The Projection Analysis of NMR Chemical Shifts Reveals Extended EPAC Autoinhibition Determinants

Rajeevan Selvaratnam,[†] Bryan VanSchouwen,[†] Federico Fogolari,[‡] Mohammad T. Mazhab-Jafari,[§] Rahul Das,[§] and Giuseppe Melacini[†§]

[†]From the Department of Chemistry and Chemical Biology, [§]and the Department of Biochemistry and Biomedical Sciences,
McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4M1, Canada

[‡]From the Department of Biomedical Science and Technology,
University of Udine, Piazzale Kolbe 4, 33100 Udine, Italy

*Running Title: Determinants of EPAC autoinhibition

To whom correspondence should be addressed: Giuseppe Melacini, Department of Chemistry and Chemical Biology, Department of Biochemistry and Biomedical Sciences, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4M1, Canada. Fax: 905-522-2509; E-mail: melacin@mcmaster.ca

SUPPLMENTARY MATERIAL:

MD Simulation Protocol - All MD simulations were performed using the NAMD 2.7 software (S1) on the Shared Hierarchical Academic Research Computing Network (SHARCNET). The CHARMm27 force field was used for all simulations and the simulations were set up to mimic experimental conditions utilized in previous solution-state studies of EPAC: a pH of 7.6; explicit water (with periodic boundary conditions) with a 50 mM concentration of NaCl; a constant temperature of 34°C (307 K); and a constant external pressure of 1 atm. Protein structure coordinate and parameter files (with hydrogen atoms) for the EPAC2 structures were constructed using the "Psfgen" module of VMD 1.8.6. Parameters for cAMP were constructed from the parameters for the adenine ribonucleotide ("ADE" in the CHARMm force field) by applying the force field's intrinsic "CY35" patch function. Amino acid substitutions present in the simulated EPAC2 mutants were also introduced using the "Psfgen" module, by applying the module's intrinsic "mutate" routine to the affected amino acid residues. In order to mimic a pH of 7.6, hydrogen atoms were added such that all His side chains were in their deionized τ-state and the N-/C-termini and all Asp, Glu, Arg and Lys side chains were in their ionized states. The structures were then immersed in a box of TIP3P water molecules using the Solvate module of VMD 1.8.6, ensuring a minimum distance of 12 Å between the protein and the edge of the solvent box. Salt ions (Na^+ and Cl^-) were added to the solvent box using the Autoionize module of VMD 1.8.6, such that the system was neutralized and the effective NaCl concentration in the solvent was 50 mM.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The analysis was carried out using a Fortran-based software package, in which the PC calculations were performed with SHARCNET's AMD ScaLAPACK library (S2). All analyses

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

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

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

SUPPLEMENTARY REFERENCES

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

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

SUPPLEMENTARY FIGURES

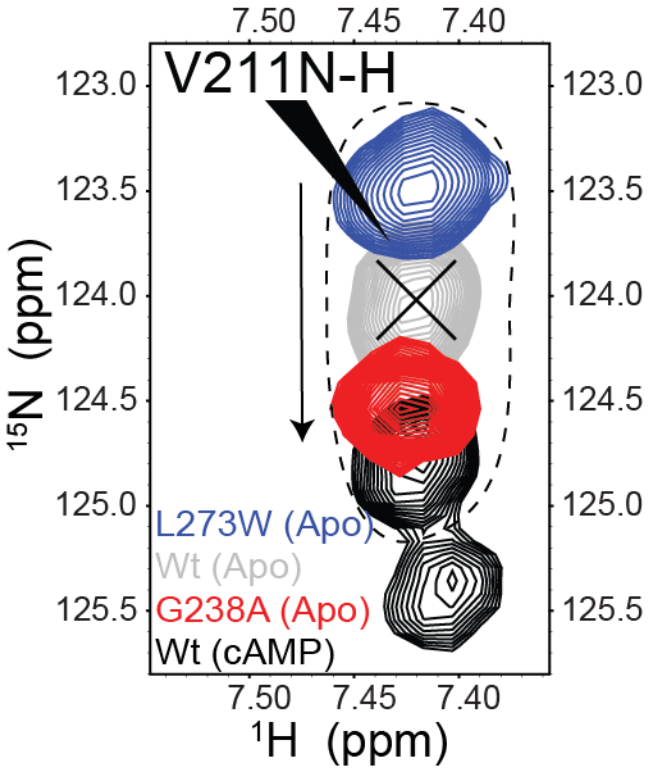

Figure S1: Representative region of the [^{15}N-^{1}H] HSQC spectra of apo-Wt (grey) and cAMP-bound (holo) Wt (black) superimposed with the [^{15}N-^{1}H] HSQC spectra of mutants: apo-L273W (blue) and apo-G238A (red).

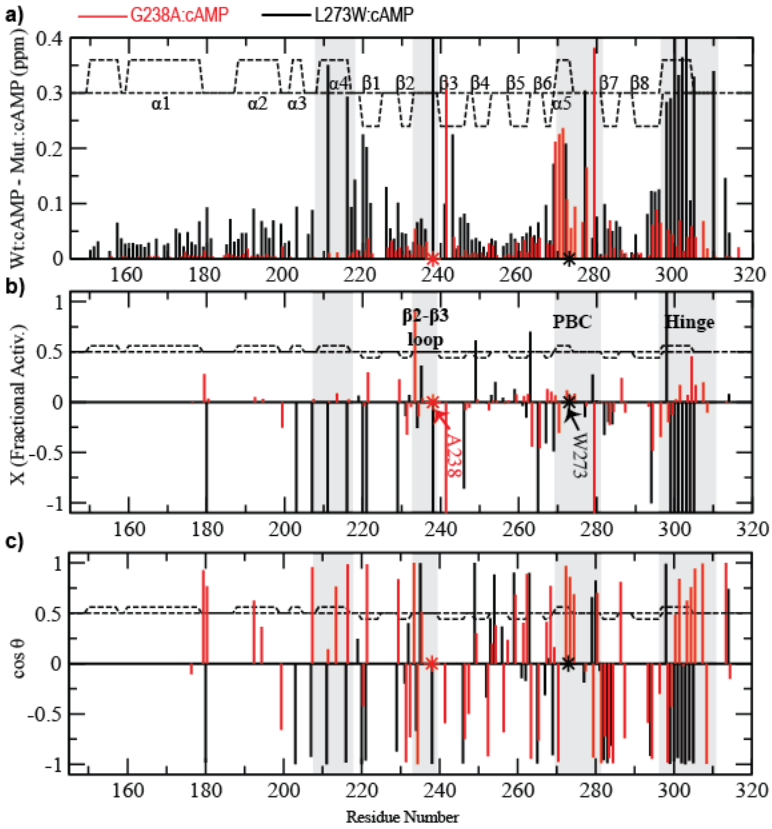

Figure S2: The effects of the L273W (black) and G238A (red) mutations in the presence of cAMP. a) The compounded chemical shift profile of the cAMP-bound mutant relative to cAMP-bound Wt. b) Fractional shift toward activation/inactivation as achieved by the mutation in the presence of cAMP. c) Projection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

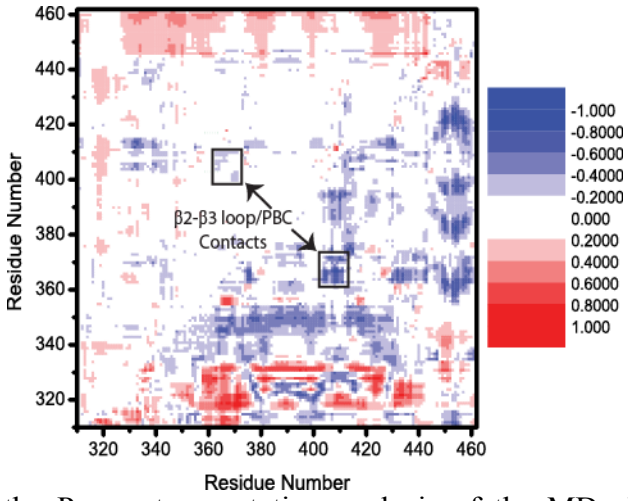

Figure S3: Results of the Procrustean rotation analysis of the MD simulations. The residue numbering refers to EPAC2. The area corresponding to the β2-β3 loop and the PBC is boxed in black lines. Dynamics in these regions are subject to quenching either due to the G238A mutation (upper triangular half)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

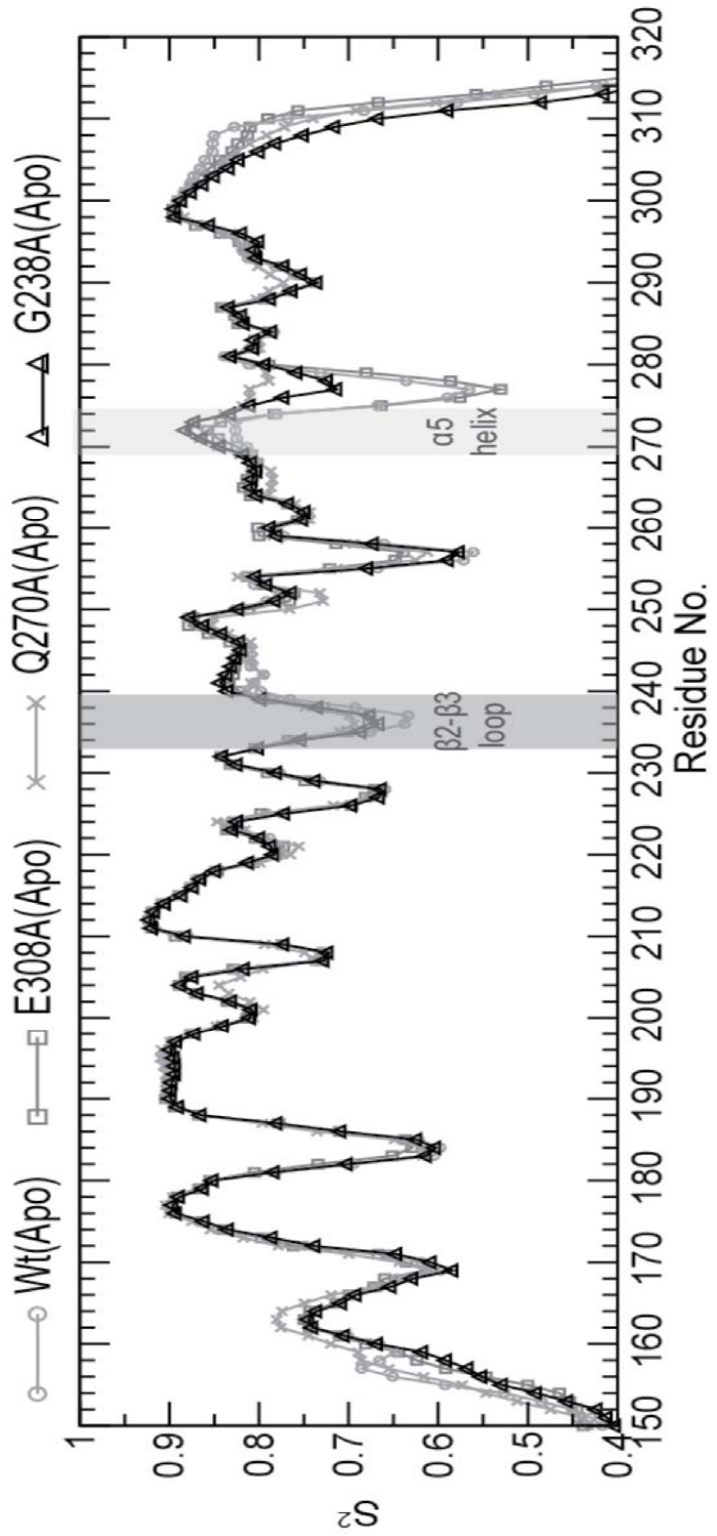

Figure S4: Order parameters (S^2) for the fast (ps-ns) dynamics of the apo-Wt, apo-G238A, apo-Q270A and apo-E308A $EPAC1_{149\text{-}318}$.

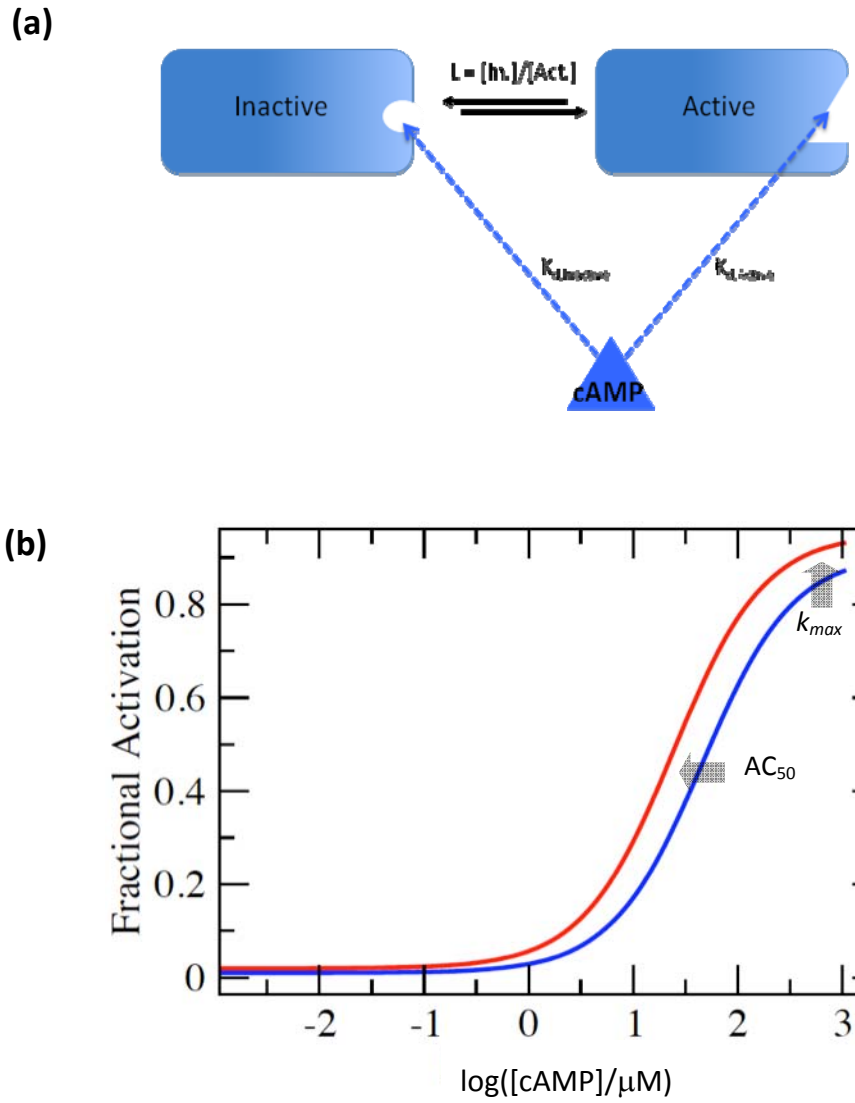

Figure S5: (a) Two-state thermodynamic model for the cAMP-dependent activation of EPAC. This model is defined by three key parameters: L = $[Inactive]_{Apo}/[Active]_{Apo}$, which is the equilibrium constant for the apo/inactive *vs.* apo/active equilibrium (*i.e.* auto-inhibitory equilibrium); $K_{d,Inactive}$ and $K_{d,Active}$, which are the dissociation constants for the binding of cAMP to the inactive and active conformations, respectively. In the case of EPAC, the exact values for L, $K_{d,Inactive}$ and $K_{d,Active}$ are not known. However, for illustrative purposes it is useful to note that setting L = 10^2, $K_{d,Inactive}$ = 500 μM and $K_{d,Active}$ = 0.5 μM results in an activation profile in the experimentally observed [cAMP] range (panel **(b);** blue curve). The activation profiles were computed based on the following equation: Fractional Activation = $1/(1+L_{app})$ with $L_{apparent} = L(1+([cAMP]/K_{d,Inactive}))/(1+([cAMP]/K_{d,Active}))$, where [cAMP] is the concentration of free cAMP. **(b)** Effect of a reduction of L on the activation profile. Simulated activation profile for wt

EPAC (blue) and for a mutant (red) with reduced L value (*e.g.* L = 50 with unchanged $K_{d,Inactive}$ and $K_{d,Active}$). A reduced L models a shift of the apo auto-inhibitory equilibrium towards activation, similarly to what is observed through the NMR projection analysis for the Q270A and the E308A CBD mutants (Figure 7). Overall, a shift of the auto-inhibitory equilibrium towards activation results in decreased AC_{50} and increased k_{max} values (grey arrows), as also observed experimentally for the Q270A and the E308A CBD mutants (Table S2).

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

|B| cut-off: 0.025				|B| cut-off: 0.05				|B| cut-off: 0.1			
Residue	*X*, when |cos (θ)| >0.90	Residue	All *X*	Residue	*X*, when |cos (θ)| >0.90	Residue	All *X*	Residue	*X*, when |cos (θ)| >0.90	Residue	All *X*
295	5.64166	295	5.64166	219	1.41048	219	1.41048	277	0.409072	277	0.409072
219	1.41048	219	1.41048	265	0.923712	265	0.923712	220	-0.317601	252	0.005202
265	0.923712	265	0.923712	277	0.409072	277	0.409072	275	-0.550629	280	-0.000168
247	0.890374	247	0.890374	263	-0.247571	246	0.408751	214	-0.584049	282	-0.022237
236	0.485858	248	0.872726	220	-0.317601	252	0.005202	307	-0.678239	301	-0.156837
277	0.409072	236	0.485858	180	-0.374095	280	-0.000168	211	-0.722264	220	-0.317601
194	0.254042	277	0.409072	213	-0.414849	282	-0.022237	303	-0.734555	299	-0.471642
263	-0.247571	246	0.408751	275	-0.550629	207	-0.085259	216	-0.740106	275	-0.550629
220	-0.317601	289	0.265345	221	-0.580723	301	-0.156837	294	-0.911939	214	-0.584049
180	-0.374095	194	0.254042	214	-0.584049	263	-0.247571	268	-0.973808	307	-0.678239
213	-0.414849	250	0.152262	307	-0.678239	220	-0.317601	305	-0.98144	211	-0.722264
275	-0.550629	293	0.021952	211	-0.722264	180	-0.374095	308	-1.12051	303	-0.734555
221	-0.580723	252	0.005202	303	-0.734555	213	-0.414849			216	-0.740106
214	-0.584049	280	-0.000168	229	-0.739103	299	-0.471642			302	-0.837122
307	-0.678239	282	-0.022237	216	-0.740106	275	-0.550629			294	-0.911939
211	-0.722264	207	-0.085259	294	-0.911939	221	-0.580723			268	-0.973808
303	-0.734555	285	-0.108554	268	-0.973808	214	-0.584049			305	-0.98144
229	-0.739103	301	-0.156837	305	-0.98144	267	-0.585202			308	-1.12051
216	-0.740106	287	-0.177324	308	-1.12051	307	-0.678239				
179	-0.850643	263	-0.247571	231	-2.31733	211	-0.722264				
294	-0.911939	178	-0.309937			303	-0.734555				
268	-0.973808	220	-0.317601			229	-0.739103				
305	-0.98144	174	-0.321848			216	-0.740106				
308	-1.12051	180	-0.374095			249	-0.790444				
243	-1.95458	213	-0.414849			302	-0.837122				
231	-2.31733	299	-0.471642			294	-0.911939				
226	-2.65122	233	-0.477403			304	-0.937142				
296	-2.71308	275	-0.550629			268	-0.973808				
230	-2.89852	176	-0.569376			305	-0.98144				
286	-3.07383	221	-0.580723			308	-1.12051				
		214	-0.584049			231	-2.31733				
		267	-0.585202								
		307	-0.678239								
		211	-0.722264								
		303	-0.734555								
		229	-0.739103								
		216	-0.740106								
		192	-0.744364								
		249	-0.790444								
		302	-0.837122								
		179	-0.850643								
		294	-0.911939								
		304	-0.937142								
		268	-0.973808								
		305	-0.98144								
		308	-1.12051								
		237	-1.8461								
		243	-1.95458								
		262	-2.03197								
		231	-2.31733								
		226	-2.65122								
		296	-2.71308								
		230	-2.89852								
		286	-3.07383								

Table S2. Functional Bioassay Data for Q270A and E308A EPAC mutants from ref. (27).		
EPAC Construct	AC_{50} / μM	*Relative* k_{max}
Wt	50	1
Q270A	40	1.7
E308A	15	3.0