575.

(13)   Kannan, N.; Wu, J.; Anand, G. S.; Yooseph, S.; Neuwald, A. F.; Venter, J. C.; Taylor, S. S. Evolution of Allostery in the Cyclic Nucleotide Binding Module. *Genome Biology* **2007**, *8* (12), R264.

(14)   McCue, L. A.; McDonough, K. A.; Lawrence, C. E. Functional Classification of CNMP-Binding Proteins and Nucleotide Cyclases with Implications for Novel Regulatory Pathways in Mycobacterium Tuberculosis. *Genome research* **2000**, *10* (2), 204–219.

(15)   Berman, H. M.; ten Eyck, L. F.; Goodsell, D. S.; Haste, N. M.; Kornev, A.; Taylor, S. S. The CAMP Binding Domain: An Ancient Signaling Module. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (1), 45–50.

(16)   Shabb, J. B.; Corbin, J. D. THE JOURNAL OF BIOLOGICAL CHEMISTRY Cyclic Nucleotide-Binding Domains in Proteins Having Diverse Functions*. *Journal of Biological Chemistry* **1992**, *267*, 5723–5726.

(17)   Diller, T. C.; Madhusudan, M.; Xuong, N. H.; Taylor, S. S. Molecular Basis for Regulatory Subunit Diversity in CAMP-Dependent Protein Kinase: Crystal Structure of the Type IIβ Regulatory Subunit. *Structure* **2001**, *9* (1), 73–82.

(18)   Kornev, A. P.; Taylor, S. S.; ten Eyck, L. F. A Generalized Allosteric Mechanism for Cis-Regulated Cyclic Nucleotide Binding Domains. *PLoS computational biology* **2008**, *4* (4).

(19)   Pantano, S.; Zaccolo, M.; Carloni, P. Molecular Basis of the Allosteric Mechanism of CAMP in the Regulatory PKA Subunit. *FEBS Letters* **2005**, *579* (12), 2679–2685.

(20)   Su, Y.; Dostmann, W. R. G.; Herberg, F. W.; Durick, K.; Xuong, N. H.; ten Eyck, L.; Taylor, S. S.; Varughese, K. I. Regulatory Subunit of Protein Kinase A: Structure of Deletion Mutant with CAMP Binding Domains. *Science (New York, N.Y.)* **1995**, *269* (5225), 807–813.

(21)   Diller, T. C.; Madhusudan, M.; Xuong, N. H.; Taylor, S. S. Molecular Basis for Regulatory Subunit Diversity in CAMP-Dependent Protein Kinase: Crystal Structure of the Type II Beta Regulatory Subunit. *Structure (London, England : 1993)* **2001**, *9* (1), 73–82.

(22)   Lee, C. H.; MacKinnon, R. Structures of the Human HCN1 Hyperpolarization-Activated Channel. *Cell* **2017**, *168* (1–2), 111-120.e11.

(23)   James, Z. M.; Zagotta, W. N. Structural Insights into the Mechanisms of CNBD Channel Function. *The Journal of General Physiology* **2018**, *150* (2), 225.

(24)   Wang, J.; Chen, S.; Siegelbaum, S. A. Regulation of Hyperpolarization-Activated HCN Channel Gating and CAMP Modulation Due to Interactions of COOH Terminus and Core Transmembrane Regions. *The Journal of general physiology* **2001**, *118* (3), 237–250.

(25)   Craven, K. B.; Zagotta, W. N. Salt Bridges and Gating in the COOH-Terminal Region of HCN2 and CNGA1 Channels. *The Journal of general physiology* **2004**, *124* (6), 663–677.

(26)   Zhou, L.; Olivier, N. B.; Yao, H.; Young, E. C.; Siegelbaum, S. A. A Conserved Tripeptide in CNG and HCN Channels Regulates Ligand Gating by Controlling C-Terminal Oligomerization. *Neuron* **2004**, *44* (5), 823–834.

(27)   Gordon, S. E.; Zagotta, W. N. Localization of Regions Affecting an Allosteric Transition in Cyclic Nucleotide-Activated Channels. *Neuron* **1995**, *14* (4), 857–864.

(28)   Varnum, M. D.; Black, K. D.; Zagotta, W. N. Molecular Mechanism for Ligand Discrimination of Cyclic Nucleotide-Gated Channels. *Neuron* **1995**, *15* (3), 619–625.

(29)   Gordon, S. E.; Oakley, J. C.; Varnum, M. D.; Zagotta, W. N. Altered Ligand Specificity by Protonation in the Ligand Binding Domain of Cyclic Nucleotide-Gated Channels. *Biochemistry* **1996**, *35* (13), 3994–4001.

(30)   Flynn, G. E.; Black, K. D.; Islas, L. D.; Sankaran, B.; Zagotta, W. N. Structure and Rearrangements in the Carboxy-Terminal Region of SpIH Channels. *Structure (London, England : 1993)* **2007**, *15* (6), 671–682.

(31)   Zhou, L.; Siegelbaum, S. A. Gating of HCN Channels by Cyclic Nucleotides: Residue Contacts That Underlie Ligand Binding, Selectivity, and Efficacy. *Structure (London, England : 1993)* **2007**, *15* (6), 655–670.

(32)   Akimoto, M.; Zhang, Z.; Boulton, S.; Selvaratnam, R.; van Schouwen, B.; Gloyd, M.; Accili, E. A.; Lange, O. F.; Melacini, G. A Mechanism for the Auto-Inhibition of Hyperpolarization-Activated Cyclic Nucleotide-Gated (HCN) Channel Opening and Its Relief by CAMP. *The Journal of biological chemistry* **2014**, *289* (32), 22205–22220.

(33)   Hofmann, F.; Ammendola, A.; Schlossmann, J. Rising behind NO: CGMP-Dependent Protein Kinases. *Journal of Cell Science* **2000**, *113* (10), 1671–1676.

(34)   Hofmann, F.; Bernhard, D.; Lukowski, R.; Weinmeister, P. CGMP Regulated Protein Kinases (CGK). *Handbook of Experimental Pharmacology* **2009**, *191*, 137–162.

(35)   Hofmann, F.; Feil, R.; Kleppisch, T.; Schlossmann, J. Function of CGMP-Dependent Protein Kinases as Revealed by Gene Deletion. *Physiological Reviews* **2006**, *86* (1), 1–23.

(36)   CorbinS, J. D.; Millerll, J. P.; Suvall, R. H.; Jastorffll, B. THE JOURNAL OF BIOLOGICAL CHEMISTRY Studies of CGMP Analog Specificity and Function of the Two Intrasubunit Binding Sites of CGMP-Dependent Protein Kinase". **1986**, *261* (3), 1208–1214.

(37)   Reed, R. B.; Sandberg, M.; Jahnsen, T.; Lohmann, S. M.; Francis, S. H.; Corbin, J. D. Fast and Slow Cyclic Nucleotide-Dissociation Sites in CAMP-Dependent Protein Kinase Are Transposed in Type Iβ CGMP-Dependent Protein Kinase *. *Journal of Biological Chemistry* **1996**, *271* (29), 17570–17575.

(38)   Das, R.; Chowdhury, S.; Mazhab-Jafari, M. T.; SilDas, S.; Selvaratnam, R.; Melacini, G. Dynamically Driven Ligand Selectivity in Cyclic Nucleotide Binding Domains. *The Journal of biological chemistry* **2009**, *284* (35), 23682–23696.

(39)   De, S.; Greenwood, A. I.; Rogals, M. J.; Kovrigin, E. L.; Kun, ‡; Lu, P.; Nicholson, L. K. Complete Thermodynamic and Kinetic Characterization of the Isomer-Specific Interaction between Pin1-WW Domain and the Amyloid Precursor Protein Cytoplasmic Tail Phosphorylated at Thr668. **2012**.

(40)   Greenwood, A. I.; Kwon, J.; Nicholson, L. K. Isomerase-Catalyzed Binding of Interleukin-1 Receptor-Associated Kinase 1 to the EVH1 Domain of Vasodilator-Stimulated Phosphoprotein. *Biochemistry* **2014**, *53* (22), 3593–3607.

(41)   Beavo, J. A.; Brunton, L. L. Cyclic Nucleotide Research -- Still Expanding after Half a Century. *Nature reviews. Molecular cell biology* **2002**, *3* (9), 710–718.

(42)   Börner, S.; Schwede, F.; Schlipp, A.; Berisha, F.; Calebiro, D.; Lohse, M. J.; Nikolaev, V. O. FRET Measurements of Intracellular CAMP Concentrations and CAMP Analog Permeability in Intact Cells. *Nature Protocols 2011 6:4* **2011**, *6* (4), 427–438.

(43)   Wolter, S.; Golombek, M.; Seifert, R. Differential Activation of CAMP-and CGMP-Dependent Protein Kinases by Cyclic Purine and Pyrimidine Nucleotides. **2011**.

(44)   VanSchouwen, B.; Selvaratnam, R.; Giri, R.; Lorenz, R.; Herberg, F. W.; Kim, C.; Melacini, G. Mechanism of CAMP Partial Agonism in Protein Kinase G (PKG). *Journal of Biological Chemistry* **2015**, *290* (48), 28631–28641.

(45)   Kawasaki, H.; Springett, G. M.; Mochizuki, N.; Toki, S.; Nakaya, M.; Matsuda, M.; Housman, D. E.; Graybiel, A. M. A Family of CAMP-Binding Proteins That Directly Activate Rap1. *Science* **1998**, *282* (5397), 2275–2279.

(46)   de Rooij, J.; Zwartkruis, F. J. T.; Verheijen, M. H. G.; Cool, R. H.; Nijman, S. M. B.; Wittinghofer, A.; Bos, J. L. Epac Is a Rap1 Guanine-Nucleotide-Exchange Factor Directly Activated by Cyclic AMP. *Nature* **1998**, *396* (6710), 474–477.

(47)   Tsalkova, T.; Blumenthal, D. K.; Mei, F. C.; White, M. A.; Cheng, X. Mechanism of Epac Activation: STRUCTURAL AND FUNCTIONAL ANALYSES OF Epac2 HINGE MUTANTS WITH CONSTITUTIVE AND REDUCED ACTIVITIES*. *The Journal of Biological Chemistry* **2009**, *284* (35), 23644.

(48)   Ni, Z.; Cheng, X. Origin and Isoform Specific Functions of Exchange Proteins Directly Activated by Camp: A Phylogenetic Analysis. *Cells* **2021**, *10* (10).

(49)   Rehmann, H.; Wittinghofer, A.; Bos, J. L. Capturing Cyclic Nucleotides in Action: Snapshots from Crystallographic Studies. *Nature reviews. Molecular cell biology* **2007**, *8* (1), 63–73.

(50)   Rehmann, H.; Das, J.; Knipscheer, P.; Wittinghofer, A.; Bos, J. L. Structure of the Cyclic-AMP-Responsive Exchange Factor Epac2 in Its Auto-Inhibited State. *Nature* **2006**, *439* (7076), 625–628.

(51)   Schmidt, M.; Dekker, F. J.; Maarsingh, H. Exchange Protein Directly Activated by CAMP (Epac): A Multidomain CAMP Mediator in the Regulation of Diverse Biological Functions. *Pharmacological Reviews* **2013**, *65* (2), 670–709.

(52)   Robichaux, W. G.; Cheng, X. Intracellular CAMP Sensor EPAC: Physiology, Pathophysiology, and Therapeutics Development. *Physiological Reviews* **2018**, *98* (2), 919–1053.

(53)   Bos, J. L. Epac Proteins: Multi-Purpose CAMP Targets. *Trends in Biochemical Sciences* **2006**, *31* (12), 680–686.

(54)   Lezoualc'H, F.; Fazal, L.; Laudette, M.; Conte, C. Cyclic AMP Sensor EPAC Proteins and Their Role in Cardiovascular Function and Disease. *Circulation Research* **2016**, *118* (5), 881–897.

(55)   Roscioni, S. S.; Elzinga, C. R. S.; Schmidt, M. Epac: Effectors and Biological Functions. *Naunyn-Schmiedeberg's Archives of Pharmacology* **2008**, *377* (4–6), 345–357.

(56)   Rehmann, H.; Schwede, F.; Doøskeland, S. O.; Wittinghofer, A.; Bos, J. L. Ligand-Mediated Activation of the CAMP-Responsive Guanine Nucleotide Exchange Factor Epac *. *Journal of Biological Chemistry* **2003**, *278* (40), 38548–38556.

(57)   Kraemer, A.; Rehmann, H. R.; Cool, R. H.; Theiss, C.; de Rooij, J.; Bos, J. L.; Wittinghofer, A. Dynamic Interaction of CAMP with the Rap Guanine-Nucleotide Exchange Factor Epac1. *Journal of molecular biology* **2001**, *306* (5), 1167–1177.

(58)   Selvaratnam, R.; Akimoto, M.; Vanschouwen, B.; Melacini, G. CAMP-Dependent Allostery and Dynamics in Epac: An NMR View.

(59)   Smock, R. G.; Gierasch, L. M. Sending Signals Dynamically. *Science (New York, N.Y.)* **2009**, *324* (5924), 198–203.

(60)   Tsai, C. J.; del Sol, A.; Nussinov, R. Allostery: Absence of a Change in Shape Does Not Imply That Allostery Is Not at Play. *Journal of Molecular Biology* **2008**, *378* (1), 1–11.

(61)   Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. **2009**.

(62)   Das, R.; Mazhab-Jafari, M. T.; Chowdhury, S.; SilDas, S.; Selvaratnam, R.; Melacini, G. Entropy-Driven CAMP-Dependent Allosteric Control of Inhibitory Interactions in Exchange Proteins Directly Activated by CAMP. *Journal of Biological Chemistry* **2008**, *283* (28), 19691–19703.

(63)   Selvaratnam, R.; Chowdhury, S.; VanSchouwena, B.; Melacini, G. Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108* (15), 6133–6138.

(64)   Mazhab-Jafari, M. T.; Das, R.; Fotheringham, S. A.; SilDas, S.; Chowdhury, S.; Melacini, G. Understanding CAMP-Dependent Allostery by NMR Spectroscopy: Comparative Analysis of the EPAC1 CAMP-Binding Domain in Its Apo and CAMP-Bound States. *undefined* **2007**, *129* (46), 14482–14492.

(65)   Selvaratnam, R.; Vanschouwen, B.; Fogolari, F.; Mazhab-Jafari, M. T.; Das, R.; Melacini, G. The Projection Analysis of NMR Chemical Shifts Reveals Extended EPAC Autoinhibition Determinants. *Biophysical Journal* **2012**, *102* (3), 630.

(66)   Rehmann, H.; Arias-Palomo, E.; Hadders, M. A.; Schwede, F.; Llorca, O.; Bos, J. L. Structure of Epac2 in Complex with a Cyclic AMP Analogue and RAP1B. *Nature* **2008**, *455* (7209), 124–127.

(67)   Rehmann, H.; Prakash, B.; Wolf, E.; Rueppel, A.; de Rooij, J.; Bos, J. L.; Wittinghofer, A. Structure and Regulation of the CAMP-Binding Domains of Epac2. *Nature Structural Biology* **2003**, *10* (1), 26–32.

(68)   Hothi, S. S.; Gurung, I. S.; Heathcote, J. C.; Zhang, Y.; Booth, S. W.; Skepper, J. N.; Grace, A. A.; Huang, C. L. H. Epac Activation, Altered Calcium Homeostasis and Ventricular Arrhythmogenesis in the Murine Heart. *Pflugers Archiv European Journal of Physiology* **2008**, *457* (2), 253–270.

(69)   Brette, F.; Blandin, E.; Simard, C.; Guinamard, R.; Sallé, L. Epac Activator Critically Regulates Action Potential Duration by Decreasing Potassium Current in Rat Adult Ventricle. **2013**.

(70)   Nattel, S.; Maguy, A.; le Bouter, S.; Yeh, Y. H. Arrhythmogenic Ion-Channel Remodeling in the Heart: Heart Failure, Myocardial Infarction, and Atrial Fibrillation. *Physiological Reviews* **2007**, *87* (2), 425–456.

(71)   Aflaki, M.; Qi, X. Y.; Xiao, L.; Ordog, B.; Tadevosyan, A.; Luo, X.; Maguy, A.; Shi, Y.; Tardif, J. C.; Nattel, S. Exchange Protein Directly Activated by CAMP Mediates Slow Delayed-Rectifier Current Remodeling by Sustained β-Adrenergic Activation in Guinea Pig Hearts. *Circulation Research* **2014**, *114* (6), 993–1003.

(72)   Domínguez-Rodríguez, A.; Ruiz-Hurtado, G.; Sabourin, J.; Maria Gómez, A.; Alvarez, J. L.; Benitah, J.-P. Proarrhythmic Effect of Sustained EPAC Activation on TRPC3/4 in Rat Ventricular Cardiomyocytes. **2015**.

(73)   Okumura, S.; Fujita, T.; Cai, W.; Jin, M.; Namekata, I.; Mototani, Y.; Jin, H.; Ohnuki, Y.; Tsuneoka, Y.; Kurotani, R.; Suita, K.; Kawakami, Y.; Hamaguchi, S.; Abe, T.; Kiyonari, H.; Tsunematsu, T.; Bai, Y.; Suzuki, S.; Hidaka, Y.; Umemura, M.; Ichikawa, Y.; Yokoyama, U.; Sato, M.; Ishikawa, F.; Izumi-Nakaseko, H.; Adachi-Akahane, S.; Tanaka, H.; Ishikawa, Y. Epac1-Dependent Phospholamban Phosphorylation Mediates the Cardiac Response to Stresses. *The Journal of Clinical Investigation* **2014**, *124* (6), 2785–2801.

(74)   Ulucan, C.; Wang, X.; Baljinnyam, E.; Bai, Y.; Okumura, S.; Sato, M.; Minamisawa, S.; Hirotani, S.; Ishikawa, Y. Developmental Changes in Gene Expression of Epac and Its Upregulation in Myocardial Hypertrophy. *American Journal of Physiology - Heart and Circulatory Physiology* **2007**, *293* (3), 1662–1672.

(75)   Morel, E.; Marcantoni, A.; Gastineau, M.; Birkedal, R.; Rochais, F.; Garnier, A.; Lompré, A. M.; Vandecasteele, G.; Lezoualc'h, F. CAMP-Binding Protein Epac Induces Cardiomyocyte Hypertrophy. *Circulation Research* **2005**, *97* (12), 1296–1304.

(76)   Laurent, A. C.; Bisserier, M.; Lucas, A.; Tortosa, F.; Roumieux, M.; de Régibus, A.; Swiader, A.; Sainte-Marie, Y.; Heymes, C.; Vindis, C.; Lezoualc'h, F. Exchange Protein Directly Activated by CAMP 1 Promotes Autophagy during Cardiomyocyte Hypertrophy. *Cardiovascular Research* **2015**, *105* (1), 55–64.

(77)   Parnell, E.; Mcelroy, S. P.; Wiejak, J.; Baillie, G. L.; Porter, A.; Adams, D. R.; Rehmann, H.; Smith, B. O.; Yarwood, S. J. Identification of a Novel, Small Molecule Partial Agonist for the Cyclic AMP Sensor, EPAC1 OPEN.

(78)   Almahariq, M.; Tsalkova, T.; Mei, F. C.; Chen, H.; Zhou, J.; Sastry, S. K.; Schwede, F.; Cheng, X. A Novel EPAC-Specific Inhibitor Suppresses Pancreatic Cancer Cell Migration and Invasion. *Molecular Pharmacology* **2013**, *83* (1), 122–128.

(79)   Courilleau, D.; Bisserier, M.; Jullian, J.-C.; Lucas, A.; Bouyssou, P.; Fischmeister, R.; Blondeau, J.-P.; Lezoualc'h, F. Identification of a Tetrahydroquinoline Analog as a Pharmacological Inhibitor of the CAMP-Binding Protein Epac. *Journal of Biological Chemistry* **2012**, *287* (53), 44192–44202.

(80)   Brown, L. M.; Rogers, K. E.; Aroonsakool, N.; McCammon, J. A.; Insel, P. A. Allosteric Inhibition of Epac: Computational Modeling and Experimental Validation to Identify Allosteric Sites and Inhibitors. *The Journal of biological chemistry* **2014**, *289* (42), 29148–29157.

(81)   Tao, X.; Mei, F.; Agrawal, A.; Peters, C. J.; Ksiazek, T. G.; Cheng, X.; Tseng, C.-T. K. Blocking of Exchange Proteins Directly Activated by CAMP Leads to Reduced Replication of Middle East Respiratory Syndrome Coronavirus. *Journal of virology* **2014**, *88* (7), 3902–3910.

(82)   Shao, H.; Mohamed, H.; Boulton, S.; Huang, J.; Wang, P.; Chen, H.; Zhou, J.; Luchowska-Stańska, U.; Jentsch, N. G.; Armstrong, A. L.; Magolan, J.; Yarwood, S.; Melacini, G. Mechanism of Action of an EPAC1-Selective Competitive Partial Agonist. *Journal of Medicinal Chemistry* **2020**, *63* (9), 4762–4775.

(83)   Zwahlen, C.; Legault, P.; Vincent, S. J. F.; Greenblatt, J.; Konrat, R.; Kay, L. E. Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage λ N-Peptide/BoxB RNA Complex. **1997**.

(84)   Melacini, G. Separation of Intra-and Intermolecular NOEs through Simultaneous Editing and J-Compensated Filtering: A 4D Quadrature-Free Constant-Time J-Resolved Approach. **2000**.

(85)   Wang, P.; Luchowska-Stańska, U.; van Basten, B.; Chen, H.; Liu, Z.; Wiejak, J.; Whelan, P.; Morgan, D.; Lochhead, E.; Barker, G.; Rehmann, H.; Yarwood, S. J.; Zhou, J. Synthesis and Biochemical Evaluation of Noncyclic Nucleotide Exchange Proteins Directly Activated by CAMP 1 (EPAC1) Regulators. *Journal of Medicinal Chemistry* **2020**, *63* (10), 5159–5184.

(86)   Neves, B. J.; Braga, R. C.; Melo-Filho, C. C.; Moreira-Filho, J. T.; Muratov, E. N.; Andrade, C. H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Frontiers in Pharmacology* **2018**, *9* (NOV), 1–7.

(87)   Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. v.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without Borders. *Chemical Society Reviews* **2020**, *49* (11), 3525–3564.

(88)   Hansch, Corwin.; Fujita, Toshio. P -σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society* **1964**, *86* (8), 1616–1626.

(89)   Golbraikh, A.; Wang, X. S.; Zhu, H.; Tropsha, A. Predictive QSAR Modeling: Methods and Applications in Drug Discovery and Chemical Risk Assessment. In *Handbook of Computational Chemistry*; Springer Netherlands: Dordrecht, 2016; pp 1–48.

(90)   Gramatica, P. On the Development and Validation of QSAR Models. In *Computational Toxicology: Volume II*; Reisfeld Brad and Mayeno, A. N., Ed.; Humana Press: Totowa, NJ, 2013; pp 499–526.

(91)   Mureşan, S.; Jastorff, B.; Kurunczi, L.; Simon, Z. Comparative Structure-Affinity Relations by MTD for Binding of Cycloadenosine Monophosphate Derivatives to Protein Kinase Receptors. *Quantitative Structure-Activity Relationships* **1995**, *14* (3), 242–248.

(92)   Su, Y.; Dostmann, W. R. G.; Herberg, F. W.; Durick, K.; Xuong, N.; Eyck, L. ten; Taylor, S. S.; Varughese, K. I. Regulatory Subunit of Protein Kinase A: Structure of Deletion Mutant with CAMP Binding Domains. *Science* **1995**, *269* (5225), 807–814.

(93)   Mracec, M.; Mracec, M.; Kurunczi, L.; Nusser, T.; Simon, Z.; Náray-Szabó, G. QSAR Study with Steric (MTD), Electronic and Hydrophobicity Parameters on Psychotomimetic Phenylalkylamines. *Journal of Molecular Structure: THEOCHEM* **1996**, *367* (1–3), 139–149.

(94)   ØGREID, D.; DOSTMANN, W.; GENIESER, H. -G; NIEMANN, P.; DØSKELAND, S. O.; JASTORFF, B. (Rp)- and (Sp)-8-Piperidino-Adenosine 3',5'-(Cyclic)Thiophosphates Discriminate Completely between Site A and B of the Regulatory Subunits of CAMP-Dependent Protein Kinase Type I and II. *European journal of biochemistry* **1994**, *221* (3), 1089–1094.

(95)   Mureşan, S.; Bologa, C.; Mracec, M.; Chiriac, A.; Jastorff, B.; Simon, Z.; Náray-Szabó, G. Comparative QSAR Study with Electronic and Steric Parameters for CAMP Derivatives with Large Substituents in Positions 2, 6 and 8. *Journal of Molecular Structure: THEOCHEM* **1995**, *342* (C), 161–171.
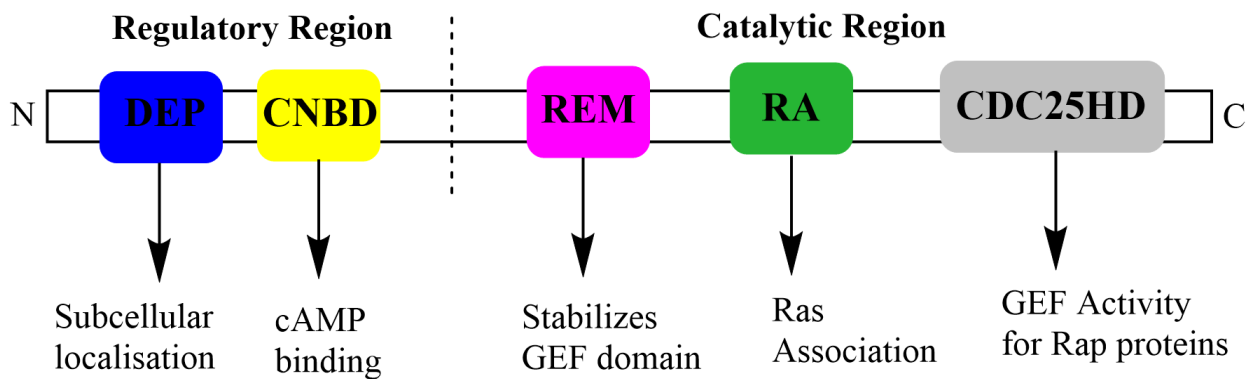
Figure 1.1: EPAC1 domain organization with corresponding function.
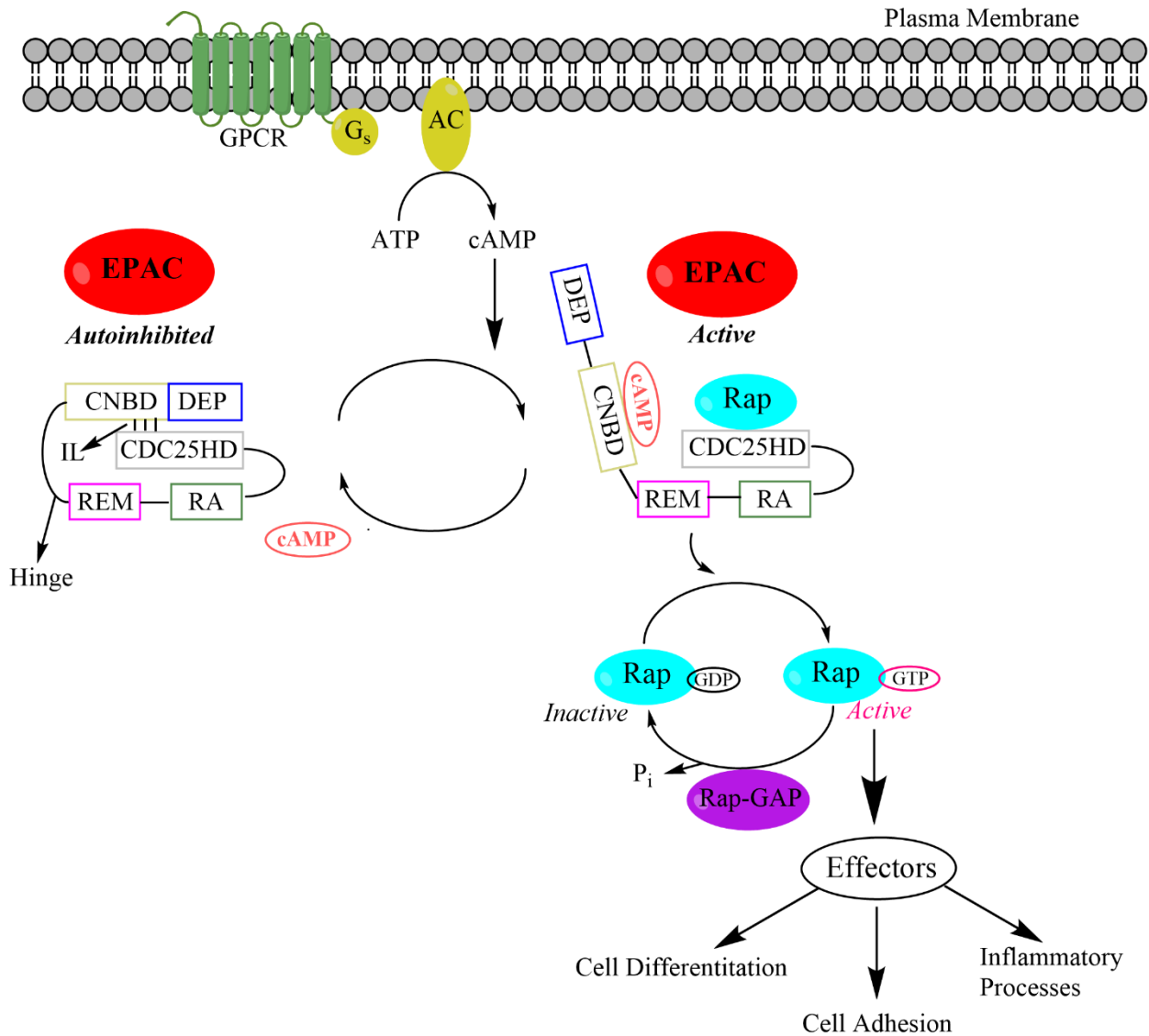
Figure 1.2: EPAC Activation Mechanism. Adenylyl cyclase (AC) produces cAMP from ATP in response to $G\alpha_s$-coupled G protein–coupled receptors (GPCRs) stimulation. Binding of cAMP to the cyclic nucleotide–binding domain (CNBD) of EPAC induces conformational changes leading to exposure of the catalytic region for binding of Rap GTPase to catalyze the exchange of GDP for GTP. Rap-GTPase–activating proteins (Rap-GAPs) enhance the slow, intrinsic GTP hydrolysis activity of Rap leading to GTPase inactivation.
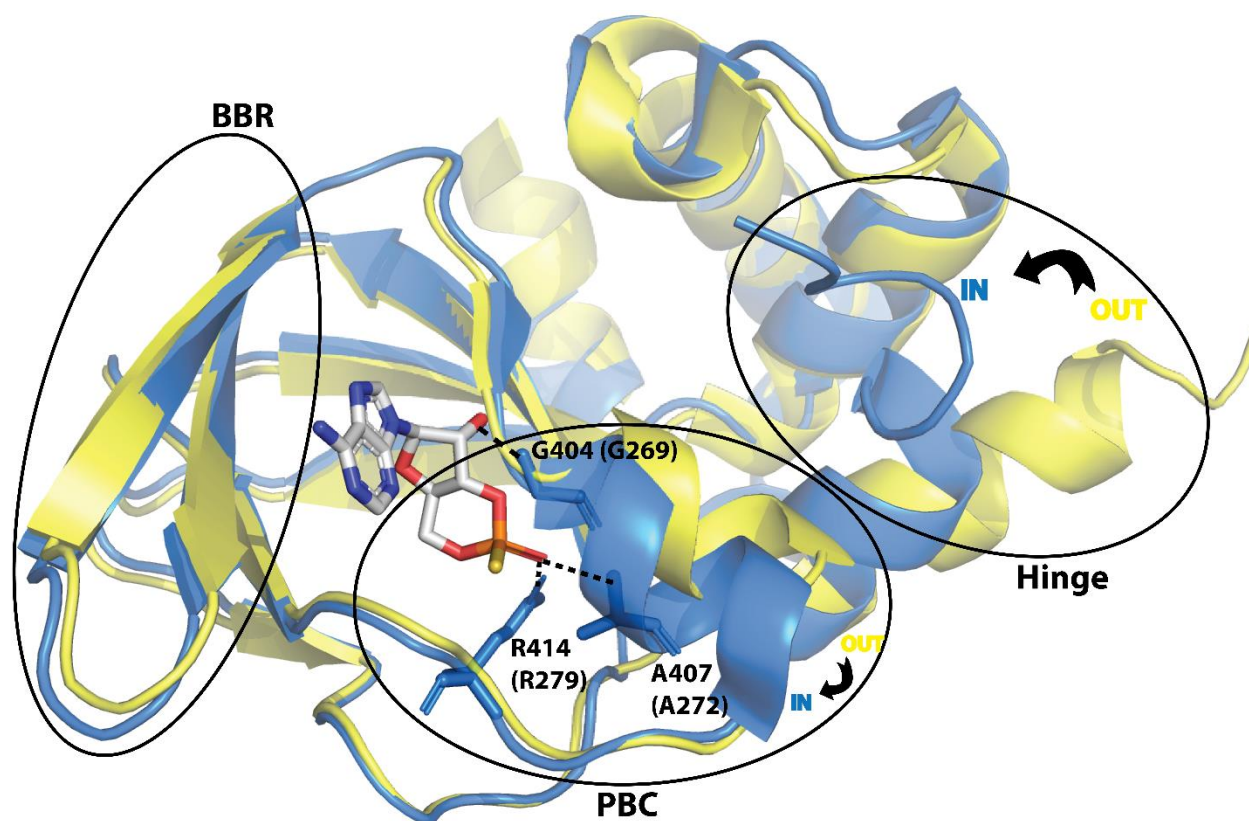
Figure 1.3: Structural differences between the apo and holo states of EPAC. The cAMP-bound EPAC2-CNBD (holo) state (blue, PDB 3CF6) is overlayed on the unbound (apo) state (yellow, PDB 1O7F). The phosphate binding cassette (PBC), the base binding region (BBR) and the hinge region are marked by black circles. The PBC and the hinge exhibit the 'out' conformation in the apo state and adopt an 'in' conformation in the holo state. cAMP, represented as a stick figure forms key interactions between its phosphate group and residues in the PBC whereby the residues equivalent to those residues in the EPAC1 structure are written in brackets. The base of the nucleotide is oriented towards the BBR, facilitating hydrophobic interactions.
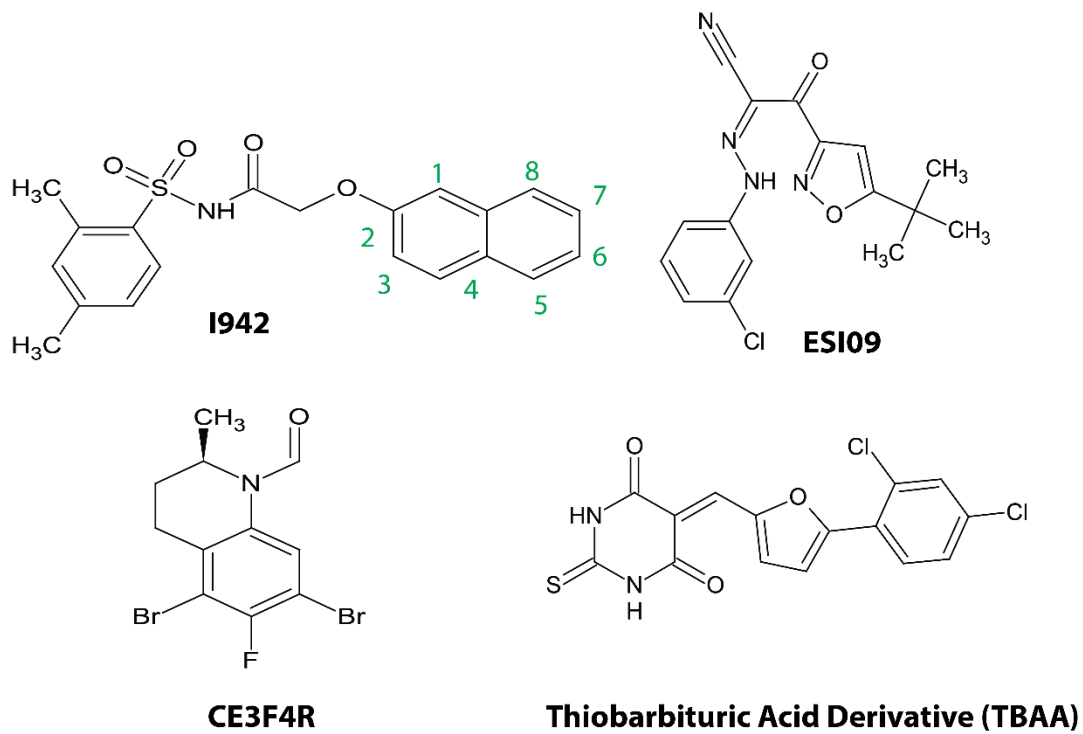
Figure 1.4: Molecular structures of non-cNMP inhibitors of EPAC1
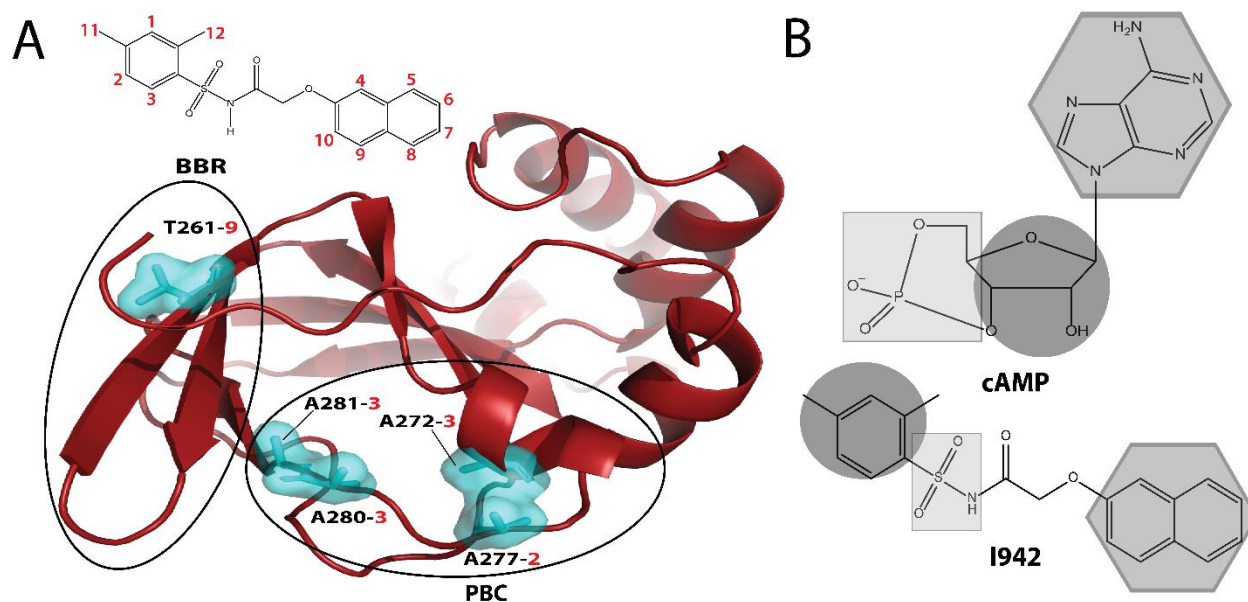
Figure 1.5: I942 binding interface and cAMP mimicry. A) A map of residues, highlighted as cyan surfaces, in the EPAC1-CNBD that showed NOE peaks towards I942 protons from the ^{13}C,^{15}N-filtered NOESY-HSQC readout. NOE peaks to protons in the phenyl group of I942 (2 and 3) originated from residues at the PBC whereas the NOE peak towards the naphthalene proton (9) originated from the BBR residue, T261. B) A scheme for the cAMP mimicry that I942 demonstrates whereby regions of the molecular structures that are highlighted by the same shape are proposed to interact in a similar fashion with EPAC1-CNBD.
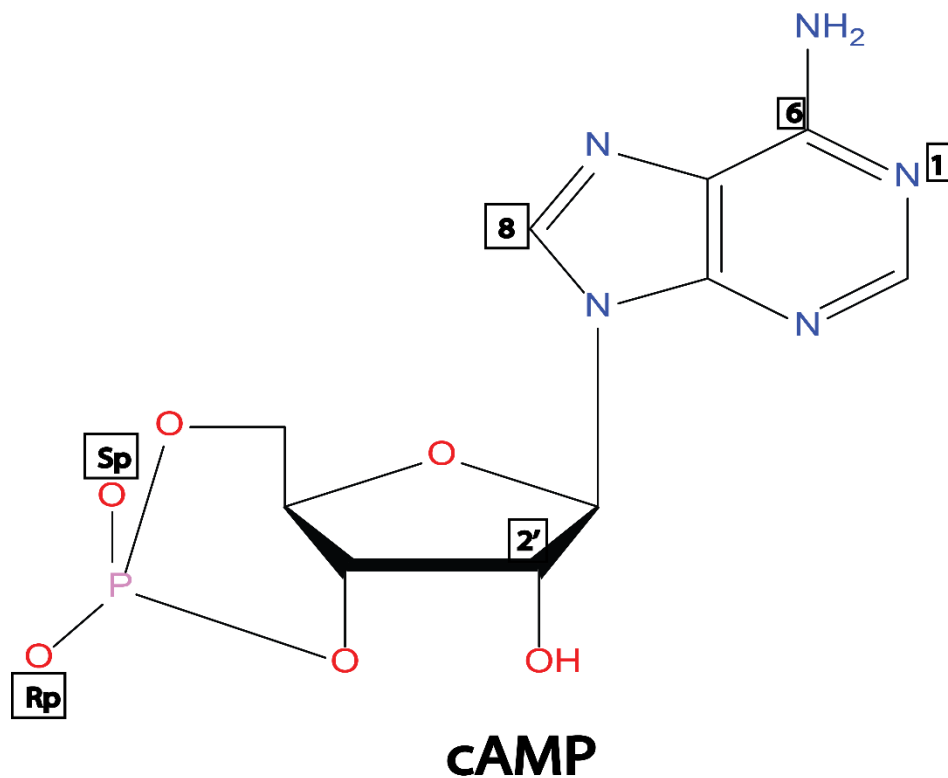
Figure 1.6: Cyclic adenosine monophosphate (cAMP) structure. Numbers in boxes mark the positions where substituents were added to generate different cAMP derivatives. Sp and Rp indicate the location of the sulfur atom in place of the axial and the equatorial oxygens, respectively.

Table 1.1: Preclinical EPAC1 Inhibitors [a]

Name	Type	IC50/µM
I942	Competitive	35 ± 1 [76]
ESI09	Competitive	3.0 ± 1 [77]
CE3F4	Uncompetitive	23 ± 3 [78]
TBAA	Non-competitive	4.0 ± 1 [79]

[a] They bind to the EPAC1 CNBD

Chapter 2

Identification of Core Allosteric Sites through Temperature- and Nucleus-Invariant Chemical Shift Covariance

2.1 Introduction

Allostery plays a central role in cellular signaling, pathological dysregulation, and drug development [1,2,3]. Due to lower evolutionary pressure for conservation at allosteric *vs.* orthosteric sites, targeting the former rather than the latter often leads to enhanced selectivity [4]. In addition, allosteric ligands often exhibit higher affinity than orthosteric binders due to the absence of competition with endogenous substrates [5–7]. It is therefore critical to reliably map residues involved in mediating allostery.

A means to identify allosteric sites in systems for which allostery relies on fast-exchanging conformational equilibria [8] (Figure 2.1A) is the NMR chemical shift covariance analysis (CHESCA) [9–13]. CHESCA maps allosteric networks underlying long-range communication between distal sites within a protein [14–17] (Figure 2.1B). CHESCA utilizes NMR chemical shift variations to single out clusters of residues exhibiting similar responses to a common perturbation library (Figure 2.1C). Such clustering relies on pairwise correlations between residue-specific chemicals shifts (Figure 2.1D-G), whereby residue pairs displaying high correlations typically are assigned to the same allosteric cluster [9].

The original implementation of CHESCA [9,18–21] is effective in generating exhaustive maps of allosteric networks. However, for the purpose of prioritizing subsequent experiments, such as mutations at allosteric sites or allosteric drug design, it is also essential to define the hierarchy of

allostery maps. As a first step in this direction, a means to demonstrate how to narrow down exhaustive CHESCA maps to 'core' allosteric residues suitable for prioritization in follow up studies, is presented. In order to identify such key allosteric sites, three different and complementary CHESCA-based methods are proposed: 1) Temperature-CHESCA (T-CHESCA), which relies on performing CHESCA analyses at different temperatures at which the protein is still folded ($T < T_{melting}$) [22] and identifying residue pairs that remain highly correlated across all temperatures; 2) CLASS-CHESCA, which relies on multiple correlations between separate ^{1}H and ^{15}N amide chemical shifts for each residue pair as opposed to the typical correlations between combined chemical shifts (CCS), which lead to projection compression artifacts (Figure 2.2A) resulting in false positives [10]; 3) The combination of the T- and CLASS-CHESCAs with complete linkage-CHESCA (CL-CHESCA), which implements a form of agglomerative clustering that, unlike single linkage, avoids chaining effects (Figure 2.2B) and related false positives.[10] Together the T-, CLASS- and CL-CHESCAs effectively implement strict filtration criteria that minimize false-positives and selectively reveal core allosteric sites useful to be prioritized in follow up mutation design and/or docking efforts. The proposed CHESCA approaches are validated by applying them to the EPAC1-CNBD.

The available comparative structural, dynamical, and mutational analyses [23,24,25] consistently point to the IL and hinge regions of the EPAC1 CNBD as ideal benchmarks to validate the CHESCA-based approaches proposed to dissect core allosteric elements. The EPAC1 CNBD serves as an excellent model system to test CHESCA methodologies also because cAMP analogs spanning a wide range of EPAC GEF activities are available to use as CHESCA perturbation libraries (Figure 2.1C). Hence, the EPAC1 CNBD model system is utilized to show how T-, CLASS- and CL-CHESCA approaches can be used to build hierarchical maps of allostery.

2.2 Methods

Sample Preparation. The EPAC1$_h$ (149-318) construct was purified according to previously established protocols [9] except for inducing the expression by adding 1 mM of Isopropyl β-D-1-thiogalactopyranoside and concentrating the final protein samples to 0.25 mM.

NMR Measurements and Processing. NMR experiments were acquired using a Bruker Avance 700 MHz spectrometer. Sensitivity and gradient-enhanced ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) spectra were acquired with 256 (t_1) and 2048 (t_2) complex points and spectral widths of 31.82 and 14.06 ppm for the ^{15}N and ^{1}H dimensions, respectively. The number of scans was 8 and the recycle delay was 1 s. The experiments were repeated for each of the five EPAC1-CNBD samples (*i.e.,* apo, cAMP-, Rp-cAMPS-, Sp-cAMP- and 2'-OMe-cAMP-bound) and at each of the five temperature points (290K, 298K, 306K, 310K to 316K). The equilibration time interval between each temperature was 15 minutes. The spectra were processed using NMRPipe [26] where the size of the real spectrum (SI) was 1024 and 512 in the ^{1}H and the ^{15}N dimensions, respectively. A window function (WDW) of sine squared was applied with a sine bell shift (SSB) value of 3 for both dimensions. Forward line prediction (LPfc) was used for the two dimensions with the number of LP coefficients being 64 for the ^{15}N dimension.

CHESCA Analyses -Temperature CHESCA (T-CHESCA). The CHESCA analyses were preformed using the NMRFAM-SPARKY plugin [27]. The HSQC peaks were first referenced to ^{15}N-acetyl glycine and were assigned by comparison starting from previously determined assignments at 306K [9,28,29]. The chemical shifts from the assigned spectra for each EPAC1 state/perturbation (Apo, cAMP, Sp-cAMPS, 2'OMe-cAMP and Rp-cAMPS) at each temperature were used to build temperature-specific correlation matrices through Sparky-CHESCA [12]. The cut-

off values for the ^{15}N and ^{1}H chemical shifts were set to 5 and 10 Hz, respectively and a scaling factor of 0.2 was utilized for the ^{15}N dimension in the computations of combined chemical shifts (CCS). The Pearson correlation coefficient (R) cut-off was set to 0.98 and the CHESCA matrices for all temperatures were generated in the same manner. Only residue pairs for which R values could be computed across all five temperatures were included in the analyses. The embedded "CHESCA-CL" feature in Sparky-CHESCA was used to generate the complete linkage clusters and the correlation coefficient cut-off between residues in the clusters was set to 0.98. The SVD feature in Sparky-CHESCA was used to generate the singular value decomposition (SVD) plot at 306K. In the SVD, Rp-cAMPS was used as the reference state to calculate the differentials of the states with respect to Rp-cAMPS (loading plot). PyMoL was used for mapping the residues identified through CHESCA on the EPAC-CNBD structure.

2.3 Results and Discussion

Rationale for Temperature-CHESCA (T-CHESCA). The central tenet of the T-CHESCA is that core allosteric couplings are preserved at increasing temperatures, provided that no appreciable thermal unfolding occurs. In CHESCA, allosteric couplings between two generic residues *i* and *j* are identified through linear chemical shift correlations [9].

$$\delta_{is} = \delta_{js}\alpha + \beta \qquad (2.1)$$

where δ_{is} and δ_{js} are typically the combined ^{15}N and ^{1}H amide chemical shifts of residues *i* and *j*, respectively, in sample *s* (*i.e.,* either the apo or the cAMP, Rp-cAMPS, Sp-cAMPS or 2'-OMe-cAMP-bound EPAC1 CNBD; Figure 2.1C), and:

$$\alpha = \frac{\delta_{i,Ac} - \delta_{i,In}}{\delta_{j,Ac} - \delta_{j,In}} \tag{2.2}$$

$$\beta = \delta_{i,In} - \alpha\delta_{j,In} \tag{2.3}$$

with $\delta_{i,In}$ ($\delta_{j,In}$) and $\delta_{i,Ac}$ ($\delta_{j,Ac}$) representing the combined ppm values (CCS) of residue i (j) in the pure inactive and active states, respectively.

If the temperature at which the CHESCA analysis is performed changes (Figure 2.3A), the $\delta_{i,In}$ ($\delta_{j,In}$) and $\delta_{i,Ac}$ ($\delta_{j,Ac}$) values are expected to change [30–32], thus possibly changing the slope (α) and intercept (β) of the correlation between δ_{is} and δ_{js}. However, based on equation (2.1), the extent to which δ_{is} and δ_{js} are linearly correlated remains unaffected. Hence, we anticipate that CHESCA correlations for residue pairs (i and j) sensing fast-exchanging concerted two-state transitions are temperature invariant.

Temperature variations may also lead to changes in the relative populations of the two exchanging states (*i.e.,* inactive, and active) and hence results in redistributions of the experimental points defining the correlation between δ_{is} and δ_{js} (Figure 2.3B). The Pearson correlation coefficient, however, is not expected to vary appreciably, provided the conditions of equations (2.1) are fulfilled, *i.e.,* two-state exchange in the fast exchange regime, ensuring that the observed chemical shifts are population-weighted linear averages of the pure state ppm values.

The temperature invariance of pairwise residue *vs.* residue CHESCA correlations is a unique property of tight allosteric couplings. If the coupling between residue i and j is weak, the likelihood that these residues sample more than two states increases, as predicted by ensemble allosteric models [1,33–36]. In this case temperature-induced variations in state-populations and/or

state-specific ppm values are expected to result in losses of CHESCA correlations. For example, a fast-exchanging three state model predicts that equation (2.1) should be modified to [9]:

$$\delta_{is} = \delta_{js}\alpha + \beta + p_s''\,(\delta_{i,Ac} - \delta_{i,In})\gamma \qquad (2.4)$$

where p_s'' is the population of the new (third) state in sample s and:

$$\gamma = \left(\frac{\varepsilon_i}{\delta_{i,Ac}-\delta_{i,In}}\right) - \left(\frac{\varepsilon_j}{\delta_{j,Ac}-\delta_{j,In}}\right) \qquad (2.5)$$

with ε_i (ε_j) is defined as the difference between the combined chemical shifts of the third state and the active state for residue i (j). The third addendum in equation (2.4) dictates the non-linearity of the δ_{is} vs. δ_{js} plot and its temperature dependency, as it includes a population factor and pure-state chemical shifts. Hence, the temperature-invariance of the linearity in CHESCA inter-residue pairwise correlations is a unique signature of residues that report exquisitely on highly concerted conformational transitions, which in turn reflect tight allosteric couplings. On this basis, it is expected that core allosteric sites should exhibit CHESCA correlations that are more resistant to temperature variations than more weakly coupled allosteric loci. This is the rationale of the temperature-CHESCA (T-CHESCA).

Implementation of T-CHESCA. In addition to the previously used temperature, 306 K,[9,28,29] lower (290K and 298K) as well as higher (310K and 316K) temperatures were utilized to implement the T-CHESCA for the EPAC1-CNBD (Figure 2.3). The structural integrity of this domain is preserved even at the higher temperatures, as indicated by the conservation of the ^{1}H chemical shift dispersion (Figure 2.3A). Hence, we proceeded with the computation of the CHESCA correlation matrices for EPAC1-CNBD at each of the five temperature values (Figure 2.3D-F; Figure 2.4) using the same perturbation library of ligands at each temperature (*i.e.,* cAMP,

Rp-cAMPS, Sp-cAMPS or 2'-OMe-cAMP; Figure 2.1C). The comparative analysis of the CHESCA correlation matrices reveals that selected correlations are maintained across the full temperature spectrum, such as those between several residues in the hinge and the adjacent α4 helix, while other correlations are markedly temperature-dependent, such as those between several residues in the a- and b-subdomains (Figure 2.3D-F and Figure 2.4).

Further insight into the temperature-dependence of the CHESCA correlation matrices is provided by the probability distribution of the respective Pearson's correlation coefficients (R) at different temperatures (Figure 2.5). Figure 2.5 shows that most R values are concentrated at the extreme end values (*i.e.,* +/- ~1). In the case of the ^{15}N ppm values both ends are populated at comparable levels (Figure 2.5A), while for the ^{1}H ppm values a net preference for positive R values is observed (Figure 2.5B). As expected, for compounded chemical shifts an intermediate pattern between the ^{15}N and ^{1}H distributions is detected (Figure 2.5C). In all three cases, it is notable that at the highest temperature, lower absolute R values start to be populated more than at the lower temperatures (Figure 2.5), suggesting a heating induced decorrelation possibly arising from the appreciable sampling of additional conformational states within the native ensemble of EPAC1-CNBD. These R value distributions do not seem to be correlated with the magnitude of the underlying chemical shift changes (Figure 2.6).

To determine which residues are involved in pairwise correlations preserved across all temperatures, we computed the mean R (<R>) and the related standard deviation (s) for each residue pair for which assigned chemical shifts were available for all five CHESCA perturbations (Figure 2.1C) at all temperature values. The resulting s *vs.* <R> plot is shown in Figure 2.7. Figure 2.7 shows that residue pairs exhibiting excellent or poor correlations (|<R>| ~ 1 or |<R>| ~ 0) tend

to remain as such throughout the temperature spectrum tested, *i.e.,* they exhibit minimal s values (Figure 2.7A-C). On the contrary, intermediate |<R>| values are subject to the highest degree of temperature-dependent variability (s) (Figure 2.7B). In fact, most of the residue pairs fall within or in the vicinity of the boundaries dictated by the dashed lines in Figure 2.7 defined as: s = |<R>| for |<R>| approaching zero and s = 1-|<R>| for |<R>| approaching unity.

To select the most conserved correlations, which are more likely to reflect tight allosteric couplings, we focused on the residue pairs meeting the conditions: <R> > 0.98 and s < (1-0.98). These correlations are displayed in the zoomed insets of Figure 2.7 (Figure 2.7A,C) and define the ensemble of proposed core allosteric sites determined through the T-CHESCA approach (Table 2.1). To further filter the core allosteric sites identified by the T-CHESCA approach, we complemented it with another proposed CHESCA variation called here the CLASS-CHESCA.

Rationale for the CLASS-CHESCA. Similarly, to the T-CHESCA, the CLASS-CHESCA also singles out tight allosteric couplings based on their invariance with respect to changes in the chemical shifts of the pure active and inactive states. However, unlike the T-CHESCA, in the CLASS-CHESCA the chemical shifts of the pure states are not changed by varying the temperature but by varying the type of nuclei selected for each residue (Figure 2.8A). For example, if ppm values for the ^{15}N and ^{1}H nuclei are available, the CLASS-CHESCA approach requires the computation of four correlation coefficients (R_{NH}, R_{HN}, R_{HH} and R_{NN}; Figure 2.8A) corresponding to the respective inter-residue correlations between the four possible nuclei pairs (*i.e.,* ^{15}N *vs.* ^{1}H, ^{1}H *vs.* ^{15}N, ^{1}H *vs.* ^{1}H and ^{15}N *vs.* ^{15}N; Figure 2.8A).

As shown in Figure 2.8A, in the CLASS-CHESCA method the nitrogen and proton chemical shifts are examined separately as opposed to the original CHESCA implementation, in

which the ^{1}H and the scaled ^{15}N ppm values are added to yield a single combined chemical shift (CCS) for each residue. While the CCS approach offers a simple means to build residue-residue correlations, it suffers from ambiguities arising from projection compression artifacts (Figure 2.2A) [10]. Due to projection compression, distinct ^{1}H-^{15}N HSQC cross-peaks may result in similar CCS values (Figure 2.2A) thus possibly leading to spurious pair-wise correlations and false-positives in the detection of allosteric sites. The CLASS-CHESCA also circumvents this limitation without requiring the acquisition of data at different temperatures.

In the CLASS-CHESCA, the strength of the overall residue-pair correlations is assessed based on how many of the nuclei-specific correlations (R_{NH}-R_{NN}) give rise to Pearson correlation coefficients above a given threshold value (R_{th}; Figure 2.8B). For the weakest correlations, defined as class 0 correlations, no R value exceeds R_{th}, while for the strongest correlations, denoted as class 4 correlations, all four R values exceed R_{th} (Figure 2.8B). For the intermediate correlations (classes 1-3), a subset of R values exceeds R_{th} (Figure 2.8B). Representative examples of residue-residue correlations for each class are shown in Figure 2.8C and the full CLASS-CHESCA correlation matrix is displayed in Figure 2.8D.

Residue pairs that belong to classes 3 and 4 define the best CLASS-CHESCA correlations and point to tight allosteric couplings. Based on the CLASS-CHESCA matrix (Figure 2.8D), class 3 and 4 residues appear in the IL, α2, α3-α4, α4 and α6 regions (Table 2.1). It is also notable that correlations along the diagonal (*i.e.,* correlations between the chemical shifts of the same residue) are often not classified as class 4 (Figure 2.8D). This is because CHESCA perturbation-induced changes in ^{1}H and ^{15}N ppm values for a given residue are not necessarily linearly correlated. The

information content of diagonal CLASS-CHESCA self-correlations is an added benefit of this type of CHESCA implementation.

CL-CHESCA. As a complement to the T-CHESCA and CLASS-CHESCA as well as a term of reference, the classical CHESCA was implemented based on CCS at a single temperature (306K) but using a stringent form of hierarchical clustering known as complete-linkage (CL) clustering. Unlike single-linkage (SL) clustering, CL clustering requires that all residue pairs within a given cluster exhibit correlation coefficients above a selected threshold (Figure 2.2B). The CL-CHESCA resulted in five clusters (Figure 2.9 and Figure 2.10) with correlations coefficients above 0.98. The CL-CHESCA clusters 1-3 and 5 (Figure 2.9 and Figure 2.10) include residues preferentially aligned along the second singular value decomposition (SVD) principal component (PC2; Figure 2.11), which reflects, primarily, allosteric rather than binding contributions. On the contrary, cluster 4 in the CL-CHESCA aligns better with PC1 as opposed to PC2, which predominantly represents binding rather than allosteric contributions. In fact, cluster 4 residues are located at the BBR and just near the PBC region and both BBR and PBC are known binding sites for cyclic nucleotides. Cluster 4 was therefore, excluded from the CL-CHESCA ensemble in order to mainly account for allosteric contributors (Table 2.1).

Comparison of T-, CLASS- and CL-CHESCAs. The T-, CLASS- and CL-CHESCAs approaches select distinct but overlapping subsets of the commonly used single-linkage (SL) CCS-CHESCA (Figure 2.12A). Comparison of the residues identified by the T-, CLASS- and CL-CHESCAs (Figure 2.12A-E; Table 2.1) demonstrates that sites shared by all three ensembles are in regions of the EPAC1-CNBD known to be critical for the allosteric control of EPAC [23,24,37–41], *i.e.,* the IL, the α3- α4 loop as well as the α6 hinge helix and the adjacent α4 helix (Figure 2.12F).

This observation suggests that the CHESCA-based approaches proposed here are effective in identifying core allosteric sites to be targeted by EPAC modulators. For example, the binding mode of CE3F4R, which is an unconventional uncompetitive allosteric inhibitor of EPAC1 [42–44], was found to be in the interface between the α and β-subdomains of the CNBD with the most significant chemical shift changes upon CE3F4R binding observed in the α4 and α6 helices [43]. Both helices are, therefore, pivotal components in the allosteric regulation of EPAC1 activity and serve as hotspots for EPAC1-targeted drug discovery.

Out of the three CHESCA approaches investigated, the CL-CHESCA method emerges as the least selective approach. The CL-CHESCA yields the largest allosteric ensemble with several residues not captured by the other two sets (Table 2.1). The percentage of residues exclusively found in the CL-CHESCA ensemble is the highest compared to the other two approaches (Table 2.1), pointing to low selectivity in identifying core allosteric sites. Improved selectivity is obtained by altering the chemical shifts of the two pure states (*i.e.,* active, and inactive conformations) either by changing the temperature within the folded range (T-CHESCA) and/or by mixing and matching the chemical shifts of spin-active nuclei for each residue involved in pairwise correlations (CLASS-CHESCA).

The CLASS-CHESCA appears as the most selective method with no residue exclusively identified by it (Table 2.1). Hence, the CL- and CLASS-CHESCA lie at opposite extremes of the selectivity spectrum in the identification of allosteric sites. While the former may suffer from false positives, the latter may introduce false negatives. The T-CHESCA emerges therefore, as a more balanced compromise to concurrently minimize both false positives and negatives (Table 2.1).

However, unlike the other two types of CHESCA, T-CHESCA requires the acquisition of NMR data and assignments at multiple temperatures.

2.4 References

(1)   Gunasekaran, K.; Ma, B.; Nussinov, R. Is Allostery an Intrinsic Property of All Dynamic Proteins? *Proteins: Structure, Function and Genetics* **2004**, *57* (3), 433–443.

(2)   Nussinov, R.; Tsai, C.-J.; Liu, J. Principles of Allosteric Interactions in Cell Signaling. *Journal of the American Chemical Society* **2014**, *136* (51), 17692–17701.

(3)   Nussinov, R.; Tsai, C. J. Allostery in Disease and in Drug Discovery. *Cell* **2013**, *153* (2), 293–305.

(4)   Ortiz Zacarías, N. V; Lenselink, E. B.; IJzerman, A. P.; Handel, T. M.; Heitman, L. H. Intracellular Receptor Modulation: Novel Approach to Target GPCRs. *Trends in pharmacological sciences* **2018**, *39* (6), 547–559.

(5)   Foster, D. J.; Conn, P. J. Allosteric Modulation of GPCRs: New Insights and Potential Utility for Treatment of Schizophrenia and Other CNS Disorders. *Neuron* **2017**, *94* (3), 431–446.

(6)   Christopoulos, A. Advances in G Protein-Coupled Receptor Allostery: From Function to Structure. *Molecular Pharmacology* **2014**, *86* (5), 463–478.

(7)   Kenakin, T. P. Biased Signalling and Allosteric Machines: New Vistas and Challenges for Drug Discovery. *British Journal of Pharmacology* **2012**, *165* (6), 1659–1669.

(8)   Xu, J.; Sarma, A. V. S.; Wei, Y.; Beamer, L. J.; van Doren, S. R. Multiple Ligand-Bound States of a Phosphohexomutase Revealed by Principal Component Analysis of NMR Peak Shifts. *Scientific Reports 2017 7:1* **2017**, *7* (1), 1–11.

(9)   Selvaratnam, R.; Chowdhury, S.; VanSchouwena, B.; Melacini, G. Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108* (15), 6133–6138.

(10)   Boulton, S.; Akimoto, M.; Selvaratnam, R.; Bashiri, A.; Melacini, G. A Tool Set to Map Allosteric Networks through the NMR Chemical Shift Covariance Analysis. *Scientific Reports* **2015**, *4* (1), 7306.

(11)   Boulton, S.; Melacini, G. Advances in NMR Methods To Map Allosteric Sites: From Models to Translation. *Chemical Reviews* **2016**, *116* (11), 6267–6304.

(12)   Shao, H.; Boulton, S.; Olivieri, C.; Mohamed, H.; Akimoto, M.; Subrahmanian, M. V.; Veglia, G.; Markley, J. L.; Melacini, G.; Lee, W. CHESPA/CHESCA-SPARKY: Automated NMR Data Analysis Plugins for SPARKY to Map Protein Allostery. *Bioinformatics* **2020**, 1–2.

(13)   Verkhivker, G. M. Making the Invisible Visible: Toward Structural Characterization of Allosteric States, Interaction Networks, and Allosteric Regulatory Mechanisms in Protein Kinases. *Current Opinion in Structural Biology* **2021**, *71*, 71–78.

(14)   Gagné, D.; Narayanan, C.; Doucet, N. Network of Long-Range Concerted Chemical Shift Displacements upon Ligand Binding to Human Angiogenin. *Protein Science* **2015**, *24* (4), 525–533.

(15) Chen, H.; Marsiglia, W. M.; Cho, M. K.; Huang, Z.; Deng, J.; Blais, S. P.; Gai, W.; Bhattacharya, S.; Neubert, T. A.; Traaseth, N. J.; Mohammadi, M. Elucidation of a Four-Site Allosteric Network in Fibroblast Growth Factor Receptor Tyrosine Kinases. *eLife* **2017**, *6*.

(16) Hajredini, F.; Ghose, R. An ATPase with a Twist: A Unique Mechanism Underlies the Activity of the Bacterial Tyrosine Kinase, Wzc. *Science Advances* **2021**, *7* (39), 5836–5858.

(17) Weber, D. K.; Reddy, U. V.; Wang, S.; Larsen, E. K.; Gopinath, T.; Gustavsson, M. B.; Cornea, R. L.; Thomas, D. D.; de Simone, A.; Veglia, G. Structural Basis for Allosteric Control of the SERCA-Phospholamban Membrane Complex by Ca2+ and Phosphorylation. *eLife* **2021**, *10*.

(18) Selvaratnam, R.; Akimoto, M.; VanSchouwen, B.; Melacini, G. CAMP-Dependent Allostery and Dynamics in Epac: An NMR View. *Biochemical Society Transactions* **2012**, *40* (1), 219–223.

(19) VanSchouwen, B.; Ahmed, R.; Milojevic, J.; Melacini, G. Functional Dynamics in Cyclic Nucleotide Signaling and Amyloid Inhibition. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2017**, *1865* (11), 1529–1543.

(20) Boulton, S.; Selvaratnam, R.; Ahmed, R.; Melacini, G. Implementation of the NMR CHEmical Shift Covariance Analysis (CHESCA): A Chemical Biologist's Approach to Allostery; 2018; pp 391–405.

(21)   Byun, J. A.; Melacini, G. NMR Methods to Dissect the Molecular Mechanisms of Disease-Related Mutations (DRMs): Understanding How DRMs Remodel Functional Free Energy Landscapes. *Methods* **2018**, *148*, 19–27.

(22)   Rees, D. C.; Robertson, A. D. Some Thermodynamic Implications for the Thermostability of Proteins. *Protein Science* **2001**, *10* (6), 1187–1194.

(23)   Tsalkova, T.; Blumenthal, D. K.; Mei, F. C.; White, M. A.; Cheng, X. Mechanism of Epac Activation: Structural and Functional Analyses of Epac2 Hinge Mutants with Constitutive and Reduced Activities. *The Journal of biological chemistry* **2009**, *284* (35), 23644–23651.

(24)   White, M. A.; Li, S.; Tsalkova, T.; Mei, F. C.; Liu, T.; Woods, V. L.; Cheng, X. Structural Analyses of a Constitutively Active Mutant of Exchange Protein Directly Activated by CAMP. *PloS one* **2012**, *7* (11), e49932.

(25)   Rehmann, H.; Rueppel, A.; Bos, J. L.; Wittinghofer, A. Communication between the Regulatory and the Catalytic Region of the CAMP-Responsive Guanine Nucleotide Exchange Factor Epac. *Journal of Biological Chemistry* **2003**, *278* (26), 23508–23514.

(26)   Delaglio, F.; Grzesiek, S.; Vuister, GeertenW.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *Journal of Biomolecular NMR* **1995**, *6* (3).

(27)   Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* **2015**, *31* (8), 1325–1327.

(28)   Mazhab-Jafari, M. T.; Das, R.; Fotheringham, S. A.; SilDas, S.; Chowdhury, S.; Melacini, G. 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 American Chemical Society* **2007**, *129* (46), 14482–14492.

(29)   Das, R.; Mazhab-Jafari, M. T.; Chowdhury, S.; SilDas, S.; Selvaratnam, R.; Melacini, G. Entropy-Driven CAMP-Dependent Allosteric Control of Inhibitory Interactions in Exchange Proteins Directly Activated by CAMP. *Journal of Biological Chemistry* **2008**, *283* (28), 19691–19703.

(30)   Trainor, K.; Palumbo, J. A.; MacKenzie, D. W. S.; Meiering, E. M. Temperature Dependence of NMR Chemical Shifts: Tracking and Statistical Analysis. *Protein Science* **2020**, *29* (1), 306–314.

(31)   Baxter, N. J.; Williamson, M. P. Temperature Dependence of 1H Chemical Shifts in Proteins. *Journal of Biomolecular NMR* **1997**, *9* (4), 359–369.

(32)   Tomlinson, J. H.; Williamson, M. P. Amide Temperature Coefficients in the Protein G B1 Domain. *Journal of Biomolecular NMR* **2012**, *52* (1), 57–64.

(33)   Kern, D.; Zuiderweg, E. R. The Role of Dynamics in Allosteric Regulation. *Current Opinion in Structural Biology* **2003**, *13* (6), 748–757.

(34)   Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508* (7496), 331–339.

(35)   Tsai, C.-J.; Nussinov, R. A Unified View of "How Allostery Works." *PLoS Computational Biology* **2014**, *10* (2), e1003394.

(36)   Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annual Review of Biophysics* **2012**, *41* (1), 585–609.

(37)   Rehmann, H.; Das, J.; Knipscheer, P.; Wittinghofer, A.; Bos, J. L. Structure of the Cyclic-AMP-Responsive Exchange Factor Epac2 in Its Auto-Inhibited State. *Nature* **2006**, *439* (7076), 625–628.

(38)   Rehmann, H.; Arias-Palomo, E.; Hadders, M. A.; Schwede, F.; Llorca, O.; Bos, J. L. Structure of Epac2 in Complex with a Cyclic AMP Analogue and RAP1B. *Nature* **2008**, *455* (7209), 124–127.

(39)   Rehmann, H.; Prakash, B.; Wolf, E.; Rueppel, A.; de Rooij, J.; Bos, J. L.; Wittinghofer, A. Structure and Regulation of the CAMP-Binding Domains of Epac2. *Nature Structural Biology* **2003**, *10* (1), 26–32.

(40)   Selvaratnam, R.; Mazhab-Jafari, M. T.; Das, R.; Melacini, G. The Auto-Inhibitory Role of the EPAC Hinge Helix as Mapped by NMR. *PLoS ONE* **2012**, *7* (11), e48707.

(41)   Li, S.; Tsalkova, T.; White, M. A.; Mei, F. C.; Liu, T.; Wang, D.; Woods, V. L.; Cheng, X. Mechanism of Intracellular CAMP Sensor Epac2 Activation. *Journal of Biological Chemistry* **2011**, *286* (20), 17889–17897.

(42)   Courilleau, D.; Bisserier, M.; Jullian, J.-C.; Lucas, A.; Bouyssou, P.; Fischmeister, R.; Blondeau, J.-P.; Lezoualc'h, F. Identification of a Tetrahydroquinoline Analog as a Pharmacological Inhibitor of the CAMP-Binding Protein Epac. *Journal of Biological Chemistry* **2012**, *287* (53), 44192–44202.

(43)   Boulton, S.; Selvaratnam, R.; Blondeau, J. P.; Lezoualc'H, F.; Melacini, G. Mechanism of Selective Enzyme Inhibition through Uncompetitive Regulation of an Allosteric Agonist. *Journal of the American Chemical Society* **2018**, *140* (30), 9624–9637.

(44)   Byun, J. A.; VanSchouwen, B.; Akimoto, M.; Melacini, G. Allosteric Inhibition Explained through Conformational Ensembles Sampling Distinct "Mixed" States. *Computational and Structural Biotechnology Journal* **2020**, *18*, 3803–3818.
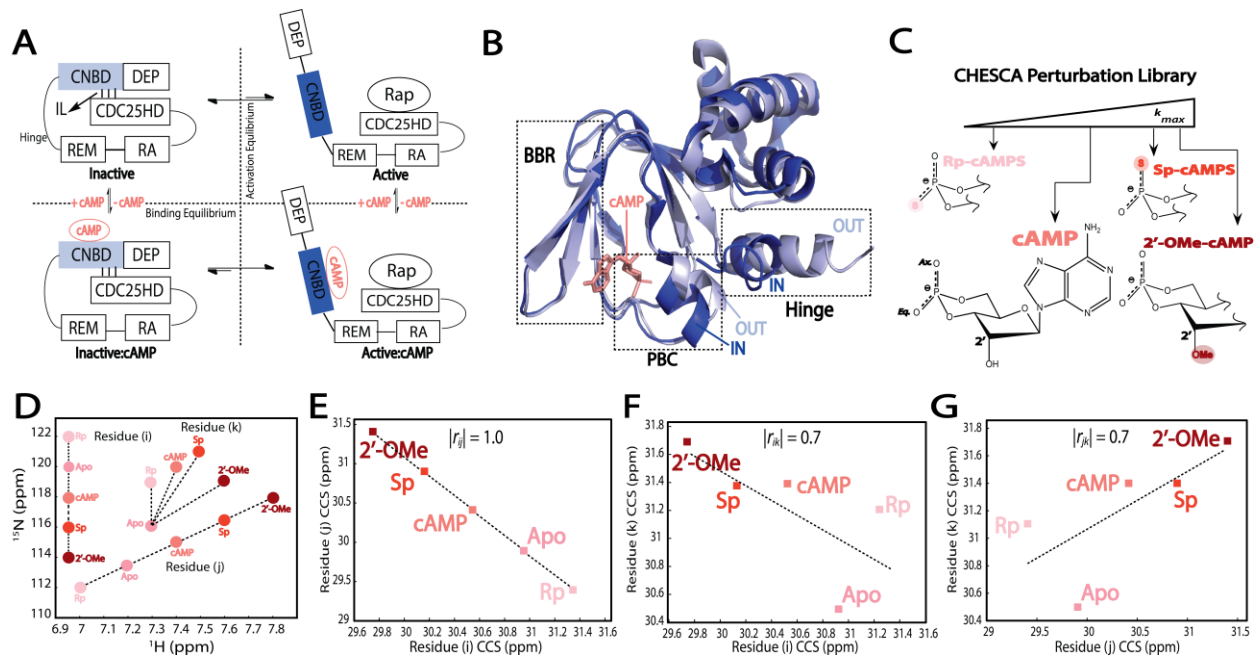
Figure 2.1: Allosteric regulation of EPAC1, CHESCA library, and pairwise residue correlations. A) Coupled activation and binding equilibria underlying the allosteric regulation of EPAC1. IL denotes the ionic latch region of EPAC1-CNBD, which is displaced upon activation by cAMP. B) Overlay of the EPAC-CNBD holo, active state (dark blue, PDB 3CF6) and the apo, inactive state (light blue, PDB 1O7F). Dashed boxes indicate the phosphate binding cassette (PBC), the base binding region (BBR) and the hinge region. The PBC and the hinge are in the 'out' conformation in the apo, inactive state whereas in the holo, active state they are in the 'in' conformation. The endogenous ligand, cAMP, is shown in stick representation docked between the BBR and the PBC. C) Molecular structures of cAMP and cAMP analogues. cAMP analogues include covalent modifications marked by circles. Sp-cAMPS and Rp-cAMPS include a sulfur atom in place of the axial and equatorial oxygens, respectively, while 2'-OMe-cAMP exhibits a methoxy group in place of the 2'-hydroxyl group. The extent of relative activation for each cyclic nucleotide is qualitatively described by the $k_{max,}$ scale above the structures. Shades of red also reflect the extent

of EPAC1 activation. D) Scheme illustrating three representative residues denoted as i, j and k subject to linear *vs.* non-linear changes in chemical shift changes across different perturbations. E) ^{1}H and ^{15}N combined chemical shift correlation plots between residues undergoing a concerted chemical shift response to the perturbations in (C) result in a high degree of correlation. F,G) lower degrees of correlation are observed between residues with chemical shift changes that deviate from linearity due to deviations from two-state fast-exchange model.
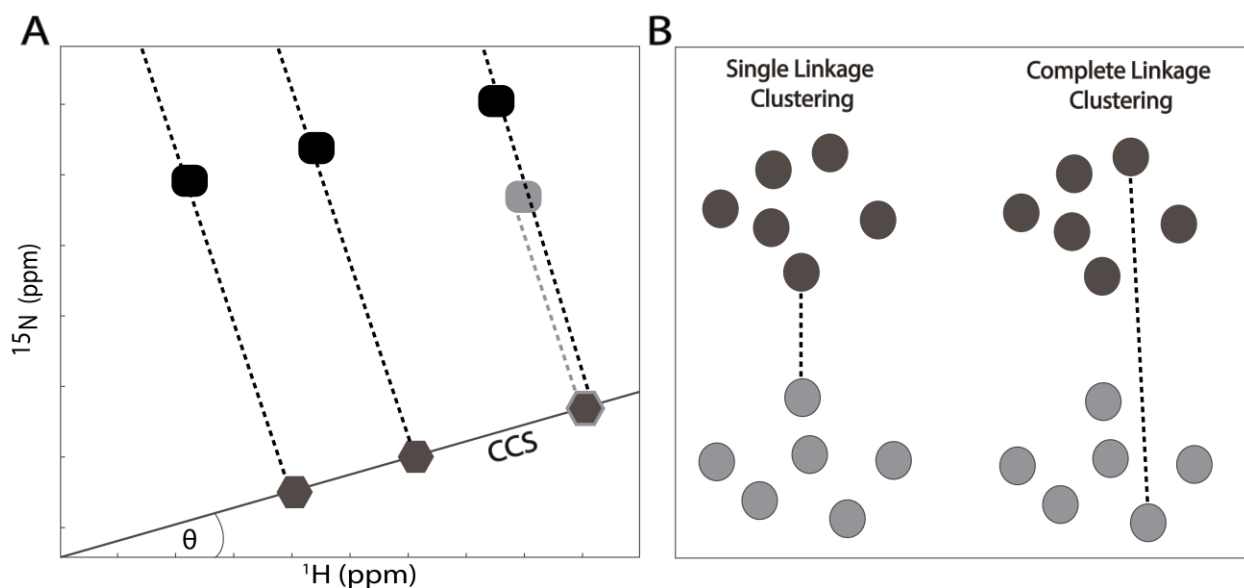
Figure 2.2: Projection compression and CL *vs.* SL clustering. A) Illustration, adapted from Boulton et al. [10], of projection compression occurring when combined ^{1}H and ^{15}N chemical shifts (CCS) are used. Using CCS is equivalent to projecting the 2D HSQC cross-peaks onto a single line with slope ('θ') depending on the relative weights in the linear combination of ^{1}H and ^{15}N chemical shifts utilized for the CCS computation. Two different 2D HSQC cross-peaks may result in the same CCS value. B) Scheme illustrating the differences between the single versus complete-linkage clustering. The black and grey spheres refer to two different sub-clusters of residues and the dotted line shows how the clusters are connected by each method of clustering. In single-linkage, it is sufficient that a single residue pair across the two sub-clusters meets the clustering criterium. In complete-linkage, all residue pairs across the two sub-clusters must meet the clustering criterium.
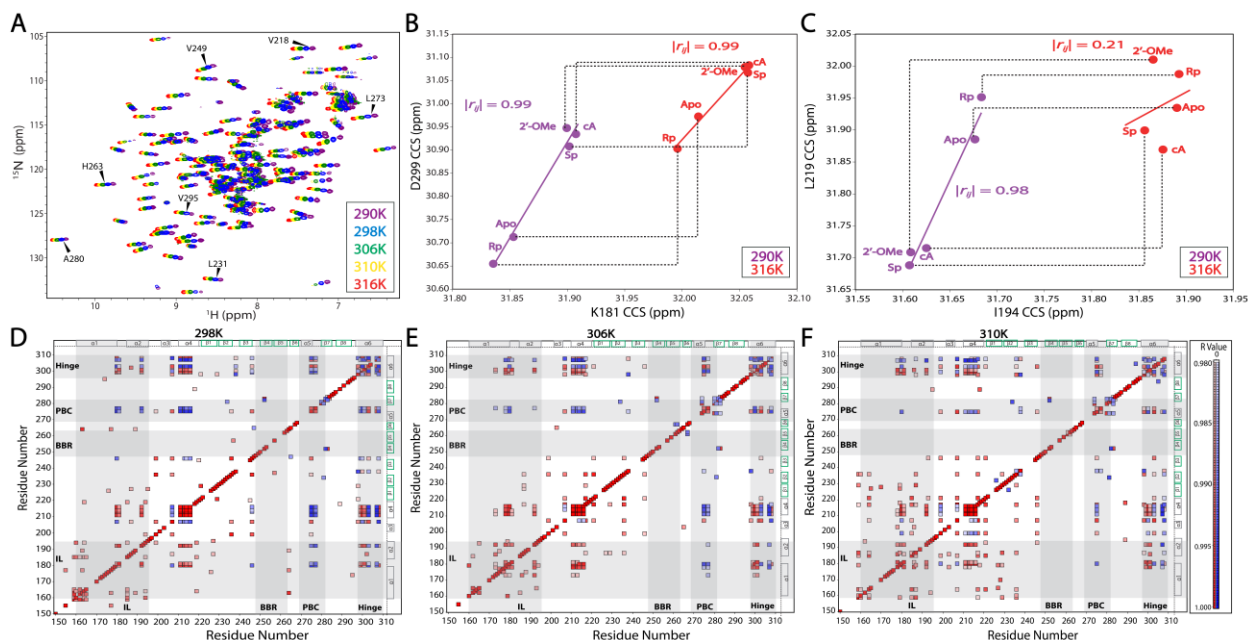
Figure 2.3: EPAC1 HSQC spectra and CHESCA matrices across 298-310K. A) Overlay of unreferenced ^{1}H-^{15}N HSQC spectra acquired for the cAMP-bound (holo) EPAC1-CNBD at different temperatures. B) Examples of inter-residue pairwise correlations for a residue pair with high Pearson correlation coefficient at low and high temperatures. C) Example of a residue pair with high correlation coefficient at low temperature but significantly lower correlation coefficient at higher temperature. Labeled circles represent the different EPAC states at 290K (purple) and 316K (red). The black dotted lines mark the difference in HSQC positions between the same state at the two temperatures. D-F) Chemical shift correlation (CHESCA) matrices for the EPAC1-CNBD at D) 298K E) 306K and F) 310K. The CHESCA R value legend in (F) indicates the color code for absolute correlation coefficients above or equal to 0.98 (red: positive and blue: negative). The secondary structure of the apo EPAC1-CNBD is mapped on the matrix, whereby grey boxes represent α helices and green boxes correspond to β sheets. Regions highlighted in grey denote the phosphate binding cassette (PBC), ionic latch (IL), base binding region (BBR) and the hinge helix.
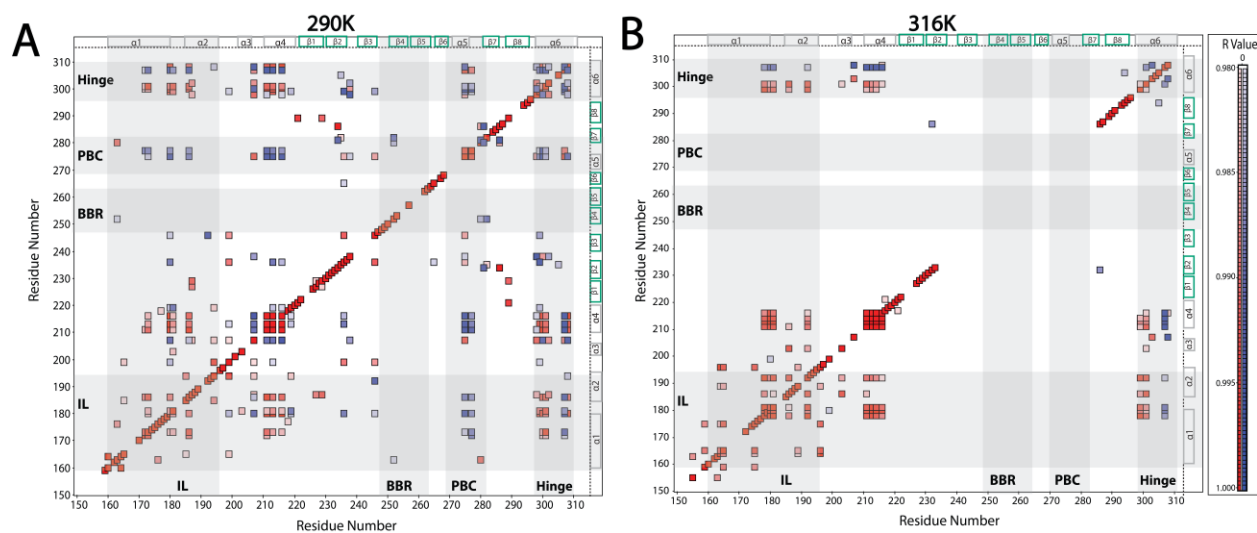
Figure 2.4: Chemical shift correlation (CHESCA) matrices for EPAC1-CNBD at A) 290K and B) 316K. The color code, secondary structure and key regions are marked as in Fig. 2.3D-F.
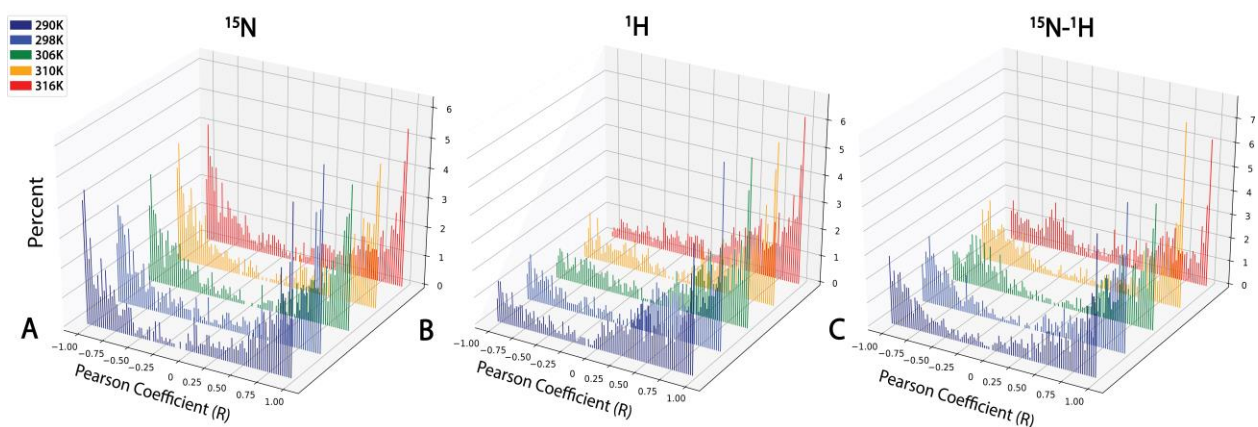
Figure 2.5: Three-dimensional bar plots showing the percentage distribution of residue pair Pearson correlation coefficients (R) at different temperatures ranging from 290K (dark blue) to 316K (red) for A) ^{15}N-only based CHESCA, B) ^{1}H-only based CHESCA and C) combined chemical shift (^{15}N-^{1}H)-based CHESCA.
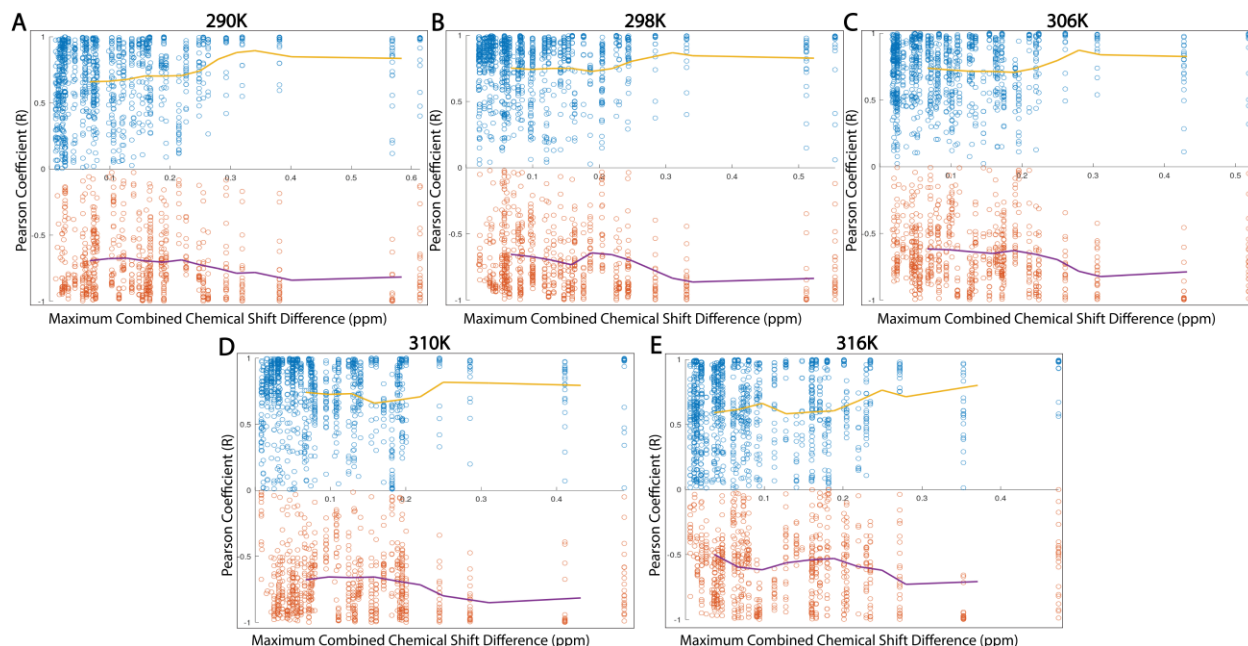
Figure 2.6: Maximum combined chemical shift difference versus Pearson coefficient (R) of EPAC1 residue pairs at different temperatures: A) 290, B) 298K, C) 306K, D) 310K and E) 316K. For every residue pair, the maximum combined chemical shift (CCS) difference corresponds to the difference between the largest and the smallest combined chemical shift value for the first residue of the residue pair. That value is plotted against the Pearson coefficient of the pair of residues. The yellow and purple lines represent the moving averages (+/- three residues) for the positive and negative R values, respectively.
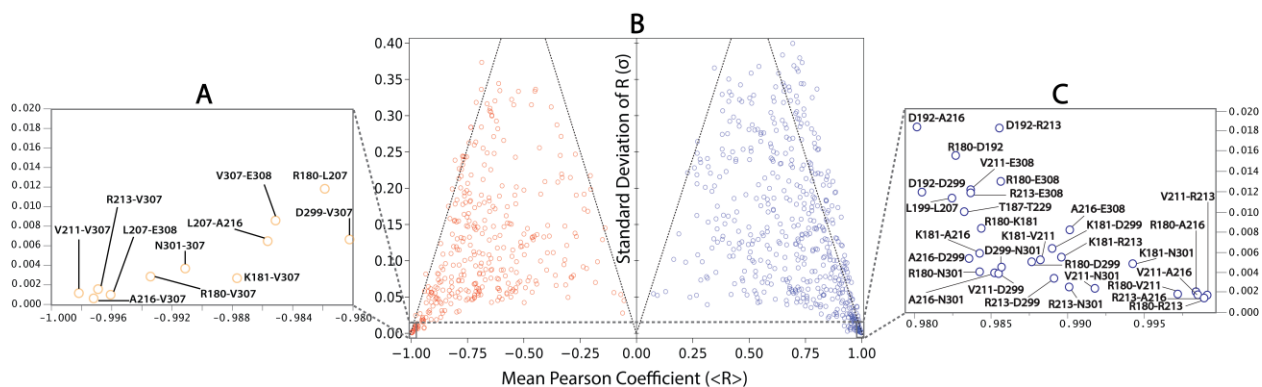
Figure 2.7: Mean Pearson coefficient against standard deviation across temperature. The average R value (<R>) for residue pair correlations was calculated across the five temperatures and plotted against the standard deviation of R (σ) in panel (B). Each circle represents a residue pair: orange (blue) circles correspond to residue pairs with negative (positive) mean R values. The black dashed lines (*i.e.,* σ = <R> if |<R>| < 0.4 and σ = 1 - <R> if |<R>| > 0.6) capture the general approximate trend: residue pairs with high <R> values consistently exhibit low standard deviations, as illustrated in the zoomed in panels (A, C), while residue pairs with lower <R> values tend to result in higher and more variable σ values.

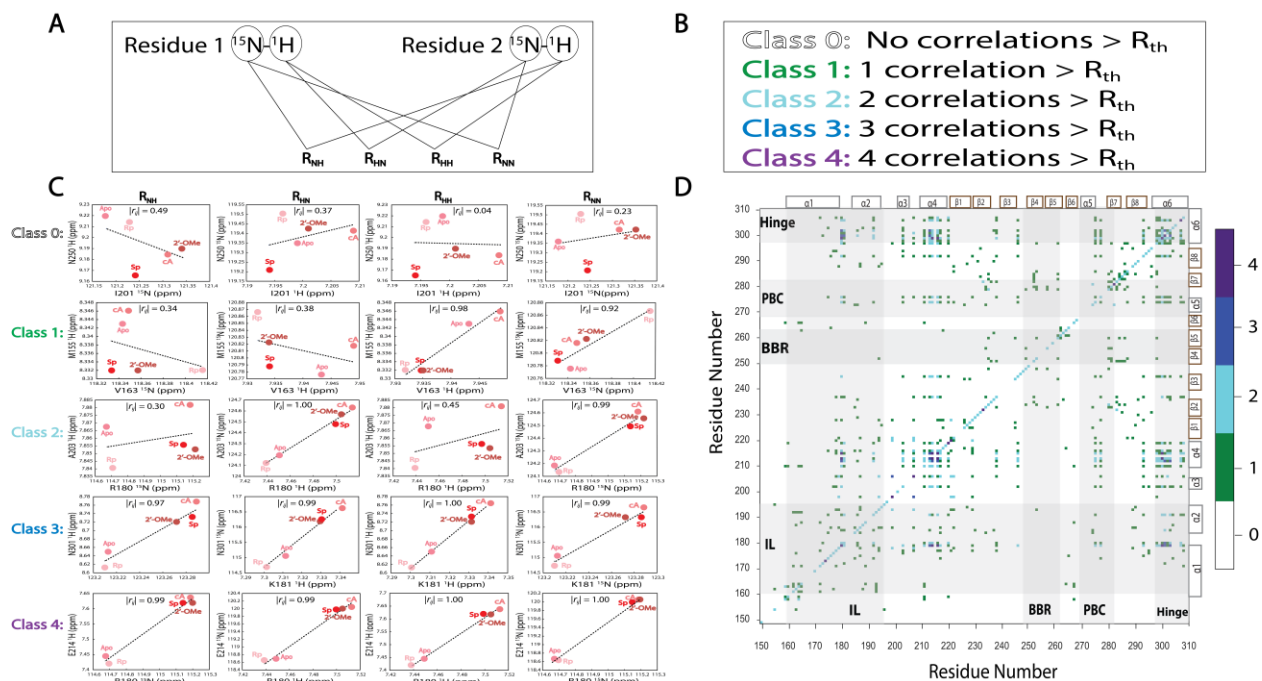**Figure 2.8:** Schematic diagram explaining the rationale of the CLASS-CHESCA. A) For each residue pairs four Person correlation coefficients (R_{NH}-R_{NN}) are calculated, where R_{NH} corresponds to the correlation coefficient between the nitrogen chemical shift of the first residue in the pair and the proton chemical shift of the second residue, R_{HN} is for the proton chemical shift of the first residue and the nitrogen chemical shift of the second residue, R_{HH} is for the proton chemical shift of the first residue and the proton chemical shift of the second residue and R_{NN} is for the nitrogen chemical shift of the first residue and the nitrogen chemical shift of the second residue. B) Residue pairs are then divided into five different classes (0 - 4) depending on how many of the correlation coefficients (R_{NH}-R_{NN}) are greater than a set threshold (R_{th}), *e.g.,* 0.98. C) Examples of residue pairs that fall under different classes. For each class, one residue pair is provided, and the four different correlations (corresponding to R_{NH}-R_{NN}) are shown in different columns. The perturbation states are indicated by color coded circles like Fig. 2.1. D) CLASS-CHESCA matrix where the different colors relate to the type of class that the residue pair correlation belongs to, as

shown by the legend on the right of the panel. The secondary structure of the EPAC1-CNBD is mapped on the matrix whereby grey boxes represent α helices and brown boxes correspond to β sheets. Regions highlighted in grey denote the phosphate binding cassette (PBC), Ionic latch (IL), base binding region (BBR) and the hinge.
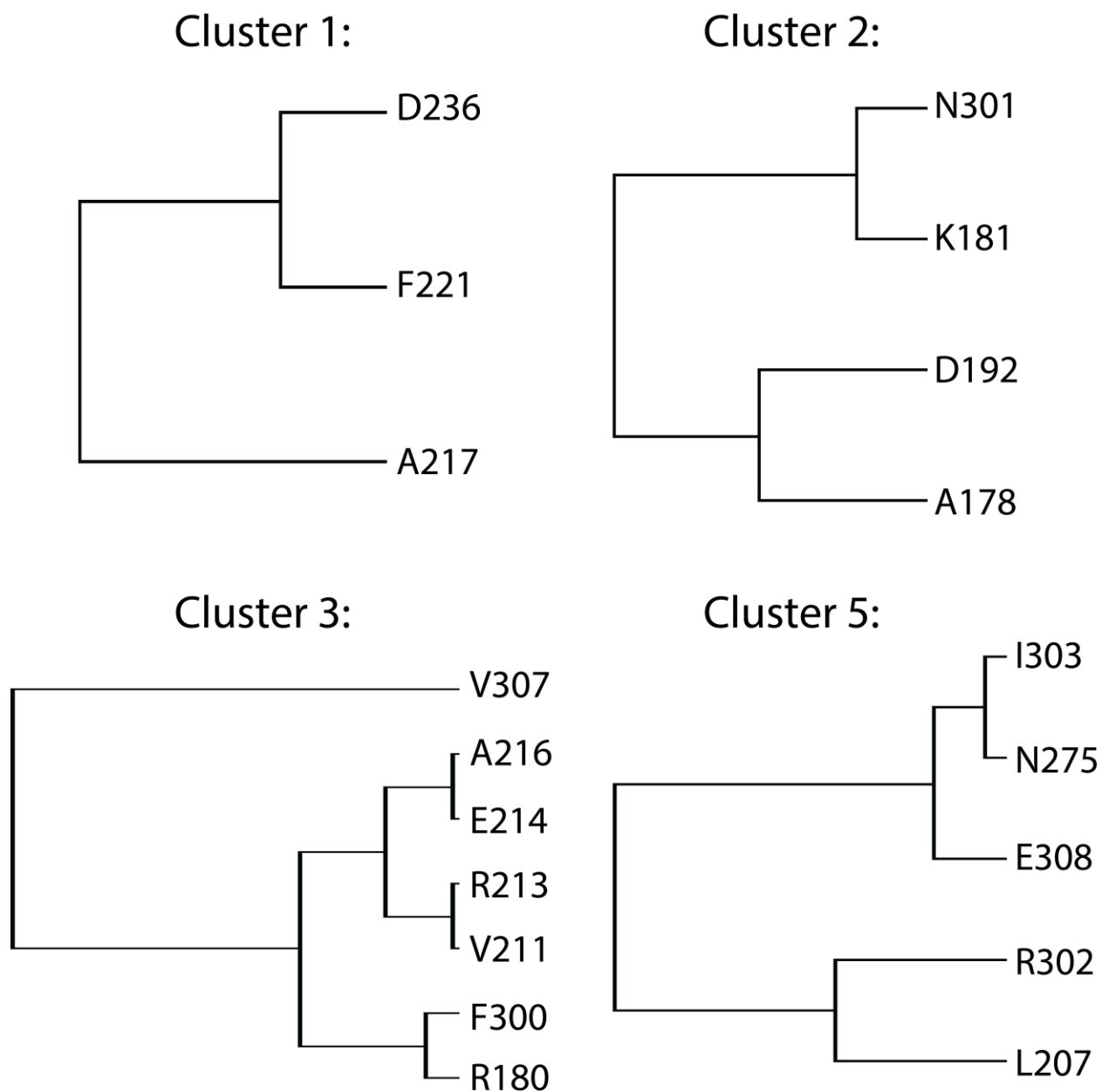
Figure 2.9: Perturbation-based agglomerative clustering of EPAC1-CNBD residues obtained from complete-linkage clustering performed at 306K.
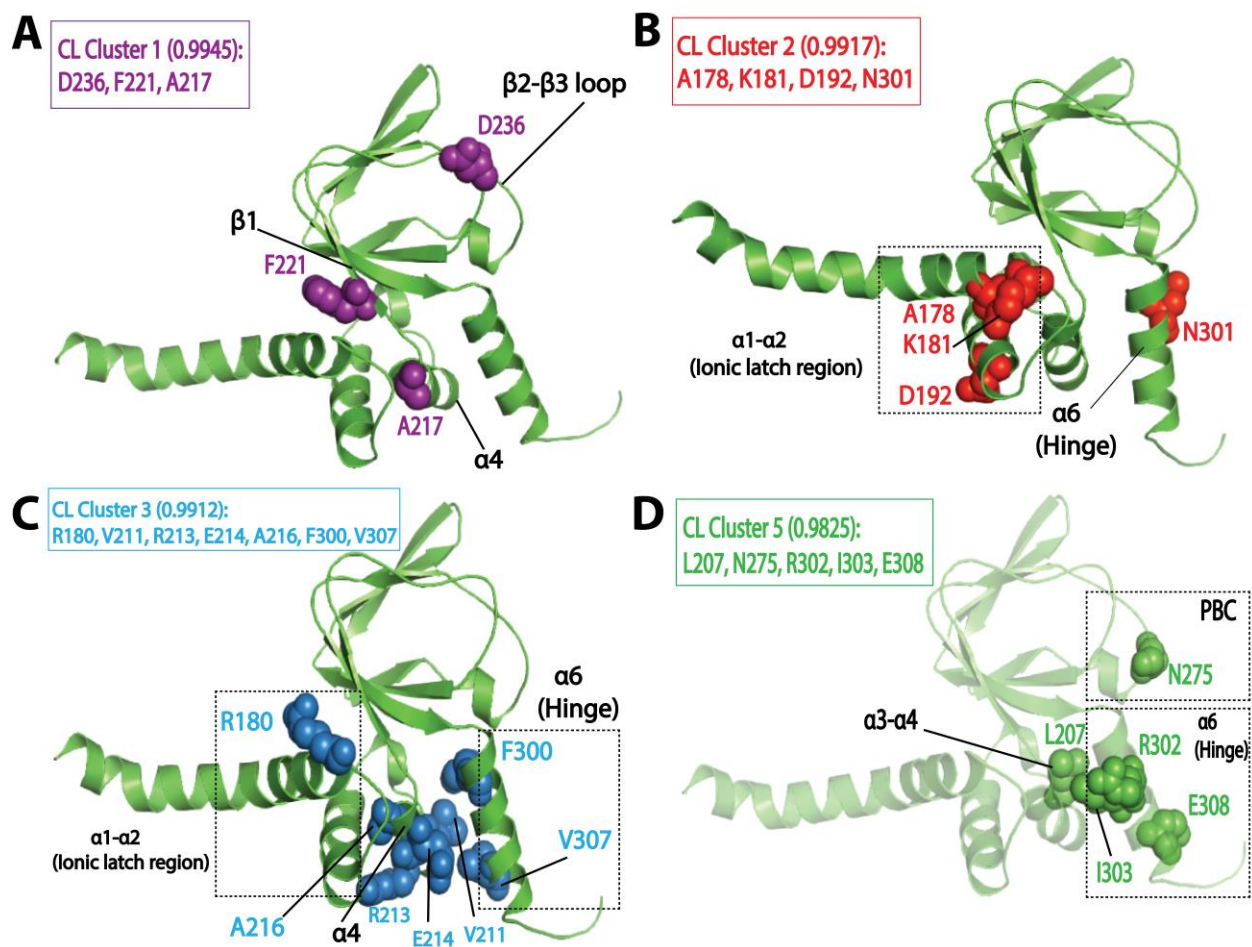
Figure 2.10: Mapping of complete-linkage clusters A) Cluster 1, B) Cluster 2, C) Cluster 3, and D) Cluster 5 as colored spheres on the structure of EPAC1-CNBD. Numbers in brackets next to the respective cluster numbers specify the highest residue-pair correlation coefficient in each complete linkage cluster.
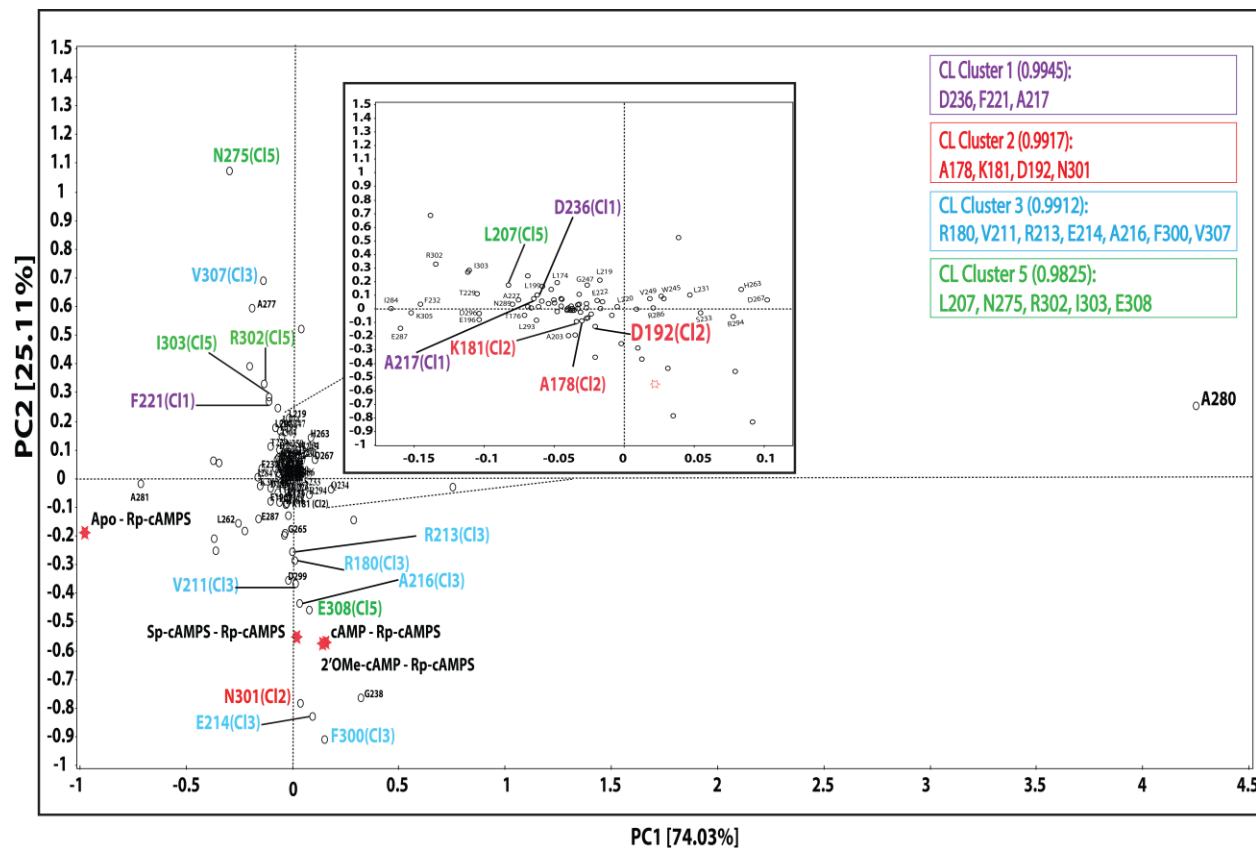
Figure 2.11: Singular value decomposition (SVD) PC1 *vs*. PC2 plot at 306K. Loadings are marked by red stars and show that PC1 is mainly a measure of binding contributions, whereas PC2 reflects, primarily, allosteric contributions. Residue-specific scores are shown as circles. Scores for residues in different complete-linkage clusters are labeled with different colors as per the legend in the top-right corner. The clusters align mostly along PC2 (allostery) rather than PC1(binding).
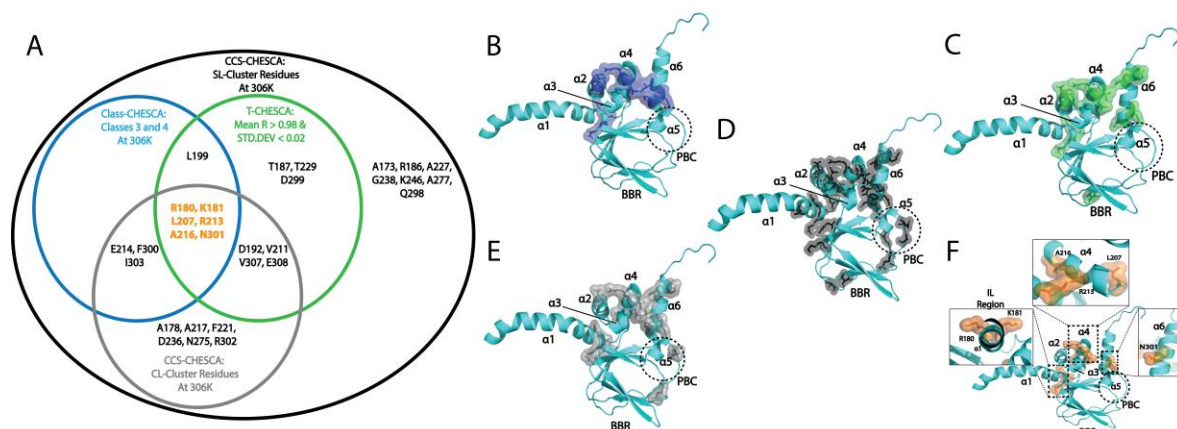
Figure 2.12: Comparison of T-, CLASS-, and CL-CHESCA residue ensembles A) Venn diagram of CHESCA-identified residues, including the proposed T-, CL- and CLASS-CHESCA ensembles aimed at identifying the core allosteric network of EPAC1. The cyan ensemble refers to classes 3 and 4 of the CLASS-based CHESCA at 306K, the green ensemble is for the set of residues from the T-CHESCA with mean Pearson coefficient greater than 0.98 and standard deviation less than 0.02 across temperatures, and the dark (light) grey ensemble describes the residues identified from complete (single)-linkage clustering, *i.e.,* CL (SL)-CHESCA, at 306K. Core residues that are common to all ensembles are highlighted in orange. B-F) The residues identified from each ensemble are mapped on the structure of EPAC1-CNBD and represented as surfaces colored as in panel A. F) Map of the core residues highlighted in orange on the Venn diagram with respective zoomed in panels. Notations for the structural elements are as in Fig. 2.1A, B.

Table 2.1: Core Allosteric Sites Identified by Different CHESCA-Based Approaches

	T-CHESCA[a]	**CLASS-CHESCA**[b,c]	**CL-CHESCA**[c,d]
Residues and 2ary Structure Elements	R180, K181: α1 T187: α1-α2 D192, L199: α2 L207: α3-α4 V211, R213, A216: α4 T229: β2 D299, N301, V307, E308: α6	R180, K181: α1 L199: α2 L207: α3-α4 R213, E214, A216: α4 F300, N301, I303: α6	A178, R180, K181: α1 D192: α2 L207: α3-α4 V211, R213, E214, A216, A217: α4 F221: β1 D236: β2- β3 N275: α5-β7 F300, N301, R302, I303, V307, E308: α6
T % [e]	100%	70%	53%
CLASS% [f]	50%	100%	47%
CL % [g]	71%	90%	100%
Unique	21%	0%	32%

[a] *Residues with $|{<}R{>}| > 0.98$ and $s < 0.02$, as shown in Figure 2.7 insets.* [b] *Classes 3 and 4, as defined in Figure 2.8.* [c] *Implemented at 306K.* [d] *Complete-linkage CHESCA.* [e] *Percentage of residues common to the T-CHESCA sites.* [f] *Percentage of residues common to the CLASS-CHESCA sites.* [g] *Percentage of residues common to the CL-CHESCA sites.* [h] *Percentage of residues unique to each ensemble.*

Chapter 3

QSAR Models of EPAC1-Selective Modulators

3.1 Introduction

Quantitative structure-activity relationships (QSAR) depend on the concept stating that a biological property/activity of a compound is related to the structure of that compound. The information encoded by the structure is described through 'molecular descriptors and the biological property is then expressed as a function of these descriptors [1].

Though the first QSAR study on a cyclic adenosine monophosphate (cAMP)-dependent protein dates to the late 1990s [2], the reported QSARs since then have been quite limited for cyclic nucleotide monophosphate (cNMP) sensors in signaling pathways [3–6].

Here, the first QSAR model for EPAC1 is demonstrated based on a series of derivatives of a novel EPAC1-selective modulator known as I942 [7–9]. I942 showed promising partial agonistic activity on the EPAC1-cyclic nucleotide binding domain (CNBD) [9] and its binding sampled a third state in addition to the active and inactive states of the CNBD. That third state was described as a 'mixed' intermediate displaying a mix of features of the active and inactive state whereby the phosphate binding cassette (PBC) is in the 'in' conformation and the hinge region is in the 'out' conformation.

The novel QSAR model was validated and then used to predict affinities for a series of I942 analogues that were 'unknown' to the model. The affinity for the most promising candidate, known as MLGM-2013, as predicted by our validated QSAR model was confirmed through fluorescence competition assays. In addition, we investigated the mechanism of action of MLGM-

2013 using NMR experiments, revealing a new avenue to design I942 analogs with enhanced potency through modifications of its phenyl moiety.

3.2 Methods

QSAR Model. The I942-based QSAR was developed using the I942 analogues synthesized by Wang et al [8]. The respective molecules were built in MolView [10], transferred to the 3D model viewer and the energy of the 3D conformers was minimized using the Jmol energy minimization based on the MMFF94 forcefield [11] and a limit of 100 minimization steps at a time. PaDEL-Descriptor[12]was used to calculate the 1D and 2D molecular descriptors from the minimized structures. Partition of the molecules into training and test sets was implemented according to an 80:20 ratio for the training *vs.* test sets, respectively, and considering their measured affinities, reported as relative fluorescence intensity (RFI) percentages [8]. Specifically, molecules in each set were chosen to sample the entire spectrum of RFI values. Following these criteria, the original dataset was divided into eleven different training and test partitions. To check for potential outlier RFI values, the Z-score was computed as [13]:

$$Z_i = \frac{y_i - \bar{y}}{s} \tag{3.1}$$

where y_i is the RFI value of a given I942 analog, $\bar{y}$ is the mean and s is the standard deviation. Molecules with $|Z_i|$ greater than 2.5 are considered outliers [13]. However, the dataset of I942 analogs did not contain Z-score outliers and therefore, all the molecules were included in the model.

RapidMiner Studio [14] was used to narrow down the number of descriptors for the QSAR model by applying the forward selection method [15] on the training set. The method entails sequential addition of molecular descriptors that improve the performance of the model, *i.e.,* descriptors leading to enhanced linear regression correlations. The stopping criteria for the sequential addition are 1) There is no improvement in model performance or 2) The maximum number of descriptors that satisfy a 5:1 ratio for number of molecules *vs.* number of descriptors was reached [16]. The descriptors chosen were then fed into RapidMiner to generate the linear regression model, which was applied to the training and test sets generating a coefficient of multiple determination [17] (R^2) for each. R^2 is calculated as:

$$R^2 = \frac{\sum_i(\hat{y}_i - \bar{y})^2}{\sum_i(y_i - \bar{y})^2} \qquad (3.2)$$

where, $\hat{y}_i$ is the calculated dependent variable, i.e., the predicted RFI value, y_i is the observed or actual RFI value and $\bar{y}$ is the mean RFI.

An additional parameter reporting on the QSAR quality, known as the root mean squared error (RMSE)[17] describes the range of error in the model's predictions and is defined as:

$$RMSE = \sqrt{\frac{\sum_i(y_i - \hat{y}_i)^2}{n}} \qquad (3.3)$$

where, y_i is an observed RFI value, $\hat{y}_i$ is the corresponding predicted RFI value and n is the number of molecules in the training set, in this case, 45. The RMSE values of both training and test sets are well below the corresponding standard deviations (σ) of the observed RFI percentages, meaning that the predictions are significantly reliable.

As an initial mean of validating the QSAR model, we relied on cross-validation (CV), which is a form of internal validation of the model's predictivity utilizing an approach called the 'Leave-Many-Out' (LMO) [18] method. LMO holds back a portion of the training set as a small test set and applies the model without that test set. The process was repeated for 10 iterations and the squared correlation obtained was represented as an average value of the multiple iterations. The descriptor selection process and QSAR workflow outlined above were repeated for each of the eleven different training and test partitions and average statistical parameters across these partitions were computed.

Protein Purification. The wild-type $EPAC1_h$ (149-318) construct was purified according to previously published protocols [9,19–22]. The protein was cultured in either Lysogeny Broth (LB) or ^{15}N-labeled M9 minimal media to prepare unlabeled or ^{15}N-labeled EPAC1-CNBD, respectively, as needed for fluorescence or NMR measurements.

Preparation of I942 and MLGM Compounds. Compounds were dissolved in deuterated $DMSO-d_6$ to prepare 10 mM stock solutions. I942 was purchased from Life Chemicals (purity > 99%) and was prepared as a 10 mM stock solution with deuterated $DMSO-d_6$.

8-NBD-cAMP Competition Assay. 8-(2-[7-nitro-4-benzofurazanyl] aminoethylthio) adenosine-3',5'-cyclic monophosphate (8-NBD-cAMP) binds to EPAC1-CNBD with high affinity and the binding can be monitored by fluorescence intensity changes [23]. Unlabeled EPAC1-CNBD was used for this assay. The K_D measurements of EPAC1-CNBD in complex with either I942, MLGM-2013, or MLGM-2017 were recorded from the decrease in fluorescence intensity as a result of 8-NBD-cAMP competitive displacement [9,23]. The compounds were added at concentrations between 0 to 300 µM to solutions of 2.5 µM and 0.5 µM of EPAC1-CNBD and 8-

NBD-cAMP, respectively. The NMR buffer (*vide* infra) was used to bring the final volume of the samples to 250 µL. Samples were added to Corning 96-well half area plates (120 µL per well) after an incubation period of at least 30 minutes at room temperature to allow for equilibration. A Cytation 5 plate reader was used to scan the plate using excitation and emission wavelengths of 485 nm and 535 nm, respectively. The equation used for fitting the competitive binding isotherms was applied as previously described [24].

NMR Measurements. NMR experiments were acquired using a Bruker Avance or NEO 700 MHz spectrometer with a TCI cryoprobe. For the HSQC experiments, 350 µM of the ligand (I942/MLGM) was added to 50 µM of EPAC1-CNBD in NMR buffer with 5% D_2O. The same volume of DMSO-d_6, present in the NMR samples with ligands, was added to the apo sample to exclude the effect of DMSO-d_6 from the chemical shift perturbation assessment. The ^{15}N-^{1}H-HSQC experiment utilized an Echo and Anti-echo PFG selection along with a water flip-back and the operating temperature was 306K. The time domain digitization points were 2048 and 128 for the ^{1}H and ^{15}N dimensions, respectively, and the spectral widths were 16.23 ppm for ^{1}H and 38 ppm for ^{15}N. The number of scans was 64 and the recycle delay was 1 second. The spectra were processed in TopSpin (Bruker), where the size of the real spectrum (SI) was 2048 and 512 for the ^{1}H and the ^{15}N dimensions, respectively. Sine bell shift (SSB) values of 2 and 3 were applied for the ^{1}H and ^{15}N dimensions, respectively, and a sine squared window function (WDW) was applied for both dimensions. Forward line prediction (LPfc) was utilized for the ^{15}N dimension, where the number of LP coefficients was 32. The chemical shifts were referenced to ^{15}N-acetyl glycine and were assigned through comparison with the apo and cAMP-bound EPAC1-CNBD at 306K that were previously acquired and assigned [9].

The compounded chemical shift differences (ΔCCS) between ligand bound EPAC1 and the apo form were calculated using the following equation:

$$\Delta CCS = \sqrt{(\Delta\delta H)^2 + (0.2 \times \Delta\delta N)^2} \qquad (3.4)$$

Samples for 1D saturation transfer difference (STD) were prepared using a 50 µM EPAC1-CNBD solution that was buffer-exchanged with a 20 mM sodium phosphate buffer containing 50 mM NaCl, pH 7.4 and 99.9% D_2O. PD-10 Desalting columns (GE Healthcare) were used to facilitate the exchange by a gravity protocol. 350 µM of MLGM-2013 (final concentration) was added to 50 µM of EPAC1-CNBD and the saturation frequency in the STD experiments was set to 0.8 ppm to saturate the region of protein peaks (*i.e.,* methyl region) that is further away from the MLGM-2013's signal. An off-resonance saturation of 30 ppm was applied to the STR experiments and the STD/STR ratios normalized to the largest value were compared to those acquired for I942 [9]. The spectra were referenced to DMSO (2.48 ppm) and the assignments of MLGM-2013 were obtained by comparison with previously established assignments for I942 [9]. The STD spectra were acquired at 298K with a time domain of 32768 points and number of scans of 128 and 1024 for STR and STD experiments, respectively. Eight dummy scans were used for both STD and STR. The spectral width was 11.7057 ppm, and the transmitter frequency was set to 4.697 ppm.

The chemical shift projection analysis (CHESPA) was implemented according to previous protocols [9,25,26] and using the NMRFAM-SPARKY plugin [27]. The reference vector is defined from the apo to the cAMP-bound EPAC1 state, while the perturbation vector is defined from the cAMP-bound form to the I942-analog ligand-bound form. The minimum cut-off for the ΔCCS values of

both vectors was set to 0.02 ppm and the cos θ and fractional activation (X) values were computed according to the following formulae:

$$\cos \theta = \frac{A \cdot B}{|A||B|} \tag{3.5}$$

$$X = \frac{|A|}{|B|} \cos \theta \tag{3.6}$$

3.3 Results and Discussion

QSAR Model Development. The QSAR models for I942 were developed according to the flowchart described in Figure 3.1 where the partitioning of training and test sets was implemented maintaining an 80:20 ratio (Figure 3.2A,B) and a balanced distribution of affinities, as quantified by relative fluorescence intensity (RFI) percentage values (Figure 3.2C). QSAR models were developed for a total of 11 distinct training *vs.* test set partitions in compliance with the same criteria (Figure 3.3 and Table 3.1). The average values of statistical parameters describing the QSAR quality were then computed across the 11 resulting QSAR models (Table 3.2). One of the primary QSAR quality descriptors is the coefficient of multiple determination, [17] referred to as R^2, which reflects the overall accuracy of the RFI values predicted by the model compared to the actual measured RFI percentages. As seen from Table 3.2, the R^2 values are high (above threshold) for both training and test sets, reflecting the ability of the model to reproduce the original data as well as to predict external data, respectively. The data points in the correlation plots are also closely arranged around the line of best fit set to have a zero intercept (Figure 3.4).

We also computed the average statistical parameters of the 11 QSAR models without imposing a zero-intercept, showing slightly improved performance with the intercept set to zero.

Interestingly, the QSAR models obtained from the 11 distinct partitions, exhibited classes of recurring molecular descriptors in the multiple linear regressions. The shared descriptors are 2D in nature and fall in the 'autocorrelation' category, which essentially captures the distribution of physicochemical properties across the spatial arrangement of atoms [28]. In our particular model, the main physicochemical properties are (a) the intrinsic state, represented by the GATS5s, AATS5s and MATS5s descriptors, and these report on the electronegativity of the atom in its valence state, as well as (b) the Sanderson electronegativity [29] represented by descriptors such as ATSC8e and AATS5e. It was interesting to observe consistently positive coefficients for the descriptors in the linear regression equations, which reflects a positive correlation between these descriptors and the RFI values.

Affinity Prediction of Unknown Compounds and Experimental Validation through Fluorescence Competition Assays. After model validation, both internal, through the training set and cross-validation R^2, and external, through the test set R^2 (Figure 3.2), the model with the highest test set R^2 value (considering both zero and non-zero intercepts) was used to predict the RFI values of a new set of I942 analogues that were not part of either the original training or test sets (Figure 3.5A). Therefore, the I942 derivatives in the new set, referred to with the 'MLGM' code in Figure 3.5A, are essentially 'unknown' to our QSAR model, but they all share the same skeleton common to other I942 analogues with a sulfonamide flanked by phenyl and linked naphthyl moieties (Figure 3.5A). Based on the RFI values predicted by our QSAR model for the new set of I942 analogs (Table 3.3), the MLGM-2013 derivative (Figure 3.5A) stood out as having

the lowest predicted RFI, pointing to better binding affinity for EPAC1 relative to I942. On the contrary, the MLGM-2017 derivative (Figure 3.5A) was predicted to exhibit the weakest EPAC1 affinity (highest RFI value) within the new set.

To confirm our predictions, a competition assay was preformed using the fluorescently tagged cAMP known as 8-(2-[7-nitro-4-benzofurazanyl] aminoethylthio) adenosine-3',5'-cyclic monophosphate (8-NBD-cAMP)[23]. The displacement of 8-NBD-cAMP by a competing ligand at increasing concentrations was used to measure the dissociation constant (K_D) of MLGM-2013, I942 as well as MLGM-2017, as a negative control. The assay clearly showed a significant enhancement of the binding affinity of MLGM-2013 relative to I942 (Figure 3.5B), while MLGM-2017 resulted in a significantly higher K_D value compared to that of I942, as expected, further confirming the validity of our QSAR model's predictions. To gain structural insight into the enhanced affinity of MLGM-2013 and its mechanism of action, we investigated the interactions of this I942 analog with the EPAC1 CNBD using NMR.

The MLGM-2013 Binding Mode. The binding of MLGM-2013 to the EPAC1 CNBD was monitored through ^{15}N-^{1}H-HSQC spectra (Figure 3.5C) and the corresponding chemical shift changes (ΔCCS) are reported in Figure 3.5D. Figure 3.5D shows major ppm variations induced by MLGM-2013 in key CNBD regions such as the BBR, the PBC and the hinge region, quite similar to the chemical shift changes observed upon cAMP binding (Figure 3.5D). The ΔCCS measured for MLGM-2013 are also quite similar to those observed for I942 (Figure 3.5E), suggesting a similar binding mode, with the notable exception of N275 located in the PBC (Figure 3.5E).

To further elucidate the difference in binding affinity between I942 and MLGM-2013, saturation transfer difference (STD) experiments were performed to map the binding epitopes of

MLGM-2013 and assess the proximity of ligand protons to the EPAC1-CNBD (Figure 3.5F) [9,30]. Interestingly, we found that the STD/STR ratios for MLGM-2013 are higher for several phenyl protons compared to I942 with the most significant increase observed for the tertiary butyl protons located at the para position of the phenyl group (Figure 3.5F and Figure 3.6). As opposed to the single methyl group at that location in I942, the additional methyls of tertiary butyl offer more contacts with the protein as seen through STD/STR ratios of 0.62 *vs.* 0.39 for MLGM-2013: EPAC1-CNBD *vs.* I942: EPAC1-CNBD, respectively (Figure 3.5F and Figure 3.6). Based on the N275 outlier observed in Figure 3.5E, we hypothesized that the enhanced contacts of the tertiary butyl in MLGM-2013 are with the PBC of the EPAC1-CNBD.

To test out hypothesis, we measured the EPAC1-CNBD compounded chemical shift changes (ΔCCS) between MLGM-2013 and MLGM-2014 which lacks any phenyl substituents, and therefore, serves as a useful reference ligand to capture the effect of the MLGM-2013 tertiary butyl para substituent (Figure 3.7A). Despite the absence of phenyl substituents, MLGM-2014, previously referred to as I178 [7], was shown to bind EPAC1 and result in an IC_{50} of ~40μM [7]. Figure 3.7A reports the residue-specific MLGM-2013 *vs.* MLGM-2014 ΔCCS values as well as the corresponding I942 *vs.* MLGM-2014 ΔCCS as a control. Although the ΔCCS values of the EPAC1 CNBD in the presence of MLGM-2013 or I942 relative to MLGM-2014 are similar, the most evident difference is observed in the PBC. MLGM-2013 yields a markedly higher ΔCCS, reflecting additional perturbations in that region due to the bulkier, tertiary butyl moiety at the para phenyl position. These results confirm our hypothesis that the tertiary butyl group of MLGM-2013 interacts with the PBC and that such contacts are unique of MLGM-2013, possibly explaining the enhanced affinity of MLGM-2013 relative to the parent I942 compound.

The MLGM-2013 Mechanism of Action. To gain more insight on the possible mechanism of action of MLGM-2013, the CHEmical Shift Projection Analysis (CHESPA)[9,25,26] was implemented for the MLGM-2013-bound EPAC1 and compared with the CHESPA of I942-bound EPAC1. The CHESPA reports on the ligand-induced shifts in the auto-inhibitory equilibria between inactive and active conformations. Using the CHESPA vector scheme in Figure 3.7B, the fractional activation (X) as well as the cos θ values were computed for MLGM-2013, revealing primarily negative values (Figure 3.7C and D) or I942 (Figure 3.7E and F). This indicates a partial but quite consistent shift towards the apo-inactive conformation of the EPAC1-CNBD, reflecting a partial agonistic activity.

When the CHESPA profiles of MLGM-2013 (Figure 3.7C, D) are compared to those of I942 (Figure 3.7E, F), one of the most notable differences is observed for PBC residues such as A272, N275 and A277 (asterisks in Figure 3.7E). These sites exhibit markedly more negative X and cos θ values for MLGM-2013 than I942, suggesting a more significant shift towards the inactive state in that region compared to I942. Additionally, MLGM-2013 demonstrates a more negative average X value at the PBC region compared to I942, whereas the average X value for the hinge region is slightly less negative compared to I942 (dotted lines in Figure 3.7C, E). These average X values suggest that MLGM-2013 bound to EPAC samples an inactive state with PBC out, hinge out and a population of around 60%, while the population of the mixed intermediate with the PBC in, hinge out [9] is negligible. Based on these results, MLGM-2013 promises to serve as a more potent EPAC1-CNBD modulator than I942 with enhanced inhibitory activity.

The QSAR model proposed here, therefore, serves as an effective tool to virtually screen compound libraries for EPAC1 binding, thus aiding the identification of novel EPAC1-selective drug candidates.

3.4 References

(1)   Peter, S. C.; Dhanjal, J. K.; Malik, V.; Radhakrishnan, N.; Jayakanthan, M.; Sundar, D.; Sundar, D. Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* **2019**, *1–3*, 661–676.

(2)   Mureşan, S.; Jastorff, B.; Kurunczi, L.; Simon, Z. Comparative Structure-Affinity Relations by MTD for Binding of Cycloadenosine Monophosphate Derivatives to Protein Kinase Receptors. *Quantitative Structure-Activity Relationships* **1995**, *14* (3), 242–248.

(3)   Chakraborti, A. K.; Gopalakrishnan, B.; Sobhia, M. E.; Malde, A. 3D-QSAR Studies on Thieno[3,2-d]Pyrimidines as Phosphodiesterase IV Inhibitors. *Bioorganic and Medicinal Chemistry Letters* **2003**, *13* (8), 1403–1408.

(4)   Chakraborti, A. K.; Gopalakrishnan, B.; Sobhia, M. E.; Malde, A. 3D-QSAR Studies of Indole Derivatives as Phosphodiesterase IV Inhibitors. *European Journal of Medicinal Chemistry* **2003**, *38* (11–12), 975–982.

(5)   Dong, X.; Zheng, W. A New Structure-Based QSAR Method Affords Both Descriptive and Predictive Models for Phosphodiesterase-4 Inhibitors. *Current Chemical Genomics* **2008**, *2*, 29–39.

(6)   Sharma, V.; Kumar, H.; Wakode, S. Pharmacophore Generation and Atom Based 3D-QSAR of Quinoline Derivatives as Selective Phosphodiesterase 4B Inhibitors. *RSC Advances* **2016**, *6* (79), 75805–75819.

(7)   Parnell, E.; Mcelroy, S. P.; Wiejak, J.; Baillie, G. L.; Porter, A.; Adams, D. R.; Rehmann, H.; Smith, B. O.; Yarwood, S. J. Identification of a Novel, Small Molecule Partial Agonist for the Cyclic AMP Sensor, EPAC1 OPEN.

(8)   Wang, P.; Luchowska-Stańska, U.; van Basten, B.; Chen, H.; Liu, Z.; Wiejak, J.; Whelan, P.; Morgan, D.; Lochhead, E.; Barker, G.; Rehmann, H.; Yarwood, S. J.; Zhou, J. Synthesis and Biochemical Evaluation of Noncyclic Nucleotide Exchange Proteins Directly Activated by CAMP 1 (EPAC1) Regulators. *Journal of Medicinal Chemistry* **2020**, *63* (10), 5159–5184.

(9)   Shao, H.; Mohamed, H.; Boulton, S.; Huang, J.; Wang, P.; Chen, H.; Zhou, J.; Luchowska-Stańska, U.; Jentsch, N. G.; Armstrong, A. L.; Magolan, J.; Yarwood, S.; Melacini, G. Mechanism of Action of an EPAC1-Selective Competitive Partial Agonist. *Journal of Medicinal Chemistry* **2020**, *63* (9), 4762–4775.

(10)   Smith, T. J. MOLView: A Program for Analyzing and Displaying Atomic Structures on the Macintosh Personal Computer. *Journal of Molecular Graphics* **1995**, *13* (2), 122–125.

(11)   Halgren, T. A. *Performance of MMFF94**; John Wiley & Sons, Inc, 1996; Vol. 17.

(12) Yap, C. W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *Journal of Computational Chemistry* **2011**, *32* (7), 1466–1474.

(13) Rousseeuw, P. J.; Hubert, M. Robust Statistics for Outlier Detection. *WIREs Data Mining and Knowledge Discovery* **2011**, *1* (1), 73–79.

(14) Kotu, V.; Deshpande, B. *Predictive Analytics and Data Mining*; Elsevier, 2015.

(15) Jobson, J. D. *Applied Multivariate Data Analysis*; Springer Texts in Statistics; Springer New York: New York, NY, 1991.

(16) Todeschini, R.; Consonni, V.; Maiocchi, A. The K Correlation Index: Theory Development and Its Application in Chemometrics. *Chemometrics and Intelligent Laboratory Systems* **1999**, *46* (1), 13–29.

(17) Gramatica, P. On the Development and Validation of QSAR Models. In *Computational Toxicology: Volume II*; Reisfeld Brad and Mayeno, A. N., Ed.; Humana Press: Totowa, NJ, 2013; pp 499–526.

(18) Gramatica, P. Principles of QSAR Models Validation: Internal and External. *QSAR and Combinatorial Science* **2007**, *26* (5), 694–701.

(19) Mazhab-Jafari, M. T.; Das, R.; Fotheringham, S. A.; SilDas, S.; Chowdhury, S.; Melacini, G. 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 American Chemical Society* **2007**, *129* (46), 14482–14492.

(20)   Boulton, S.; Selvaratnam, R.; Blondeau, J. P.; Lezoualc'H, F.; Melacini, G. Mechanism of Selective Enzyme Inhibition through Uncompetitive Regulation of an Allosteric Agonist. *Journal of the American Chemical Society* **2018**, *140* (30), 9624–9637.

(21)   Selvaratnam, R.; Chowdhury, S.; VanSchouwena, B.; Melacini, G. Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108* (15), 6133–6138.

(22)   Das, R.; Chowdhury, S.; Mazhab-Jafari, M. T.; SilDas, S.; Selvaratnam, R.; Melacini, G. Dynamically Driven Ligand Selectivity in Cyclic Nucleotide Binding Domains. *The Journal of biological chemistry* **2009**, *284* (35), 23682–23696.

(23)   Tsalkova, T.; Mei, F. C.; Cheng, X. A Fluorescence-Based High-Throughput Assay for the Discovery of Exchange Protein Directly Activated by Cyclic AMP (EPAC) Antagonists. *PLoS ONE* **2012**, *7* (1).

(24)   Wang, Z. X. An Exact Mathematical Expression for Describing Competitive Binding of Two Different Ligands to a Protein Molecule. *FEBS letters* **1995**, *360* (2), 111–114.

(25)   Selvaratnam, R.; VanSchouwen, B.; Fogolari, F.; Mazhab-Jafari, M. T.; Das, R.; Melacini, G. The Projection Analysis of NMR Chemical Shifts Reveals Extended EPAC Autoinhibition Determinants. *Biophysical Journal* **2012**, *102* (3), 630–639.

(26)   Byun, J. A.; Melacini, G. NMR Methods to Dissect the Molecular Mechanisms of Disease-Related Mutations (DRMs): Understanding How DRMs Remodel Functional Free Energy Landscapes. *Methods* **2018**, *148*, 19–27.

(27)   Shao, H.; Boulton, S.; Olivieri, C.; Mohamed, H.; Akimoto, M.; Subrahmanian, M. V.; Veglia, G.; Markley, J. L.; Melacini, G.; Lee, W. CHESPA/CHESCA-SPARKY: Automated NMR Data Analysis Plugins for SPARKY to Map Protein Allostery. *Bioinformatics* **2021**, *37* (8), 1176–1177.

(28)   Todeschini, R.; Consonni, V. Molecular Descriptors for Chemoinformatics. *Molecular Descriptors for Chemoinformatics* **2010**, *2*, 1–252.

(29)   Sanderson, R. T. Electronegativity and Bond Energy. *Journal of the American Chemical Society* **2002**, *105* (8), 2259–2261.

(30)   Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *Journal of the American Chemical Society* **2001**, *123* (25), 6108–6117.

(31)   Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C. P.; Agrawal, R. K. Validation of QSAR Models - Strategies and Importance. *International Journal of Drug Design and Disocovery* **2011**, *2* (3), 511–519.

(32)   Roy, K.; Mitra, I. On Various Metrics Used for Validation of Predictive QSAR Models with Applications in Virtual Screening and Focused Library Design. *Combinatorial Chemistry & High Throughput Screening* **2011**, *14* (6), 450–474.
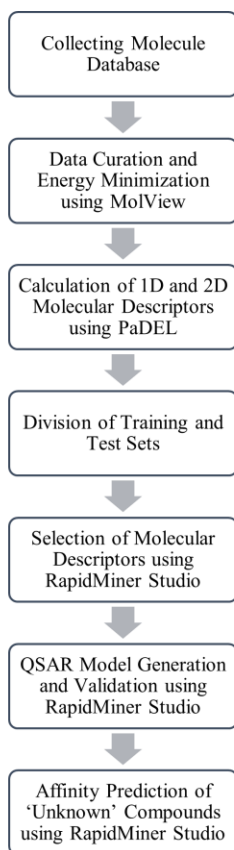
Figure 3.1: Flowchart describing the workflow of the QSAR model generation and validation
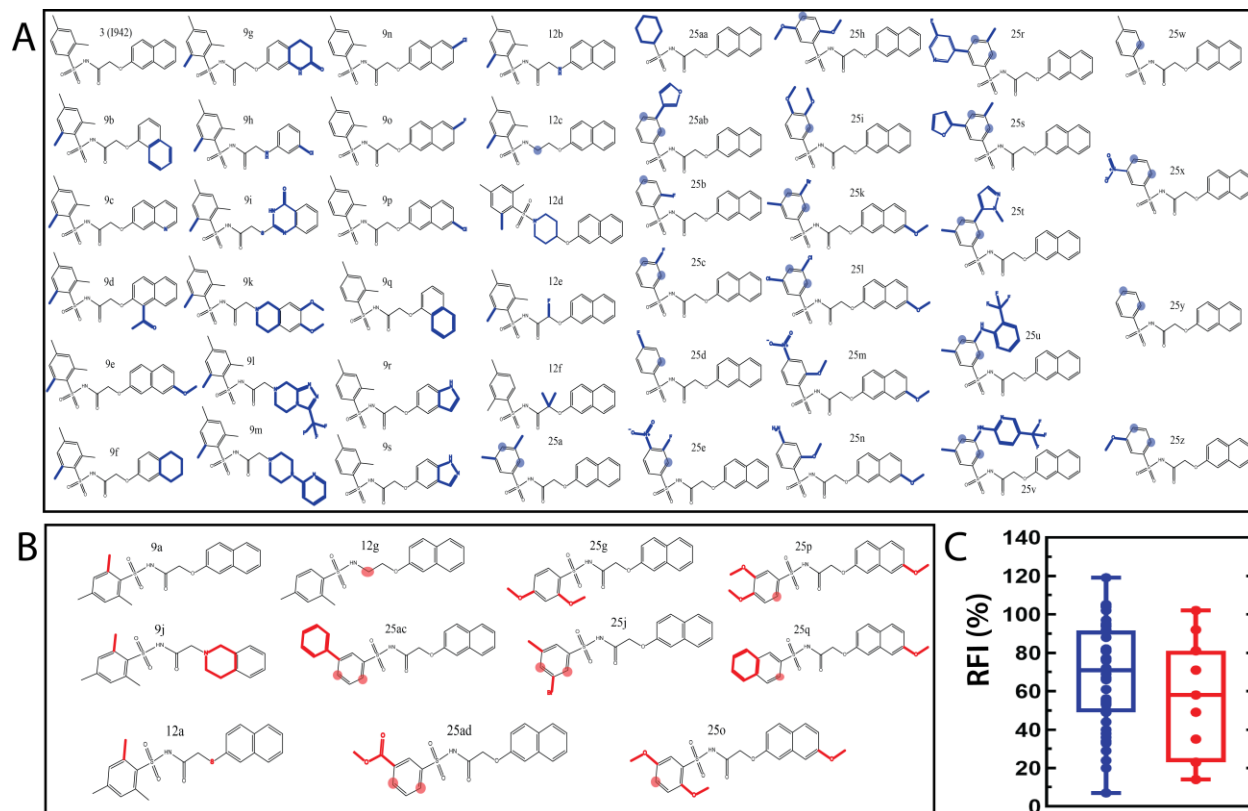
Figure 3.2: Molecular structures of the I942 analogues in the A) training and B) test sets. The training and test sets include 45 and 11 molecules, respectively. The modifications, relative to I942 (3), are marked by blue for the training and red for the test set. The shaded circles highlight the positions that lack substituents originally found in I942. C) Box plot representation for the distribution of the RFI values in the training (blue) and test (red) sets.
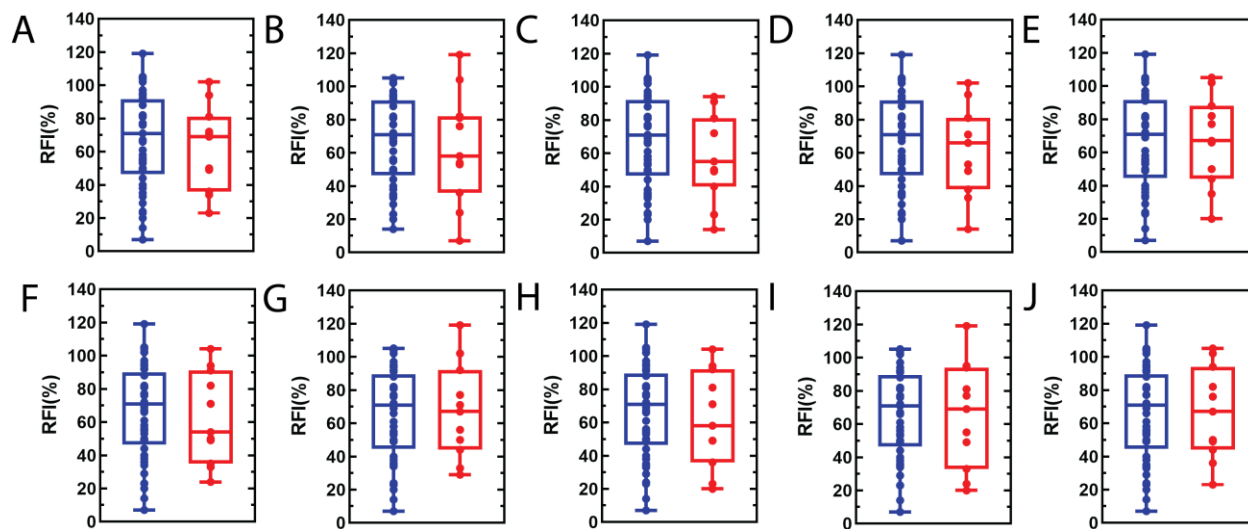
Figure 3.3: (A-J) Box plot representations for the range of RFI values of different training (blue) and test (red) set divisions. The number of training set molecules is 45 and test set molecules is 11 for all the divisions.
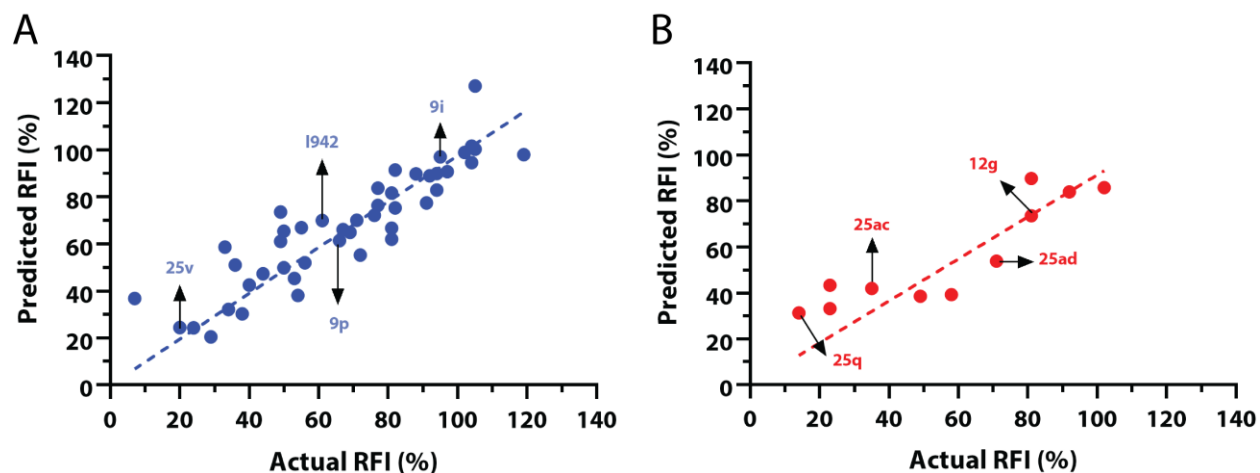
Figure 3.4: Predicted vs. measured relative fluorescence intensities (RFI) correlation plots for the A) training and B) test sets of I942 analogues shown in Figure 3.2. Representative molecules are marked with black arrows and labeled as in Figure 3.2.
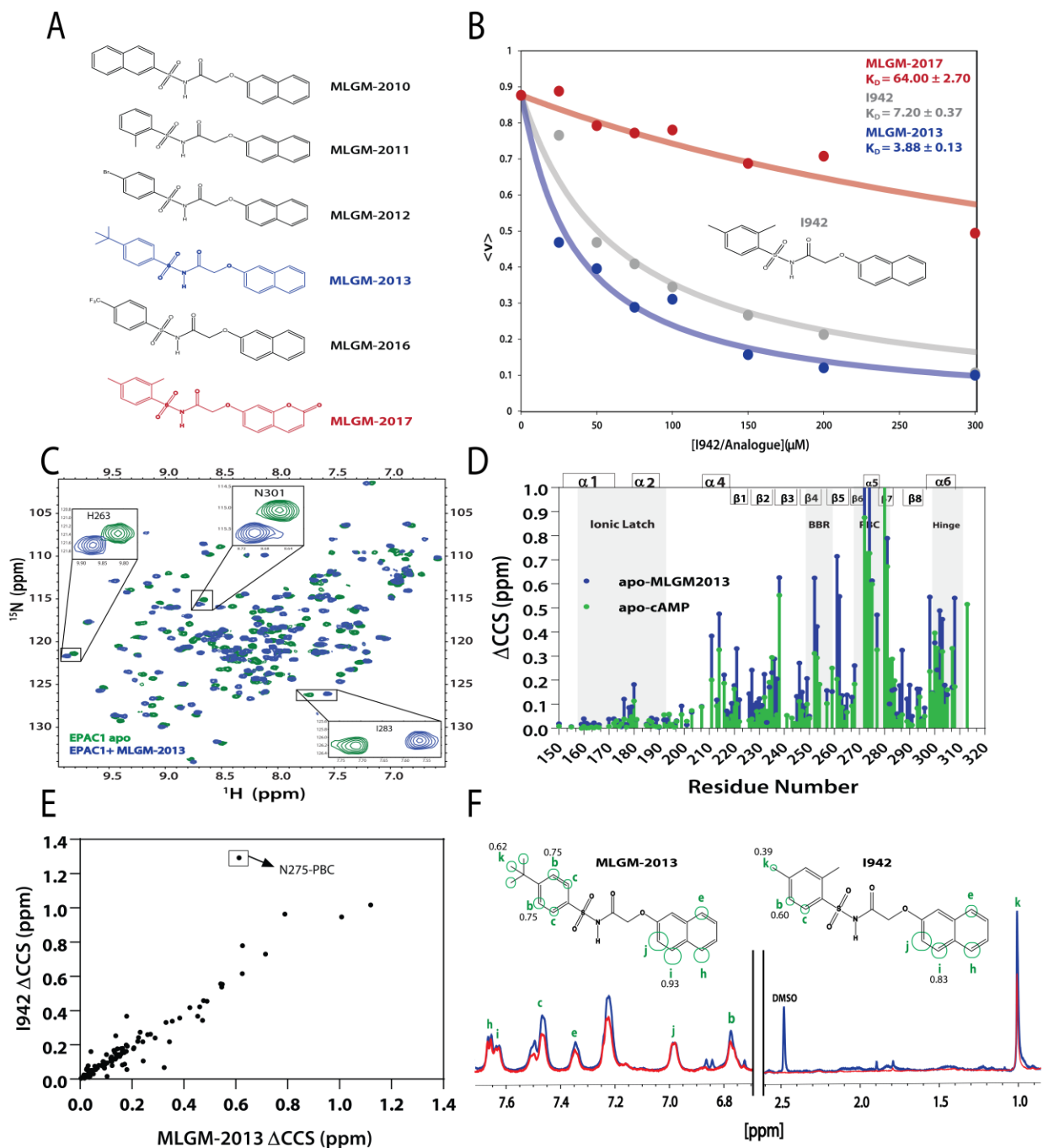
Figure 3.5: Independent validation of the QSAR model. A) Molecular structures of new synthesized I942 analogues that are 'unknown' to the QSAR model. MLGM-2013, which was predicted to have the highest affinity towards EPAC1-CNBD (Table 3.3), is highlighted in blue and MLGM-2017, the compound with the lowest predicted affinity (Table 3.3) is shown in red.

B) EPAC1-CNBD binding isotherm for I942 (grey), MLGM-2013 (blue), and MLGM-2017 (red) measured through the 8-NBD-cAMP fluorescence-based competition assay. The resulting measured dissociation constants are included in the top right corner. The percentage of 8-NBD-cAMP bound to EPAC1 is represented by <v> on the y-axis. C) The chemical shift differences between apo EPAC1-CNBD (green) and EPAC1-CNBD bound to 350 µM of MLGM-2013 (blue) were monitored by ^{15}N-^{1}HSQC spectra. D) The compounded chemical shift variations between apo EPAC1-CNBD and EPAC1-CNBD bound to MLGM-2013 (350 µM) or cAMP (1mM) are plotted as blue and green bars, respectively. The secondary structure is shown on the top of the plot in boxes and key regions are highlighted in grey. E) Compounded chemical shift differences of EPAC1-CNBD in the presence of MLGM-2013 are plotted against the compounded chemical shift differences of EPAC1-CNBD in the presence of I942 at 350 µM. F) 1D saturation-transfer reference (STR, blue) spectrum of EPAC1-CNBD: MLGM-2013 overlayed with the scaled saturation transfer difference (STD, red) spectrum. The assigned protons are marked in green and represented as circles on the structure of MLGM-2013 where the size of the circles reflects the relative STD/STR ratios (normalized to proton j with the highest STD/STR ratio). The structure of I942 with the previously determined STD/STR ratios [9] are shown for comparison whereby the STD/STR ratios with the most significant differences are reported near the corresponding proton.
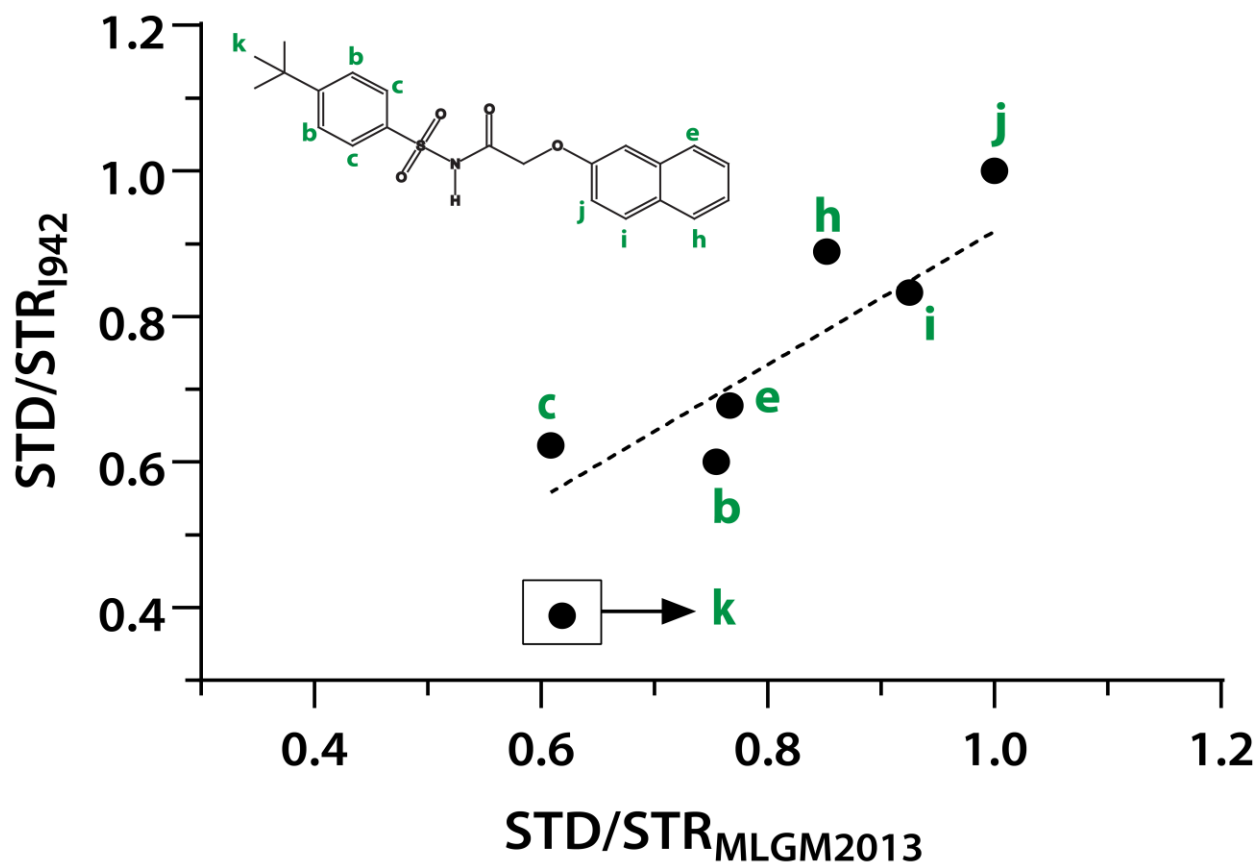
Figure 3.6: Correlation plot for the normalized STD/STR intensity ratios of protons of MLGM-2013 against protons of I942. The proton marked by a box and labeled as 'k', corresponds to the proton with the greatest difference in STD/STR ratio.

Figure 3.7: Effect of MLGM-2013's tertiary butyl moiety on EPAC1 residues and CHESPA analysis. A) Residue specific compounded chemical shift variations of between MLGM-2013-bound (green) or I942-bound (blue) and MLGM-2014-bound EPAC1 CNBD. Structural differences between the ligands are highlighted with corresponding color codes. B) Vector

representation of the CHESPA analysis. C) Fractional activation values and D) cos θ of MLGM-2013-bound EPAC1 relative to cAMP-bound EPAC1 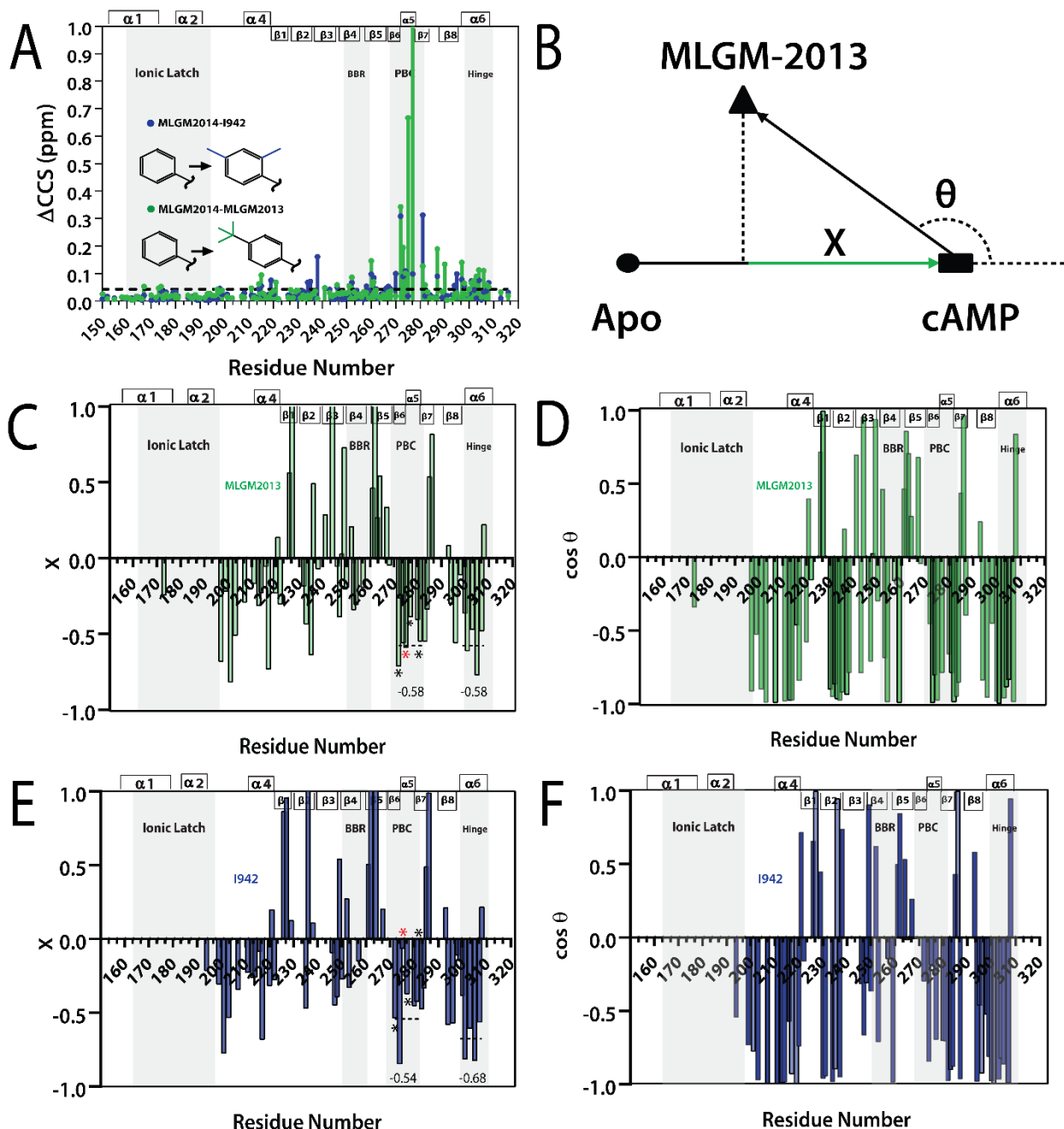where values greater than 1 or less than -1 are not within the scale of the plot. E) Fractional activation and F) cos θ values of I942-bound EPAC1 measured under the same conditions as that of MLGM-2013-bound EPAC1. The asterisks correspond to the residues in the PBC which are more negative in the MLGM-2013-bound structure and the red asterisk marks N275, which exhibits the greatest change. Dotted lines in C and E represent the average X values in the PBC and hinge region. The secondary structure of EPAC1-CNBD is shown in the same way as Figure 3.5D.

Table 3.1: Code names for the test set molecules used for each of the 10 dataset partitions. The training set molecules are the remaining part of the dataset (~ 80%).

Division 1	Division 2	Division 3	Division 4	Division 5	Division 6	Division 7	Division 8	Division 9	Division 10
25ad	12e	12g	12g	12d	25ac	12a	25ad	25aa	12c
25b	12f	25aa	25a	25ac	25c	25i	25e	25b	12e
25e	25ab	25g	25ab	25c	25i	25k	25j	25m	12f
25h	25e	25h	25d	25v	25l	25u	25o	25v	25g
25p	25j	25i	25q	25w	25m	25w	25p	25z	25o
25r	25l	25q	25y	25x	25z	25x	25r	9b	25r
25t	25m	25s	25z	9f	9e	25z	25v	9e	25w
9b	25n	25y	9i	9h	9l	9g	25y	9g	25x
9e	25r	9o	9k	9j	9q	9k	9b	9i	9j
9j	9l	9r	9p	9n	9r	9m	9c	9m	9n
9n	9m	9s	9q	9p	9s	9q	9d	9o	9s

Table 3.2: Parameters for the QSAR model developed for the I942 analogues*

	Training Set	Test Set	Cross-Validation (CV)	Threshold [31,32]
R^2	0.972 ± 0.008	0.929 ± 0.021	0.772 ± 0.055	$R^2 > 0.600$ and > 0.500 for CV
σ	27.69 ± 0.60	28.69 ± 2.40	-	-
RMSE	11.83 ± 1.81	19.09 ± 3.13	12.78 ± 1.50	$RMSE < \sigma$
k	0.972 ± 0.008	0.960 ± 0.090	-	$0.850 \leq k \leq 1.150$
* Standard deviations were computed using data from eleven different partitioning of training *vs.* test sets.				

Table 3.3: Predicted RFI values for a series of I942 analogues 'unknown' to the QSAR Model

Compound Name	Predicted RFI(%)
MLGM-2013	25.87
MLGM-2010	48.73
MLGM-2016	52.32
MLGM-2011	59.36
I942	69.97
MLGM-2012	79.08
MLGM-2017	82.22

Chapter 4

Discussion and Future Directions on CHESCA Toolset and QSAR Modeling in the Signaling Field

4.1 Conclusions of the Proposed CHESCA Methods

Two new CHESCA variations were proposed, *i.e*., the T- and CLASS-CHESCAs, to selectively identify critical allosteric sites. Both T- and CLASS-CHESCAs are based on the invariance of pairwise CHESCA correlations to variations in the chemical shifts of two fast-exchanging states (*i.e.,* active, and inactive conformations). The T- and CLASS-CHESCAs together with the more classical CL-CHESCA were implemented for the EPAC1-CNBD. Residues common to the three CHESCA ensembles were found in known EPAC allosteric core sites. These results suggest the proposed CHESCA toolset is effective in prioritizing sites for further targeting through allosteric modulators.

4.2 Conclusions of QSAR Study on EPAC1-Selective Competitive Sulfonamide Inhibitors

A novel QSAR model for a series of EPAC1-specific sulfonamide modulators was developed using the multiple linear regression approach and it showed promising correlation coefficients between the actual and predicted affinities for both the training and test sets. The model was used to predict the affinities of a set of compounds different from the sets used to train the model and based on our QSAR predictions, a new I942 analog denoted as MLGM-2013 was chosen as a promising candidate with a better predicted affinity relative to I942. 8-NBD-cAMP fluorescence competition assays confirmed the QSAR prediction that MLGM-2013 exhibits a

significantly lower K_D value than the parent compound, I942. NMR analyses further investigated the binding mode and mechanism of action of MLGM-2013 and the compound was shown to share a similar binding mode to I942, with significant chemical shift perturbations at PBC residues specifically. Based on the CHESPA analysis, MLGM-2013 was proposed to be more inhibitory compared to I942. We anticipate that the proposed QSAR model will serve as a tool to virtually screen libraries of compounds, which will aid in identifying novel EPAC1-selective drug candidates.

4.3 Future Directions

4.3.1 Applicability of the CHESCA Toolset

The T-, CLASS- and CL-CHESCA offer an effective toolset to prioritize allosteric sites to be further probed through mutations and functional assays. In addition, the allosteric networks identified through such CHESCA toolset may assist in the mechanistic understanding of disease-related mutations as well as aid in the process of developing new effectors targeting allosteric sites. It is also worth noting that the EPAC1-CNBD is a model system for this toolset and so, the parameters used rely on the behaviour of the protein under the influence of imposed conditions, such as temperature variations. Other allosterically regulated proteins may be more/less stable under the same conditions used for EPAC1. Therefore, the selection of temperatures, or the cut-off values used (*e.g.,* the mean Pearson correlation coefficient cut-off) should be fine tuned and adapted to the system of interest.

4.3.2 QSAR Models in the cNMP-signaling Field

While the QSAR model presented in this thesis addresses the affinity predictions of I942 derivatives [1], the model can essentially be extended to other series of EPAC-selective compound, some of which have been extensively studies with the purpose to build structure-activity relationships. For example, recent studies have reported a library of ESI09 [2] derivatives with corresponding inhibitory activities (IC_{50}) against both EPAC1 and EPAC2 [3,4]. The inhibition of EPAC by ESI09 has been shown to suppress the migration and invasion of pancreatic cells [2] as well as shed the light on the roles of EPAC in fatal rickettsioses [5] and T-cell-mediated immunosuppression [6]. QSAR models are anticipated to be useful in the design of more potent EPAC-specific inhibitors based on the promising ESI09 scaffold. Moreover, a further area of EPAC QSAR applications entails EPAC2-specific compounds, which can be both cNMP-like [7] as well as non-cNMP molecules [8].

As a result of the versatility of the QSAR approach, similar models can additionally be applied to the CNBDs of other members of the cNMP-signaling pathways, such as PKA, cyclic-nucleotide-gated ion channels (such as HCN) and PKG, provided enough data on potencies or possibly efficacies are available to train the model [9,10]. The current QSAR approach, which relies on planar structures of molecules, can be further leveraged to include three-dimensional geometric molecular descriptors as well, given the availability of known bioactive conformations for the ligands [11–13].

4.4 References

(1)   Wang, P.; Luchowska-Stańska, U.; van Basten, B.; Chen, H.; Liu, Z.; Wiejak, J.; Whelan, P.; Morgan, D.; Lochhead, E.; Barker, G.; Rehmann, H.; Yarwood, S. J.; Zhou, J. Synthesis and Biochemical Evaluation of Noncyclic Nucleotide Exchange Proteins Directly Activated by CAMP 1 (EPAC1) Regulators. *Journal of Medicinal Chemistry* **2020**, *63* (10), 5159–5184.

(2)   Almahariq, M.; Tsalkova, T.; Mei, F. C.; Chen, H.; Zhou, J.; Sastry, S. K.; Schwede, F.; Cheng, X. A Novel EPAC-Specific Inhibitor Suppresses Pancreatic Cancer Cell Migration and Invasion. *Molecular Pharmacology* **2013**, *83* (1), 122–128.

(3)   Ye, N.; Zhu, Y.; Liu, Z.; Mei, F. C.; Chen, H.; Wang, P.; Cheng, X.; Zhou, J. Identification of Novel 2-(Benzo[d]Isoxazol-3-Yl)-2-Oxo-N-Phenylacetohydrazonoyl Cyanide Analoguesas Potent EPAC Antagonists. *European Journal of Medicinal Chemistry* **2017**, *134*, 62–71.

(4)   Ye, N.; Zhu, Y.; Chen, H.; Liu, Z.; Mei, F. C.; Wild, C.; Chen, H.; Cheng, X.; Zhou, J. Structure-Activity Relationship Studies of Substituted 2-(Isoxazol-3-Yl)-2-Oxo-N′-Phenyl-Acetohydrazonoyl Cyanide Analogues: Identification of Potent Exchange Proteins Directly Activated by CAMP (EPAC) Antagonists. *Journal of Medicinal Chemistry* **2015**, *58* (15), 6033–6047.

(5)   Gong, B.; Shelite, T.; Mei, F. C.; Ha, T.; Hu, Y.; Xu, G.; Chang, Q.; Wakamiya, M.; Ksiazek, T. G.; Boor, P. J.; Bouyer, D. H.; Popov, V. L.; Chen, J.; Walker, D. H.; Cheng, X. Exchange Protein Directly Activated by CAMP Plays a Critical Role in Bacterial

Invasion during Fatal Rickettsioses. *Proceedings of the National Academy of Sciences* **2013**, *110* (48), 19615–19620.

(6)   Almahariq, M.; Mei, F. C.; Wang, H.; Cao, A. T.; Yao, S.; Soong, L.; Sun, J.; Cong, Y.; Chen, J.; Cheng, X. Exchange Protein Directly Activated by CAMP Modulates Regulatory T-Cell-Mediated Immunosuppression HHS Public Access. **2015**, *465* (2), 295–303.

(7)   Schwede, F.; Bertinetti, D.; Langerijs, C. N.; Hadders, M. A.; Wienk, H.; Ellenbroek, J. H.; de Koning, E. J. P.; Bos, J. L.; Herberg, F. W.; Genieser, H.-G.; Janssen, R. A. J.; Rehmann, H. Structure-Guided Design of Selective Epac1 and Epac2 Agonists. *PLOS Biology* **2015**, *13* (1), e1002038.

(8)   Wild, C. T.; Zhu, Y.; Na, Y.; Mei, F.; Ynalvez, M. A.; Chen, H.; Cheng, X.; Zhou, J. Functionalized N,N-Diphenylamines as Potent and Selective EPAC2 Inhibitors. *ACS Medicinal Chemistry Letters* **2016**, *7* (5), 460–464.

(9)   Sassone-Corsi, P. The Cyclic AMP Pathway. *Cold Spring Harbor Perspectives in Biology* **2012**, *4* (12), a011148–a011148.

(10)   Hofmann, F.; Ammendola, A.; Schlossmann, J. Rising behind NO: CGMP-Dependent Protein Kinases. *Journal of Cell Science* **2000**, *113* (10), 1671–1676.

(11)   Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *Journal of the American Chemical Society* **1988**, *110* (18), 5959–5967.

(12)   Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules To Correlate and Predict Their Biological Activity. *J. Med. Chem* **1994**, *37*, 4130–4146.

(13)   Sippl, W. 3D-QSAR – Applications, Recent Advances, and Limitations. *Challenges and Advances in Computational Chemistry and Physics* **2010**, *8*, 103–125.