Computational analysis of long-range allosteric communications in CFTR

Ersoy Ayca[1,2], Altintel Bengi [1,2*], Livnat Levanon Nurit[3*], Ben-Tal Nir[4], Haliloglu Turkan,[1,2], and Lewinson Oded[3].*

[1] Department of Chemical Engineering, Bogazici University, 34342 Istanbul, Turkey

[2] Polymer Research Center, Bogazici University, 34342 Istanbul, Turkey

[3] Department of Molecular Microbiology, Bruce and Ruth Rappaport Faculty of Medicine, Technion-Israel Institute of Technology

[4] Department of Biochemistry and Molecular Biology, Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv, Israel

* Equal contribution

Classification Physical science

Keywords Cystic fibrosis, CFTR, ABC transporter, allosteric regulation, conformational dynamics.

Abstract. Malfunction of the CFTR protein results in cystic fibrosis, one of the most common hereditary diseases. CFTR functions as an anion channel, the gating of which is controlled by long-range allosteric communications. Allostery also has direct bearings on CF treatment: the most effective CFTR drugs modulate its activity allosterically. Herein, we integrated Gaussian Network Model, Transfer Entropy, and Anisotropic Normal Mode-Langevin dynamics and investigated the allosteric communications network of CFTR. The results are in remarkable agreement with experimental observations and mutational analysis and provide extensive novel insight. We identified residues that serve as pivotal allosteric sources and transducers, many of which correspond to disease causing mutations. We find that in the ATP-free form, dynamic fluctuations of the residues that comprise the ATP binding sites facilitate the initial binding of the nucleotide. Subsequent binding of ATP then brings to the fore and focuses dynamic fluctuations that were present in a latent and diffuse form in the absence of ATP. We demonstrate that drugs that potentiate CFTR's conductance do so not by directly acting on the gating residues, but rather by mimicking the allosteric signal sent by the ATP binding sites. We have also uncovered a previously undiscovered allosteric "hotspot" located proximal to the docking site of the phosphorylated Regulatory (R) domain, thereby establishing a molecular foundation for its phosphorylation-dependent excitatory role. This study unveils the molecular underpinnings of allosteric connectivity within CFTR and highlights a novel allosteric "hotspot" that could serve as a promising target for the development of novel therapeutic interventions.

Introduction. Cystic fibrosis (CF) is a multifaceted disorder impacting the epithelial tissues of various organs, including the lungs, intestinal tract, and pancreas. It stands as one of the prevailing hereditary afflictions, affecting approximately one in every 3,000 newborns(1). Notwithstanding the significant advancements in therapeutic interventions, individuals with cystic fibrosis (CF) continue to endure pronounced morbidity and a reduced life expectancy. (2). The disease is caused by mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR) protein. Most CF patients carry a deletion of Phe 508 in NBD1 which leads to protein instability and premature degradation. More than 2000 other disease-associated mutations have been documented (3), many of which remain uncharacterized. CFTR functions as a chloride channel, the excretion of which is important for mucus hydration. Malfunction of CFTR leads to mucus dehydration and a subsequent decrease in mucus clearance. The over-accumulation of mucus leads to persistent bacterial infections, chronic inflammations, and to organ failure (4). CFTR is a member of the ABC transporter protein family, considered to be one of the largest protein families of any proteome (5). CFTR is a unique family member, as all other ABC transporters function as active pumps, moving biomolecules against their concentration gradients (6,7). The domain organization of CFTR is similar to that of canonical ABC transporters, comprising two transmembrane domains (TMDs) and two cytosolic nucleotide binding domains (NBDs). The TMDs form the translocation channel and the NBDs control channel opening and closing through processive cycles of ATP binding, hydrolysis, and release. A unique feature of CFTR is the regulatory domain (R) which contains multiple phosphorylation sites and connects the two NBDs giving rise to a TMD1–NBD1–R–TMD2–NBD2 topology. Decades of structure-function studies enabled the formalization of a mechanistic model for the function of CFTR: In the resting (non-conducting) state, the funnel-shaped pore is inward-facing and is solvent accessible only to the cytoplasm. The NBDs are separated, and the R domain is wedged between them, preventing their dimerization (8,9). Upon phosphorylation of the R domain by PKA, it disengages from its inhibitory position and shifts aside (10,11). The NBDs are now free to bind ATP and form a ‘‘head-to-tail’’ dimer (6,12,13), with the two ATP molecules sandwiched at composite binding sites formed at the dimer interface. In CFTR, like in some other ABC transporters, only one of the ATP sites is hydrolysis-competent while the other (termed the degenerate site) can only bind ATP but not hydrolyze it (14–16). Formation of the NBDs dimer leads to rearrangements of transmembrane gating residues and subsequent channel opening.

Hydrolysis of ATP at the catalytic site and release of hydrolysis products leads to opening of the NBDs, channel closure, and resetting of the system. This mechanism bears resemblance to that of *bona fide* ABC transporters, where ATP binding and hydrolysis at the NBDs drive the conversion of the TMDs between inward- and outward-facing conformations (17–19).

It is now broadly recognized that most proteins and enzymes are allosteric, where ligand binding at one site affects the functional properties of another (often distant) site (20–23). Two conceptual models dominated the field of allostery for the past 50 years. According to Koshland, Nemethy, and Filmer (KNF model, (24)), the protein exists in a single conformation (an inactive one) to which the allosteric ligand binds, inducing a conformational change that leads to the active conformation. In contrast, Monod, Wyman, and Changeux (MWC model, (25)) introduced the concept of "conformational selection", according to which proteins pre-exist in multiple conformation, and the allosteric ligand preferentially binds to the active conformer, leading to a redistribution of the conformational distributions towards the active one (26). The simplicity of these models is appealing, yet we now begin to understand that they alone cannot explain the complexity of allosteric signal generation and transduction. More recently, an alternative allosteric mechanism was demonstrated, termed "dynamic mechanism" (27,28), where allostery is mediated by changes in dynamic fluctuations around a mean conformation in the absence of a noticeable conformational change (29,30). In some proteins, as we believe is the case of CFTR, multiple allosteric mechanisms combine to generate the observed allosteric phenomenon.

There are abundant examples that long-range allosteric communications are essential to the function of CFTR. One such example is the control of channel opening and closing, where the binding site of the effector molecule (ATP) and the affected gating residues are separated by 60-70Å (8,12). Very recently, Levring *et al.*, demonstrated that ATP-dependent dimerization of the NBDs is insufficient for channel opening which additionally depends on an allosteric pathway that connects the NBDs and the channel pore (31). Based on rate-equilibrium free-energy relationship analysis Sorum *et al.*, reached similar conclusions, and suggested that this allosteric signal originates at the NBDs and gradually propagates towards the gating residues along a clear longitudinal protein axis (32). Another example of long-range allostery in CFTR is the finding that point mutations at the cytosolic loops increase channel open probability (even in the absence of ATP) by shifting a pre-existing conformational equilibrium towards the ligand bound conducting conformation (33). Notably, allostery in CFTR has direct therapeutic relevance: The potentiator

drug ivacaftor and the dual-function modulator elexacaftor do not bind in the proximity of the gating residues (34,35) yet increase the open probability (Po) of the channel (36,37). These and other studies (38–40) significantly advanced our understanding of the function of CFTR and its allosteric control.

However, some important features of the underlying allosteric network remain unknown. For example, the complete allosteric map of CFTR remains to be determined, as are the causal relations (driving vs. driven) between residues and domains. Similarly, the allosteric mechanism for modulation of CFTR by commercial drugs is only partially understood. Herein, we sought to further our understanding of the allosteric network of CFTR. For this, we combined two complementary computational approaches: Gaussian network model-Transfer Entropy (GNM-TE (41,42)), and ANM-LD simulations which combine Anisotropic Network Model (ANM) (43–46) and Langevin Dynamics (LD) (47).

GNM reduces the complexity of a protein polymer to a network of nodes (representing the main chain α carbons) connected by springs with a spring constant of unity. Further simplifications are the assumptions of force linearity and isotropy of all fluctuations. These simplifications are the source of both its advantages and limitations. On the one hand, GNM is computationally economical, allowing for rapid analytical derivation of conformational dynamics even of large proteins. This computational economy, while offering valuable insights, does entail a limitation. GNM proves highly effective in predicting fluctuations around a predefined native state and offers valuable information regarding allosteric communication and causality. Nevertheless, due to its reliance on topology, GNM is unsuitable to monitor the conformational transitions from one native state to another. However, it does have the capacity to suggest potential modes of motion accessible within a given protein topology that may contribute to such conformational changes. In addition, GNM cannot provide any information on side chains dynamics as these are omitted from the model. The robustness of the model makes GNM a highly effective tool in studying global dynamics. Despite, or because of its simplicity GNM based approaches have been very successful in describing global conformational dynamics of proteins (48–50). More details on the fundamentals of GNM, its advantages and limitations can be found here (51). To supplement GNM we integrated it with Transfer Entropy calculations. The concept of Transfer Entropy (TE) is derived from information theory and is used to measure the amount of information transferred

between two processes (41). In the context of proteins' conformational changes, TE can be used to determine whether two residues are allosterically connected (42). TE for each pair of *i* and *j* residues can be viewed as the degree to which the present movement of residue *i* decreases the uncertainty regarding the future movement of residue *j* within a specified time delay τ. If TEij (τ) > TEji (τ), then the dynamics of residue *i* affects the dynamics of residue *j*, representing a causal relationship between residues *i* and *j*. If a residue transfers information to many other residues in the protein, it is considered as an information source; if it accepts information from much of the protein, it is considered as an information receiver or a 'sink' (for more details about the correlation between TE and allostery see (52)). In the present work, we performed TE calculations based on Gaussian Network Model (GNM), which considers harmonic interactions between residue pairs (41,50). Unlike simulation trajectory approaches that infer cause-and-effect relations from the observed sequence of events (which may be circumstantial), GNM-based TE provides a direct measure, at the linear approximation, of causality relations between residues and domains. Quantitative inference of cause-and-effect relations from simulation trajectory approaches is in principle possible using enhanced kinetic sampling techniques but is highly computationally demanding for a large transmembrane transporter such as CFTR. Here we opted to use a complimentary GNM-based TE approach that readily provides a quantitative measure of sources and receivers of allosteric signals.

To complement GNM, we used a hybrid molecular simulations approach that integrates global elastic modes with atomic-scale information derived from all-atom simulations. Several such approaches have emerged in recent years (*e.g.*, collective MD (CoMD), Perturb-Scan-Pull, eBDIMS, H-REMD) (53–56) with the aim of finding a "sweet spot" between the precision of whole atom simulations and the computational efficacy of coarse-grained elastic networks. The novelty of our approach, ANM-LD, is that it integrates a low-resolution α-carbon-based anisotropic network model with stochastic all-atom implicit solvent Langevin dynamics (LD) simulations (57). ANM-LD reflects only internal dynamic modes, free of any external bias. More importantly, ANM-LD provides experimentally testable predictions, and has been successfully used to compute conformational trajectories of proteins (50,58).

By combining ANM-LD and GNM-TE we present here the allosteric network of CFTR. The results align very well with experimental and clinical data, disclose the allosteric mechanism

of CFTR, and identify a novel allosteric hotspot that may be used for therapeutic intervention.

Results

Transfer Entropy calculations identify allosteric hotspots in CFTR.

First, we performed GNM-TE calculations on non-phosphorylated ATP-free human CFTR (PDB ID 5UAK, (8)) using the ten slowest GNM "modes". As explained in the methods section and elsewhere (50,59), GNM decomposes a complex motion (such as a conformational change, or a person walking) to discrete dynamic building blocks, historically termed GNM "modes", the ensemble of which comprise the complex motion. These include modes of motion that involve large parts of the protein and lead to major conformational changes (termed "slow modes"). The "slow modes" often display the greatest degree of collective movements. Modes that involve local conformational adjustments and a smaller number of residues and display fewer collective movements are considered "fast modes". We expected that the ten slowest modes, those comprising the largest of motions, will be the main allosteric effectors, and their ensemble will suffice to capture the overall dynamics of CFTR. Indeed, Eigenvalue analysis demonstrated that the ten slowest modes of CFTR encompass its complete dynamic spectrum (Figure S1).

Figure 1 shows the residues identified as the main allosteric sources in ATP-free human CFTR, obtained by using an interaction cutoff distance (R_{cut}) of 10Å. Very similar results were obtained when R_{cut} of 7Å was used instead (Figure S2), suggesting that the results are not very sensitive to this parameter. The residues that GNM-TE identified as main allosteric determinants showed partial overlap with the hinge residues identified by GNM: Of the 60 residues that serve as main allosteric sources, 25 (~41%) were located within 4Å of a hinge residue, and this co-localization was highly statistically significant (Figure S3). This is perhaps not unexpected, as hinge residues likely play an important role in transmitting the strain energy induced by binding of allosteric ligands.

As GNM-based TE has not been extensively used to study membrane proteins as large and complex as CFTR we began by evaluating its utility in extracting meaningful functional information. Allosteric residues often coincide with functional epitopes, *i.e.,* residues that are involved in ligand binding, catalysis, or serve as conformational hinges (21,60). Therefore, we first evaluated the performance of GNM-based TE by examining its ability to identify such sites of functional importance. For this, we first compared the locations of the GNM-TE identified

allosteric sources to the positions of 14 residues that have been experimentally demonstrated to be directly responsible for the function of CFTR. This subset of 14 residues includes the most common CFTR gating mutations (G178, S549, G551, G970, G1244, S1251, S1255, G1349), residues of the intracellular loops that are essential for conformational changes (M265, N268, Y275, W1063, L1065), and the NBD2 Walker B aspartate D1370 which is essential for ATP hydrolysis (8,15,32,38). As shown in Figure 1 we observed that these 14 residues tightly colocalize with the main allosteric peaks. To assess the statistical significance of this co-segregation we compared it to a random identification of residues of functional importance, as follows: GNM-TE identified 60 residues (out of 1480) that are the main allosteric sources (*i.e.,* allosteric peaks) in dephosphorylated ATP-free CFTR (Figure 1). Accordingly, we generated 100,000 sets of 60 randomized positions and counted how many times the positions of the functionally essential residues were correctly predicted using this random allocation. As shown in Figure S4A, the random allocation has a mean value of 0.7 correct guesses while the GNM-TE based identification has a mean value of 8 ($p=9\times e^{-13}$). Next, we expanded this analysis to include not only exact matches (between the random or GNM-TE assignment of functional sites and their real location) but also first- or second- coordination sphere interactions (*i.e.,* peaks that are located <4Å or <7Å from the functionally essential residues). Using these cutoffs, we also observed a highly significant correlation between the location of the GNM-TE peaks and that of the functionally essential residues (Figure S4B-C). We then compiled a second set of reference residues, one which includes all the residues that provide the first ATP coordination sphere in ATP-bound CFTR (12). We reasoned that since CFTR gating is controlled by ATP binding and hydrolysis these residues (that are allosteric by definition) should be readily identified by the GNM-TE analysis. Indeed, as shown in Figure 1 the ATP residues also closely segregated to the allosteric peaks identified by GNM-TE, and this co-segregation outperformed the one obtained via a random allocation of allosteric peaks (Figure S5A). This co-segregation was observed whether we considered only exact matches (Figure S5A), or whether we included also first- or second- coordination sphere interactions (Figures S5B and S5C, respectively). Next, we considered a more challenging reference set of residues, one which includes the 93 CFTR missense point mutations that lead to cystic fibrosis (3). This reference set of residues holds some inherent limitations: many of the mutations that lead to CF have no direct functional role *per se*, rather, they cause CF due to compromised mRNA synthesis/stability, protein processing, or protein targeting (61). The inclusion of such residues in

the benchmark dataset will lead to apparent false-negative identifications of the GNM-TE analysis. In addition, it is likely that not all the allosteric residues have been naturally mutated, leading to false-positive identifications. Nevertheless, despite these limitations, we expected to see some degree of colocalization between the identified allosteric hotspots and the disease-causing mutations. From the 93 mutations listed in the cfrt2 database (3) we removed the cluster of seven N-terminal residues (S13, L15, W60, E56, E60, P67, G85) that are known to be involved in interactions with cellular filamins or ER retention (62). As shown in Figure S6A we observed considerable association between the positions of 86 disease-causing mutations and the allosteric peaks identified by GNM-TE. To test the statistical significance of this association we generated 100,000 sets of randomly positioned "allosteric peaks". As shown in Figure S6B, the random distribution is roughly normal with a mean of 4.5 incidents of exact matches. As shown in Figure S6B the GNM-TE analysis is significantly different from the random distribution, with a mean of 9 exact matches. As observed with the previous two reference set of residues (Figures S4 and S5) the co-segregation between the allosteric peaks identified by GNM-TE and the positions of the disease-causing mutations was also observed when we included also first- and second-coordination sphere interactions (Figures S6C and S6D, respectively). In further support of the biological relevance of the GNM-TE calculations, and despite the structural similarity between the two NBDs, a clear asymmetry can be observed between them, with NBD2 which harbors the catalytic Walker B glutamate providing the higher allosteric signaling (Figures 1 and 2A). This observed asymmetry is in line with a recent single molecule conformational study where the degenerate and catalytic ATP sites were shown to asymmetrically affect NBD dimerization and gating (31). That GNM-TE identifies this asymmetry is quite remarkable, as it is based solely on the positions of the α-carbons and does not consider the side chains. Taken together, these results (Figures 1, 2A, S2, S4, S5, and S6) suggest that GNM-based TE can provide meaningful functional information on CFTR.

As shown in Figures 1 and 2A, the entropy sources residues (*i.e.*, residues that drive allostery) are found predominantly on the NBDs and are largely absent from the TMDs. The only TMDs residues that serve as entropic sources are those found at or near intracellular loops 2 and 3 (Figure 2A). Notably, these observations were made for de-phosphorylated ATP-free CFTR. This demonstrates that even when the NBDs are completely separated and the composite ATP binding sites are yet to form, the dynamic infrastructure for nucleotide binding is already present, poised

for binding of the ligand. This pre-existing dynamic infrastructure likely facilitates ATP binding and will come into full play at a later stage of the conformational change, following phosphorylation and binding of ATP. Interestingly, some of the main entropy peaks (*i.e.*, entropy sources) located in the NBDs are slightly shifted from the ATP binding residues (Figure 1) and correspond to adjacent residues. This shift disappears once ATP is bound, and the entropy source role is transferred from the neighboring residues to the ATP binding residues themselves (see next results section dealing with the effects of ATP binding).

The results shown in Figure 2A were obtained using the 10 slowest GNM modes. However, during the calculations the entropic contribution of the slowest modes may overshadow that of the faster (less collective) ones, which may nevertheless have important functional roles (for more details see (42)). To circumvent this problem, one may repeat the calculations while omitting the slowest GNM modes, enabling the detection of latent allosteric interactions that may otherwise be hidden in the global fluctuations. For ATP-free human CFTR, the greatest effect was observed upon removal of the first and second slowest modes: Upon removal of the slowest GNM mode, a new cluster of residues that serve as allosteric sources was revealed (Figure 2B). These residues, which are distributed roughly symmetrically in the TMDs, include the putative gating residues (*e.g.*, P99, T338, S341, (8)) and additional residues that line the ion permeation pathway. The residues that form the ivacaftor binding site (34) were also revealed as entropy sources by removal of the slowest GNM mode (Figure 2B), demonstrating the utility of this approach. Removal of the next slowest GNM mode reveals another cluster of TMD residues that serve as entropy/allostery sources. Unlike the above-described symmetric cluster, this cluster is highly asymmetric, and is comprised solely of residues in TMD1 (Figure 2C). Interestingly, this cluster of residues is located at and just above the future site for docking of the regulatory domain ((63,64) and see also later in Figure 3C-D).

The above analysis revealed two latent entropy hotspots: one surrounding the ion permeation pathway (Figure 2B) and the other adjacent to the docking site of the regulatory domain (Figure 2C). Both sites were hidden by the more dominant global fluctuations of the protein (Figure 2A). Their identification in this conformation, which is non-conducting and where the R domain is still wedged between the NBDs (8), highlights once again the pre-existence of a dynamic infrastructure which will assume a central role following phosphorylation and repositioning of the R domain.

ATP binding and phosphorylation rewires and focuses the allosteric connectivity of CFTR

To study the effects of ATP binding and phosphorylation we conducted Transfer Entropy analysis of phosphorylated ATP-bound CFTR (PDB ID 6MSM, (12)). We observed that phosphorylation and binding of ATP greatly rewires the overall allosteric connectivity in CFTR (Figure S7 and compare Figures 1 and 3A): residues that serve as entropy sources are now found in all the domains (TMD1, NBD1, TMD2, NBD2) and the entropy peaks are sharp, sharper than those observed for the unphosphorylated ATP-free form (compare Figures 1 and 3A), indicating that information is originating from distinct residues/cluster of residues. This suggests that phosphorylation and binding of ATP focuses the dynamics of CFTR, which pre-existed in a more diffuse manner in the unphosphorylated ATP-free state. This observation suggests that this conformation is the preferred template/target for structure-guided development of allosteric modulators.

As discussed above and shown in Figure 1, in the ATP-free state the entropy peaks at the NBDs are slightly shifted from the ATP binding residues. This offset largely disappears upon binding of ATP, and the entropy peaks in the NBDs now closely match the location of the ATP binding residues (compare Figures 1 and 3A). In the TMDs, the residues that serve as main entropy sources are those that surround the ion permeation channel (Figure 3B). This means that upon phosphorylation and binding of ATP the gating and pore lining residues assume a more dominant role in allosteric signaling. Notably, these transmitting TMD residues were identified as latent entropy sources in the unphosphorylated ATP-free state (Figure 2B). This result again highlights that the underlying dynamics in CFTR pre-exist, prior to phosphorylation and ATP binding. To test if the effect of ATP binging is specific to human CFTR and/or is a result of the specific experimental conditions in which its ATP-free and ATP-bound structures were determined (8,12) we conducted similar calculations with the corresponding conformations of zebrafish CFTR (9,65). As shown in Figure S8, binding of ATP had a very similar effect in both CFTR homologues, suggesting that the allosteric effects of ATP binding are robust and conserved.

As explained above (see analysis of the ATP-free conformation), to identify latent entropy sources we repeated the TE calculations for the phosphorylated ATP-bound state after removing the two slowest modes. Several striking features were revealed by these calculations (Figure 3C). First, the entropy sources in the NBDs have now migrated from the ATP binding sites at the dimer interface towards the intracellular loops, perfectly positioned to transmit the signal to the TMDs.

Second, the TMDs allosteric peaks surrounding the gating residues and the ion permeation pathway have migrated towards the extracellular exit, as if to prepare for ion exit. Lastly, a cluster of residues that serve as strong entropy sources appeared in the TMDs in an asymmetric manner (Figures 3C-D). These residues are found exclusively in TMD1 and comprise or are adjacent to the site of interaction of the regulatory (R) domain (63,64). This observation suggests that the role of the R domain may not be limited to steric obstruction of NBD dimer formation in the unphosphorylated ATP-free state. The presence of this cluster of information-transmitting residues in the vicinity of the docking site of the R domain may explain previous observations that deletion of the R domain leads to fewer channel opening events, and that exogenous addition of the phosphorylated R domain peptide increases channel open probability of CFTR construct deleted of the R domain (66,67).

Concerted movements in CFTR. The Transfer Entropy (TE) analysis presented above provides a measure of information transfer between residues and domains. However, it gives no information of the physical nature of the conformational changes that are involved. To complement the GNM-TE calculations, we used ANM-LD, a molecular simulations approach which computes the trajectory of a conformational change between two known states of a protein (43,47). We simulated the transition between non-phosphorylated ATP-free and phosphorylated ATP-bound human CFTR (8,12). To understand the allosteric connectivity that underlies this conformational change, we analyzed the degree of correlated movements between all CFTR residues. At the basis of this analysis is the assumption that residues that are allosterically connected will move in synchrony in both space and time. An extreme example of such synchronization is the perfect correlation of each residue with itself (diagonal in Figure 4A). Similarly, two residues that are sequential in the amino acid sequence will also display high correlation. However, when highly synchronized movements are observed for residues or domains that are distant in sequence and in 3-D space, this likely represents allosteric connectivity.

The 2-D cross correlation map obtained for the transition between non-phosphorylated ATP-free and phosphorylated ATP-bound human CFTR is shown in Figure 4A. In this map red colors indicate high positive correlations and blue colors high negative correlations. Red colors thus indicate residues that move in parallel vectors in space and time, and blue colors indicate residues that move in anti-parallel vectors in space and time. The map is dominated by large and

continuous patches of red (positive correlation) and blue (negative correlation) areas and is very similar to the 2-D cross correlation map calculated for PglK (50), a lipid-linked oligosaccharide flippase that adopts the canonical ABC exporter 3-D fold (68). These observations suggest that, as inferred from the cryo-EM structures of CFTR (8,12), global rigid body motions with a relatively small number of hinges underlie this conformational change, and that the conformational changes of CFTR are similar to those of canonical ABC exporters. The two NBDs display asymmetric allosteric connectivity: NBD1 shows a tighter correlation with the residues that surround the permeation pathway (Figure 4B), while NBD2 (harboring the catalytic Walker B glutamate) is more strongly correlated with ECL1 (extra cellular loop 1) (Figure 4C) which is known to stabilize the conducting state of the channel (69).

Sequence of allosteric transduction in CFTR. The ANM-LD simulations offer an opportunity to examine mechanistic features that are difficult to determine experimentally. One such feature is the pre-equilibrium kinetics of allosteric interactions: if residues A and B are allosterically connected, does the conformational change of A follow that of B, or *vice versa*? To investigate the sequence of allosteric transduction in CFTR we re-analyzed the simulations while imposing a time-delay between the conformational changes of allosterically coupled residues (or domains) of interest. For any two allosterically coupled residues, if we define that the conformational change in A occurs before that of B, and still observe a positive correlation in the 2-D correlations map, this means that the conformational change of A precedes that of B. If the correlation is lost, this means that B likely precedes A, which can be confirmed by imposing the opposite time-delay (*i.e.*, imposing that the change in B precedes) and verifying that the positive correlation is again observed (see *Methods* for a mathematical expression of these relations). This approach can be used to track the entire route of allosteric signal transduction as detailed below.

Rate-equilibrium free-energy relationship analysis and single molecule FRET studies both suggest that the allosteric signal originates at the NBDs and propagates to the TMDs (32,70). Our results agree with these observations: in parallel simulations we repeatedly observed that the conformational change from the dephosphorylated ATP-free state to the phosphorylated ATP-bound one begins with the movement of the NBDs. In addition, the TE analysis of the inward-facing ATP-free state demonstrated that in this conformation the allosteric peaks are located almost exclusively at the NBDs (Figure 1). Of the two NBDs, the catalytic NBD2 seems to provide the

greater allosteric input (Figures 2A and S7A) and therefore as detailed below we began tracing the allosteric signal at NBD2. As shown above (Figure 4C), in the absence of a time delay the movements of NBD2, ICL2/3, TMD2 (including TM8 up to its hinge) and ECL1 are highly synchronized. However, this correlation persists only if NBD2 leads this motion and ICL2/3, TMD2/TM8 and ECL1 follow (Figure S9 green rectangles, Figure 5 step I). This observation strongly suggests that the movement of NBD2 is the leading event and those of ICL2/3, TMD2/TM8 and ECL1 follow, and that the allosteric signal propagates from NBD2 to ICL2/3, TMD2/TM8, and ECL1 (Figure 5 center, green arrows). Notably, based on their recent single molecule studies Levring *et al.*, proposed similar cause and effect relations between the catalytic ATP site and TM8 (31).

From ECL1 the allosteric signal propagates to NBD1: The movements of these two domains are highly correlated, but only if ECL1 leads and NBD1 follows (Figure S9 magenta rectangles, Figure 5 step II, and Figure 5 center, magenta arrow). Similar considerations suggest that from NBD1 the allosteric signal propagates to the gating residues and to the TM helices that surround the permeation pathway: The movements of these domains are only synchronized if the former leads and the latter follows (Figure S9 blue rectangles, Figure 5 step III, and Figure 5 center, blue arrow). In turn, the movements of the gating residues and the permeation pathway TM helices are correlated with those of NBD2, but only if the TM helices lead and NBD2 follows (Figure S9 orange rectangles, Figure 5 Step IV). We therefore conclude that the allosteric signal propagates from the permeation pathway to NBD2, thus completing the cycle of allosteric transduction (Figure 5 center, orange arrow).

Modulation of CFTR by drugs. In the past decade, the development of small-molecule CFTR modulators revolutionized cystic fibrosis treatment and led to a dramatic reduction in patient morbidity and increased life expectancy (71). There are currently four FDA-approved drugs that are used to treat CF: The type I correctors lumacaftor (VX-809) and tezacaftor (VX-661) stabilize mutant CFTR variants (predominantly the most common ΔF508 mutation) reducing premature protein degradation and improving surface presentation of the mature channel (72,73). The potentiator ivacaftor (VX-770) enhances the activity of conductance-defective mutants by increasing the open probability (Po) of the channel (31,34,36), and the dual-function type III corrector elexacaftor (VX-445) stabilizes unstable variants and increases channel open probability

(37,74). The recently published structures of WT and ΔF508CFTR bound to various drugs (35) enabled to us to investigate the allosteric effects of the drugs. First, we conducted GNM-TE analysis on ΔF508CFTR bound to Trikafta (PDB ID 8EIQ, (35)), a trivalent drug which combines ivacaftor, tezacaftor, and elexacaftor and is currently the most advanced cystic fibrosis treatment. As shown in Figure 6A, the type I correctors lumacaftor and tezacaftor bind at 'valleys', *i.e.,* to residues that are not allosteric sources. These results are consistent with the mechanism of these drugs which improve the function of CFTR by lowering its $\Delta G_{(folding)}$ rather than by modulating its activity (73). In contrast, the residues that comprise the binding sites for ivacaftor and elexacaftor also include ones that serve as main allosteric sources (Figure 6A), alluding to their role as allosteric modulators.

In CFTR, channel opening depends on phosphorylation of the regulatory domain and subsequent binding of ATP by the NBDs. Accordingly, mutations that interfere with binding of ATP reduce channel opening probability and result in cystic fibrosis (CF). The potentiator drug ivacaftor is used to treat CF patients who carry mutations in the NBDs (most commonly the G551D mutation at the highly conserved LSGGQ ATP binding/signature motif) (34,36,75,76). Ivacaftor increases channel opening probability despite impaired ATP binding and hydrolysis. The molecular basis for the beneficial effects of ivacaftor is not fully understood, especially since it does not bind in the vicinity of either the ATP- or ion-coordinating residues but rather peripherally at the protein-lipid interface (34). In addition, unlike the folding correcting drugs, ivacaftor does not induce a significant conformational change in ΔF508 CFTR (35). To better understand the mechanism underlying the CFTR's modulation by ivacaftor we mapped the residues that receive information from the ivacaftor binding site. Surprisingly, the gating residues and those that line the permeation pathway were not among the residues that receive information from the drug binding site (Figure 6B), leaving the molecular basis of the drug's action unclear. Thus, in attempt to resolve this issue, we compared the profile of the residues that receive information from the ivacaftor binding site to those that receive information from the ATP binding sites. As shown in Figure 6C, the two profiles showed remarkable overlap: The ivacaftor docking site and the ATP binding site send information to a very similar set of residues (Pearson's correlation coefficient of 0.83). These observations suggest that ivacaftor increases the opening probability not by directly affecting the gating residues but rather indirectly by mimicking the allosteric signaling evoked by ATP binding. Further insight for the mechanism of action of ivacaftor can be obtained from the structural representation of the

allosteric signaling shown in Figure 6D-F. The gating region (*e.g.*, gating residue F337) (12)) sends information mostly to the extracellular side of TM helices 2 and 11 and to the cytoplasmic side of helices 3, 4, 7 including intracellular loops 2 and 3 (Figure 6D). In comparison, G551 of the ATP binding signature motif (mutations in which are treated with ivacaftor) sends information predominantly to the intracellular side of TM helices 3, 4, 7, 8 and to the C-terminal intracellular helix of NBD2 (Figure 6E). Superimposition of these effects, which represents the combined allosteric signaling of the ion-coordinating and ATP-binding residues (shown in Figure 6F) is very similar to the allosteric signaling sent by the ivacaftor binding residue F312 (Figure 6B). This result suggests that ivacaftor exerts its potentiating effect by mimicking the combined allosteric output of the ATP-binding and gating residues, compensating for any potential allosteric transduction short circuit cause by mutations.

Discussion – the allosteric mechanism of CFTR

Allostery in proteins is most often studied through the lenses of the classical KNF or MWC models, and more recently the dynamic mechanism (see introduction and (25,27,28,77)). The results presented herein suggest that none of these models can singly explain the allosteric mechanism of CFTR.

In the absence of phosphorylation and ATP, CFTR adopts solely the NBDs-open conformation and spontaneous conversions to the NBDs-closed conformation are not seen (70). In addition, multiple phosphorylation events are required to dislodge the R domain from its obstructing position, presenting a very large energetic barrier (8,78). This conformational homogeneity of dephosphorylated ATP-free CFTR suggests that the first step of the conformational change, that of movement of the R domain, cannot efficiently proceed via an ensemble selection mechanism. In the absence of definitive experimental/computational evidence, we currently propose that this step occurs by an induced fit mechanism, where the reversible covalent phosphorylation is the allosteric ligand (Figure 7, step I). The next step of the conformational change, binding of ATP and closure of the NBDs, is more complex. On the one hand, single molecule studies showed that even after phosphorylation the NBDs solely adopt the open conformation and do not spontaneously close unless ATP is added. Once ATP was added, all the molecules converted to the closed-NBD conformation and were never seen to spontaneously open their NBDs (70). We observed very similar conformational homogeneity and stability with

another member of the ABC transporter family (58). Along the same lines, MD simulations showed that the closed state with dimerized NBDs is stable over hundreds of nanoseconds (79). Taken together, these results strongly argue in favor of an induced-fit mechanism, and against MWC/ensemble-selection/population shift models which assume pre-existence of multiple conformations and a spontaneous dynamic equilibrium between them). Direct experimental evidence supporting an induced fit mechanism was also provided by macroscopic current recordings combined with double mutant cycles (80). On the other hand, when the NBDs are separated the ATP binding sites are yet to form it is hard to envision how the nucleotide binds to its non-existing binding site. Such large conformational changes are considered to occur only via an ensemble selection mechanism (81). We suggest that our results provide the solution: In the ATP free form, despite the separation of the NBDs, the main allosteric sites are located near the ATP binding sites (Figures 1 and 2A). Moreover, many of the dynamic features that manifest following ATP binding exist in a latent form in the ATP-free conformation (Figure 2B and C). This means that the ATP binding residues fluctuate considerably in the absence of nucleotide, sampling conformations in the vicinity of the NBDs-separated native state. Some of these conformations likely provide an initial state for the interaction with the nucleotide (Figure 7, step II). ATP will selectively bind to these conformers, as in ensemble selection allosteric mechanism. Once the docking of the nucleotide is initiated by this combination of dynamic fluctuations and ensemble selection closure of the NBDs can proceed by an induced fit mechanism (Figure 7, step III) as suggested by experimental evidence (80). The resulting conformational changes do not yet lead to channel opening and conductance: The phosphorylated ATP bound structure observed in the cryo-EM studies is non conducting: the pore is fully dehydrated, and an ion translocation pathway is missing. Therefore, phosphorylation, ATP binding, and closure of the NBDs are essential yet insufficient for channel opening. Several lines of evidence suggest that the final missing component that enables conductance are the dynamic fluctuations of the NBDs-closed state: MD simulations of chloride conductance showed that a major conformational change is not required for channel opening. Instead, slight side chain re-orientations of pore residues enable sporadic pulses of chloride conductance (79). Single molecule measurements also failed to identify any conformational changes between the conducting and non-conducting states (70). Along the same lines, allosteric potentiators that increase CFTR's channel-open probabilities do so without noticeable effects on its conformation (34). These studies suggest that the conformations of the

conducting and non-conducting states are very similar and that a major conformational change is not required to open the channel. Rather, transition to the conducting state occurs by a dynamic allosteric mechanism with rapid conformational sampling around the phosphorylated ATP bound NBDs-closed state (Figure 7, step IV).

With this combination of allosteric models in mind, we adopted a structural/dynamic view to study the molecular basis of allosteric connectivity in CFTR. The premise of the structural/dynamic perspective is that strain energy generated by binding of the allosteric ligand (phosphorylation and ATP in CFTR) is mechanically transmitted through a network of interacting residues to be relieved by a conformational switch. To identify this network of interacting residues we used two elastic network approaches: ANM-LD is a trajectory molecular simulation approach which identifies residues that interact during a conformational change (50). It is therefore particularly useful for studying induced fit mechanisms that assume the existence of such networks of interacting residues. GNM-TE on the other hand models the intrinsic dynamics of a protein around its mean native conformations and provides a useful tool to study the dynamic/ensemble selection components of CFTR's allostery. Both approaches are computationally economical, and for a protein as large and complex as CFTR provide an attractive alternative to atomistic MD simulations.

Despite their simplifications, and without any input of functional information, the ANM-LD/ GNM-TE combination clearly distinguishes the three main functional domains of the protein (*i.e.*, NBDs, gating residues, docking site of the R domain) from the rest of the protein (Figures 2 and 3). In further support of the utility of the calculations we observe a remarkable correlation between the residues identified as main allosteric sources and the positions of functionally essential residues (Figure S4), ATP-binding residues (Figure S5), and disease-causing mutations (Figure S6). In addition, despite the high structural similarity between the NBDs the calculations identify the functional singularity of the catalytic NBD2 and identifies it as the dominant allosteric source (Figure 2A). Taken together, these results show that the computations can capture the structural dynamics of CFTR.

The allosteric effects of ATP binding and hydrolysis are not always easily distinguishable. For example, Liu *et al.*, investigated the effects of the force originating from ATP hydrolysis on the chaperonin GroEL. They applied elastic network models and principal component analysis on a series of GroEL structures and found that in this processive enzyme ATP binding is insufficient

to induce the conformational changes, which require nucleotide hydrolysis (82). In CFTR, a large body of evidence suggests that the allosteric driver is binding *per-se* and not hydrolysis: binding of ATP under non hydrolyzing conditions is sufficient to dimerize the NBDs and to open the channel (12,70). In addition, mutations and experimental conditions that stabilize the ATP-bound state by reducing hydrolysis rates stabilize the NBDs-closed conformation and conducting state of the channel. At first glance, the allosteric connectivity networks of the ATP-free and -bound conformations seem very different (Figure S7). However, a closer examination reveals that binding of ATP fine-tunes and brings to the fore dynamics that pre-existed in the ATP-free conformation. For example, in the dephosphorylated ATP-free state conformation, the allosteric signaling originates from many residues in the NBDs, mostly from those that surround the future ATP binding sites (Figures 1, 2). Binding of ATP focuses these relatively diffuse dynamics to the exact location of the ATP binding residues (Figure 3A). Similarly, in the ATP-free conformation the residues that line the permeation pathway and the gating residues have a minor allosteric role and are detected only once the more dominant fluctuations of the NBDs were omitted from the calculations (Figure 2B). Following ATP binding the allosteric role of these residues becomes much more dominant (Figure 3B). We propose that in CFTR the role of ATP binding is 1) To induce the conversion to the NBDs-closed conformation via a combination of ensemble selection and induced fit and mechanisms and 2) To modulate conformational dynamics that are present to different degrees in other conformations via a dynamic allosteric mechanism. In CFTR, ATP binding, rather than hydrolysis is sufficient to exert these effects: binding of ATP is sufficient to induce the closure of the NBDs and to open the channel (70). ATP hydrolysis and product release are not required for channel opening, but rather lead to separation of the NBDs and termination of conductance. This differs from the role of ATP in processive enzymes such as the chaperonin GroEL, where the exothermic energy released by ATP hydrolysis is required to drive the conformational changes (82).

A new allosteric hotspot was identified in TMD1, in a cleft formed by TM helices 2, 3 and 6 (Figures 3C-D). Its proximity to the interaction site of the phosphorylated R domain may provide the molecular basis for the dual role of the R domain: an inhibitory role while dephosphorylated, and a conductance-stimulatory role once phosphorylated (66,83,84). Perhaps more importantly, this allosteric cleft is highly druggable and provides a new target for structure-guided screening of small molecules.

Materials and methods

GNM-based Transfer Entropy

GNM is a minimalist 1D elastic network model that assumes a Gaussian distribution of fluctuations that is isotropic in nature, operating within the confines of a given protein's topology. The GNM model relies on a pair of parameters: force constant governing harmonic interactions and the cutoff distance within which these interactions occur. In this model, the force constant is uniform for all interactions and is conventionally set to unity. The range of interaction is defined by a cutoff distance, typically ranging between 7 to 10 Å, with a value of 10 Å being employed in this study. Transfer Entropy (TE) formulation (52) which is based on GNM (59,85) reveals the direction of information flow between two residues i and j. Given a certain movement in residue i, TE is a measure of the reduction of the uncertainty in the movements of residue j with a time delay τ between the movements of both residues.

$TE_{i,j}(\tau)$ between residues i and j at time τ is formulated as (41)

$$T_{i\to j}(\tau) = S\left(\Delta R_j(t+\tau)\middle|\Delta R_i(t)\right) - S\left(\Delta R_j(t+\tau)\middle|\Delta R_i(t), \Delta R_j(t)\right) \quad (1)$$

where the *S*s are conditional entropies, given by,

$$S\left(\Delta R_j(t+\tau)\middle|\Delta R_i(t)\right) = -\langle \ln p\left(\Delta R_i(0), \Delta R_j(\tau)\right)\rangle + \langle \ln p\left(\Delta R_i(0)\right)\rangle \quad (2)$$

$$S\left(\Delta R_j(t+\tau)\middle|\Delta R_i(t), \Delta R_j(t)\right) = -\langle \ln p\left(\Delta R_j(0), \Delta R_j(\tau)\right)\rangle + \langle \ln p\left(\Delta R_i(0), \Delta R_j(0), \Delta R_j(\tau)\right)\rangle \quad (3)$$

For more details of the definitions of GNM time correlation of the residue fluctuations in Eqs. 4 and 5 see (41,59).

The net TE from residue i to j at a given τ is described as

$$\Delta T_{i\to j}(\tau) = T_{i\to j}(\tau) - T_{j\to i}(\tau) \quad (4)$$

where $\Delta T_{i\to j}(\tau)$ estimates the direction of Transfer Entropy between residues i and j in a certain time delay τ. Thus, the TE calculations reveal entropy/information sources and receivers: Entropy

sources have positive net TE values and send information to many other residues, and entropy receivers have negative net TE values and receive information from the rest of the structure.

The choice of τ between the movement of residues is important to identify causal interrelations: when τ is too small, only local cause-and-effect relations (between adjacent amino acids) will be revealed. if τ is big enough, few (if any) long range cause-and-effect relations will manifest. In a previous work (50), we studied in detail the effects of choosing different τ values and found that an optimal value of τ which maximizes the degree of collectivities of net TE values is in most cases $3\times \tau_{opt}$ (τ_{opt} is the time window in which the total TE of residues is maximized). (See also Figure S1 in (50)).

Degree of Collectivity and TECol Score

The degree of collectivity is a measure of the number of residues that are influenced by the movement of one residue, *i.e.*, how collective is the information transfer? Utilizing Bruschweiler's study(86)(87), collectivity (K) values of residues in information transfer are calculated based on the positive net TE values (42) as

$$\mathrm{K}_{i,s} = \frac{1}{N}\exp\left(-\sum_{j=1}^{N} \alpha\left(\Delta T_{ij,s}(\tau)\right)^2 log(\alpha\left(\Delta T_{ij,s}(\tau)\right)^2\right) \quad (5)$$

where s is a selected subset of slow GNM modes, N is the total number of residues, and α is the normalization factor which is determined as

$$\sum_{i=1}^{N} a\ (\Delta \mathrm{T}_{ij,s}(\tau))^2 = 1 \quad (6)$$

Here, s covers three subsets of the ten slowest modes: Modes 1-10 (*i.e.*, all ten), modes 2-10, and modes 3-10. τ is taken as three times the optimum tau value that maximizes the collective information transfer(42)(42). The 10 slowest modes are considered since the eigenvalue distribution of the dynamic modes shows they fully represent the slow end of the dynamic spectrum (Figure S1).

TECol Score of each residue i is calculated by the multiplication of its cumulative positive net TE value (the sum of positive net TE values that residue i sends to the other residues) with its collectivity $K_{i,s}$ value in each subset s as

$$TECol\ Score_{i,s} = K_{i,s} \cdot \sum_{j=1}^{N} \Delta T_{ij,s} \quad (7)$$

The TECol Score, combining TE and collectivity values, is used in the determination of the most functionally plausible global information source residues that are powerful effectors.

ANM-LD Simulations

The conformational change between dephosphorylated ATP-free (PDB ID: 5UAK)(8) and ATP-bound (PDB ID: 6MSM)(12)(12) human CFTR was simulated using the ANM-LD approach (50,58)(50,58). ANM-LD combines a low-resolution alpha-carbon based Anisotropic Network Model (ANM) with stochastic all-atom implicit solvent Langevin Dynamics (LD) simulations. Intrinsic dynamics characterized by ANM dynamic modes and followed short-time LD simulations guide the conformational transition from the initial to the target state. That is, the conformational transition pathway is successfully evolved from the dephosphorylated ATP-free to the phosphorylated ATP-bound CFTR, starting with 7.5 Å and converging to 2.3 Å of RMSD. Due to the lack of a characterized structure, the regulatory domain is not included in the ANM-LD simulations. The time window in ANM–LD simulations does not reflect real time of simulated transitions yet provides information on transition trajectory and sequence of events, using accessible internal dynamics as the only bias towards the target conformation.

ANM-LD Simulation Flow

The hessian matrix (**H**) with a distance threshold radius (R_{cut}=13 Å) is constructed at each cycle of the ANM-LD simulation to calculate 3N-6 ANM dynamic modes, i.e., eigenvectors $\mathbf{U}_k$ and eigenvalue λ_k, k = 1, 3N-6, where N is the number of residues(43). The ANM mode that best overlaps with the difference vector of the aligned initial/intermediate and target conformations is selected ($\mathbf{U}_{best}$) and the initial conformation is deformed along this vector using a deformation factor (DF: 0.4 Å). Energy minimization is performed for 500 steps and 100 steps of LD simulations are performed with a time step of 0.2 fs at a temperature of 310 K using the Amber 11-Sander 4 biomolecular simulation program(88)(89).

The simplified position conformational sampling of ANM-LD cycles is given below as

$$\boldsymbol{R}_{new} = \boldsymbol{R}_{old} + \Delta\boldsymbol{R}_{ANM} + \Delta R_{Eng.Min} + \Delta\boldsymbol{R}_{LD} \quad (8)$$

where $\Delta\boldsymbol{R}_{\boldsymbol{ANM}} = \boldsymbol{U}_{\boldsymbol{best}} \cdot DF$.

$\Delta\boldsymbol{R}_{Eng.Min}$ and $\Delta\boldsymbol{R}_{LD}$ are energy minimization and LD simulations contributions to the sampling, respectively.

The ANM steps and LD steps are iteratively performed for a predetermined number of ANM-LD cycles or until the RMSD between intermediate and target states converges. In the present ANM-LD simulations of CFTR as the RMSD converged to a value of 2.3 Å.

Equal Time and Time Delay Cross Correlations

Each residue has a fluctuation vector $\Delta\boldsymbol{R}$ that corresponds to the difference between its current position to the mean position of its alpha carbons. The mean positions are calculated from the selected time window of the ANM-LD simulation cycles. The pseudo correlation between the fluctuations of residues *i* and *j* is calculated as

$$C_{ij}(\tau) = \frac{<\Delta\boldsymbol{R}_i(t)\,.\,\Delta\boldsymbol{R}_j(t+\tau)>}{\sqrt{<\Delta\boldsymbol{R}_i(t)^2>}\sqrt{<\Delta\boldsymbol{R}_j(t+\tau)^2>}} \quad (9)$$

where *t* represents an instantaneous pseudo time and τ is an additive time delay in ANM-LD cycles. When τ equals to 0, *Cij* reduces to standard correlation between the movements of residues *i* and *j*, with values between -1 and 1, where 0 corresponding to no correlation, and positive and negative values corresponding to movement in the same or opposite directions, respectively. When τ is greater than 0, the correlation between the fluctuations of residues *i* at time *t* and *j* at $t+\tau$ indicates that the movement of residue *i* is leading and of residue *j* is following in the case of $C_{ij}(\tau) > C_{ji}(\tau)$. The time window in the conformational transition (part of the ANM-LD simulation trajectory from which the conformations are extracted) and the time delay τ are the two main parameters in defining the correlation behavior.

The time delay, τ, an order parameter in a running time window, is a measure of changes in the dynamics that allows a temporal coarse graining in the transition pathway, i.e. quantifying directional correlations. Leading/following events, depending on the length of the time delay and

the time window, reveal cues about the complex hierarchical reorganization of residues in their motion. Here, a time delay of 16 cycles is taken in a transition window of 48 cycles from the ATP-free to the ATP-bound states of CFTR. This time delay corresponds to a time window at which the autocorrelation average over all residues shifts from a positive value to a negative value with respect to the mean conformation within 48 cycles.

Statistical Analysis

In dephosphorylated ATP-free human CFTR (PDB ID 5UAK) we identified a total of 60 allosteric peaks with (a) above average TECol score and (b) above average peak prominence. We then counted how many times the positions of the functionally essential residues or ATP binding sites spatially correlate with these peaks using cut-off distances of 0 Å (i.e., exact match), 4 Å, or 7 Å.

To evaluate the statistical significance of these predictions, we generated 100,000 sets of 60 randomly localized allosteric peaks and counted how many times the positions of the functionally essential residues or ATP sites were in the vicinity of these random allocations. Thus, Z scores of the predictions were determined using the definition of

$$Z = \frac{X - \mu}{\sigma} \tag{10}$$

where X is the count of correct guesses of our prediction while μ and σ are the mean and the standard deviation of the correct matches of random samples, respectively. From the obtained Z scores, p-values of the predictions were calculated by one-tailed hypothesis test using a significance level of 0.05.

The same statistical analysis is performed for the spatial correlation of the hinge positions with the entropy source peaks using cut-off distances of 4 Å or 7 Å.

Visualization

The 3D structural figures presented in this work are created using PyMol 2.5.4 (The PyMOL Molecular Graphics System).

References

1. Farrell PM. The prevalence of cystic fibrosis in the European Union. Journal of Cystic Fibrosis. 2008 Sep;7(5):450–3.
2. McBennett KA, Davis PB, Konstan MW. Increasing life expectancy in cystic fibrosis: Advances and challenges. Pediatr Pulmonol. 2022 Feb 1;57(Suppl 1):S5.
3. Cftr2.org. The Clinical and Functional TRanslation of CFTR (CFTR2) [Internet]. 2011 [cited 2021 Dec 6]. Available from: https://cftr2.org/
4. Fahy J V., Dickey BF. Airway Mucus Function and Dysfunction. New England Journal of Medicine. 2010 Dec 2;363(23):2233–47.
5. Dean M, Rzhetsky A, Allikmets R. The Human ATP-Binding Cassette (ABC) Transporter Superfamily. Genome Res. 2001 Jul 1;11(7):1156–66.
6. Rees DC, Johnson E, Lewinson O. ABC transporters: the power to change. Nat Rev Mol Cell Biol. 2009 Mar;10(3):218–27.
7. Gadsby DC, Vergani P, Csanády L. The ABC protein turned chloride channel whose failure causes cystic fibrosis. Vol. 440, Nature. Nature Publishing Group; 2006. p. 477–83.
8. Liu F, Zhang Z, Csanády L, Gadsby DC, Chen J. Molecular Structure of the Human CFTR Ion Channel. Cell [Internet]. 2017 Mar 23 [cited 2020 Jul 19];169(1):85-95.e8. Available from: https://pubmed.ncbi.nlm.nih.gov/28340353/
9. Zhang Z, Chen J. Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator. Cell. 2016 Dec 1;167(6):1586-1597.e9.
10. Picciotto MR, Cohn JA, Bertuzzi G, Greengard P, Nairn AC. Phosphorylation of the cystic fibrosis transmembrane conductance regulator. J Biol Chem. 1992;267(18):12742–52.
11. Tabcharani JA, Chang XB, Riordan JR, Hanrahan JW. Phosphorylation-regulated Cl- channel in CHO cells stably expressing the cystic fibrosis gene. Nature. 1991;352(6336):628–31.
12. Zhang Z, Liu F, Chen J. Molecular structure of the ATP-bound, phosphorylated human CFTR. Proc Natl Acad Sci U S A [Internet]. 2018 Dec 11 [cited 2021 Nov 19];115(50):12757–62. Available from: /pmc/articles/PMC6294961/
13. Mense M, Vergani P, White DM, Altberg G, Nairn AC, Gadsby DC. In vivo phosphorylation of CFTR promotes formation of a nucleotide-binding domain heterodimer. EMBO J. 2006 Oct 18;25(20):4728–39.
14. Hohl M, Briand C, Grütter MG, Seeger MA. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nat Struct Mol Biol. 2012 Mar 25;19(4):395–402.
15. Csanády L, Vergani P, Gadsby DC. Structure, gating, and regulation of the CFTR anion channel. Physiol Rev. 2019 Jan 1;99(1):707–38.
16. Stockner T, Gradisch R, Schmitt L. The role of the degenerate nucleotide binding site in type I ABC exporters. Vol. 594, FEBS Letters. Wiley Blackwell; 2020. p. 3815–38.
17. Johnson ZL, Chen J. ATP Binding Enables Substrate Release from Multidrug Resistance Protein 1. Cell. 2018 Jan 11;172(1–2):81-89.e10.
18. Kim Y, Chen J. Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. Science. 2018 Feb 23;359(6378):915–9.

731 19. Nguyen PT, Lai JY, Lee AT, Kaiser JT, Rees DC. Noncanonical role for the binding protein in
732 substrate uptake by the MetNI methionine ATP Binding Cassette (ABC) transporter.
733 Proceedings of the National Academy of Sciences. 2018 Nov 6;115(45):E10596–604.
734 20. Kessel A, Ben-Tal N. Introduction to Proteins: Structure, Function, and Motion. 2018th
735 ed. CRC Press; 2018.
736 21. Liu J, Nussinov R. Allostery: An Overview of Its History, Concepts, Methods, and
737 Applications. PLoS Comput Biol. 2016 Jun 1;12(6).
738 22. Gerek ZN, Ozkan SB. Change in allosteric network affects binding affinities of PDZ
739 domains: Analysis through perturbation response scanning. PLoS Comput Biol.
740 2011;7(10).
741 23. Campitelli P, Modi T, Kumar S, Banu Ozkan S. The Role of Conformational Dynamics and
742 Allostery in Modulating Protein Evolution. Vol. 49, Annual Review of Biophysics. Annual
743 Reviews Inc.; 2020. p. 267–88.
744 24. Koshland DE, Nemethy JG, Filmer D. Comparison of experimental binding data and
745 theoretical models in proteins containing subunits. Biochemistry [Internet]. 1966 Jan 1
746 [cited 2023 Sep 14];5(1):365–85. Available from:
747 https://pubmed.ncbi.nlm.nih.gov/5938952/
748 25. Monod J, Wyman J, Changeux JP. ON THE NATURE OF ALLOSTERIC TRANSITIONS: A
749 PLAUSIBLE MODEL. J Mol Biol [Internet]. 1965 [cited 2023 Sep 14];12(1):88–118.
750 Available from: https://pubmed.ncbi.nlm.nih.gov/14343300/
751 26. Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. Protein Eng
752 [Internet]. 1999 [cited 2023 Sep 14];12(9):713–20. Available from:
753 https://pubmed.ncbi.nlm.nih.gov/10506280/
754 27. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery [Internet]. Vol.
755 508, Nature. Nature Publishing Group; 2014 [cited 2021 Nov 19]. p. 331–9. Available
756 from: /pmc/articles/PMC4224315/
757 28. Nussinov R, Tsai CJ. Allostery without a conformational change? Revisiting the paradigm.
758 Curr Opin Struct Biol [Internet]. 2015 Feb 1 [cited 2023 Sep 14];30:17–24. Available from:
759 https://pubmed.ncbi.nlm.nih.gov/25500675/
760 29. Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically driven protein allostery. Nat
761 Struct Mol Biol [Internet]. 2006 Sep [cited 2023 Sep 14];13(9):831–8. Available from:
762 https://pubmed.ncbi.nlm.nih.gov/16906160/
763 30. Tsai CJ, del Sol A, Nussinov R. Allostery: Absence of a Change in Shape Does Not Imply
764 that Allostery Is Not at Play. J Mol Biol. 2008 Apr 18;378(1):1–11.
765 31. Levring J, Terry DS, Kilic Z, Fitzgerald G, Blanchard S, Chen J. CFTR function, pathology and
766 pharmacology at single-molecule resolution. Nature. 2023;
767 32. Sorum B, Czégé D, Csanády L. Timing of CFTR Pore Opening and Structure of Its Transition
768 State. Cell. 2015 Oct 22;163(3):724–33.
769 33. Wang W, Wu J, Bernard K, Li G, Wang G, Bevensee MO, et al. ATP-independent CFTR
770 channel gating and allosteric modulation by phosphorylation. Proc Natl Acad Sci U S A.
771 2010 Feb 23;107(8):3888–93.
772 34. Liu F, Zhang Z, Levit A, Levring J, Touhara KK, Shoichet BK, et al. Structural identification
773 of a hotspot on CFTR for potentiation. Science (1979) [Internet]. 2019 [cited 2021 Nov
774 19];364(6446):1184–8. Available from: https://pubmed.ncbi.nlm.nih.gov/31221859/

775 35. Fiedorczuk K, Chen J. Molecular structures reveal synergistic rescue of D508 CFTR by
776 Trikafta modulators. Science (1979). 2022 Oct 21;378(6617):284–90.
777 36. Eckford PDW, Li C, Ramjeesingh M, Bear CE. Cystic fibrosis transmembrane conductance
778 regulator (CFTR) potentiator VX-770 (ivacaftor) opens the defective channel gate of
779 mutant CFTR in a phosphorylationdependent but ATP-independent manner. Journal of
780 Biological Chemistry. 2012 Oct 26;287(44):36639–49.
781 37. Shaughnessy CA, Zeitlin PL, Bratcher PE. Elexacaftor is a CFTR potentiator and acts
782 synergistically with ivacaftor during acute and chronic treatment. Sci Rep. 2021 Dec
783 1;11(1).
784 38. Sorum B, Töröcsik B, Csanády L. Asymmetry of movements in CFTR's two ATP sites during
785 pore opening serves their distinct functions. Elife. 2017 Sep 25;6.
786 39. Scholl D, Sigoillot M, Overtus M, Martinez RC, Martens C, Wang Y, et al. A topological
787 switch in CFTR modulates channel activity and sensitivity to unfolding. Nat Chem Biol.
788 2021 Sep 1;17(9):989–97.
789 40. Wei S, Roessler BC, Chauvet S, Guo J, Hartman IV JL, Kirk KL. Conserved allosteric hot
790 spots in the transmembrane domains of Cystic Fibrosis Transmembrane Conductance
791 Regulator (CFTR) channels and Multidrug Resistance Protein (MRP) pumps. Journal of
792 Biological Chemistry. 2014;289(29):19942–57.
793 41. Hacisuleyman A, Erman B. Causality, transfer entropy, and allosteric communication
794 landscapes in proteins with harmonic interactions. Proteins: Structure, Function, and
795 Bioinformatics [Internet]. 2017 Jun [cited 2019 Aug 27];85(6):1056–64. Available from:
796 http://www.ncbi.nlm.nih.gov/pubmed/28241380
797 42. Altintel B, Acar B, Erman B, Haliloglu T. Subsets of Slow Dynamic Modes Reveal Global
798 Information Sources as Allosteric Sites. J Mol Biol [Internet]. 2022 Sep 15 [cited 2023 Mar
799 22];434(17). Available from: https://pubmed.ncbi.nlm.nih.gov/35644497/
800 43. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of
801 Fluctuation Dynamics of Proteins with an Elastic Network Model. Biophys J [Internet].
802 2001 Jan [cited 2017 Apr 4];80(1):505–15. Available from:
803 http://www.ncbi.nlm.nih.gov/pubmed/11159421
804 44. Wang Y, Rader AJ, Bahar I, Jernigan RL. Global ribosome motions revealed with elastic
805 network model. J Struct Biol. 2004 Sep;147(3):302–14.
806 45. Keskin O, Durell SR, Bahar I, Jernigan RL, Covell DG. Relating molecular flexibility to
807 function: A case study of tubulin. Biophys J. 2002;83(2):663–80.
808 46. Keskin O. Binding induced conformational changes of proteins correlate with their
809 intrinsic fluctuations: A case study of antibodies. BMC Struct Biol. 2007;7.
810 47. Brünger A, Brooks CL, Karplus M. Stochastic boundary conditions for molecular dynamics
811 simulations of ST2 water. Chem Phys Lett. 1984 Mar 30;105(5):495–500.
812 48. Yang LW, Bahar I. Coupling between catalytic site and collective dynamics: a requirement
813 for mechanochemical activity of enzymes. Structure [Internet]. 2005 Jun [cited 2023 Sep
814 14];13(6):893–904. Available from: https://pubmed.ncbi.nlm.nih.gov/15939021/
815 49. Li H, Sakuraba S, Chandrasekaran A, Yang LW. Molecular binding sites are located near
816 the interface of intrinsic dynamics domains (IDDs). J Chem Inf Model [Internet]. 2014 Aug
817 25 [cited 2023 Sep 14];54(8):2275–85. Available from:
818 https://pubmed.ncbi.nlm.nih.gov/25089914/

50. Acar B, Rose J, Aykac Fas B, Ben-Tal N, Lewinson O, Haliloglu T. Distinct Allosteric Networks Underlie Mechanistic Speciation of ABC Transporters. Structure [Internet]. 2020 Apr 2 [cited 2020 Jul 19];28(6):651-663.e5. Available from: https://pubmed.ncbi.nlm.nih.gov/32320672/
51. Cui Q, Bahar Ivet. Normal mode analysis : theory and applications to biological and chemical systems [Internet]. Chapman & Hall/CRC; 2006 [cited 2019 Sep 2]. 406 p. Available from: https://www.crcpress.com/Normal-Mode-Analysis-Theory-and-Applications-to-Biological-and-Chemical/Cui-Bahar/p/book/9781584884729
52. Hacisuleyman A, Erman B. Entropy Transfer between Residue Pairs and Allostery in Proteins: Quantifying Allosteric Communication in Ubiquitin. PLoS Comput Biol [Internet]. 2017 Jan 1 [cited 2022 Dec 27];13(1). Available from: https://pubmed.ncbi.nlm.nih.gov/28095404/
53. Krieger JM, Doruker P, Scott AL, Perahia D, Bahar I. Towards gaining sight of multiscale events: utilizing network models and normal modes in hybrid methods. Curr Opin Struct Biol [Internet]. 2020 Oct 1 [cited 2023 Sep 14];64:34–41. Available from: https://pubmed.ncbi.nlm.nih.gov/32622329/
54. Gur M, Zomot E, Bahar I. Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions. J Chem Phys [Internet]. 2013 Sep 28 [cited 2023 Sep 14];139(12). Available from: https://pubmed.ncbi.nlm.nih.gov/24089724/
55. Jalalypour F, Sensoy O, Atilgan C, Atilgan C. Perturb-Scan-Pull: A Novel Method Facilitating Conformational Transitions in Proteins. J Chem Theory Comput [Internet]. 2020 Jun 9 [cited 2023 Sep 14];16(6):3842–55. Available from: https://pubmed.ncbi.nlm.nih.gov/32324386/
56. Kandzia F, Ostermeir K, Zacharias M. Global Dynamics of Yeast Hsp90 Middle and C-Terminal Dimer Studied by Advanced Sampling Simulations. Front Mol Biosci [Internet]. 2019 Sep 1 [cited 2023 Sep 14];6(SEP). Available from: https://pubmed.ncbi.nlm.nih.gov/31681792/
57. Brünger A, Brooks CL, Karplus M. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. Chem Phys Lett. 1984 Mar 30;105(5):495–500.
58. Yang M, Livnat Levanon N, Acar B, Aykac Fas B, Masrati G, Rose J, et al. Single-molecule probing of the conformational homogeneity of the ABC transporter BtuCD. Nat Chem Biol [Internet]. 2018 Jul 18 [cited 2018 Dec 10];14(7):715–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29915236
59. Haliloglu T, Bahar I, Erman B. Gaussian Dynamics of Folded Proteins. Phys Rev Lett [Internet]. 1997 Oct 20 [cited 2018 Dec 11];79(16):3090–3. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.79.3090
60. Wodak SJ, Paci E, Dokholyan N V., Berezovsky IN, Horovitz A, Li J, et al. Allostery in Its Many Disguises: From Theory to Applications. Vol. 27, Structure. Cell Press; 2019. p. 566–78.
61. Chen Q, Shen Y, Zheng J. A review of cystic fibrosis: Basic and clinical aspects. Animal Model Exp Med. 2021 Sep 1;4(3):220–32.

861 62. Thelin WR, Chen Y, Gentzsch M, Kreda SM, Sallee JL, Scarlett CO, et al. Direct interaction
862 with filamins modulates the stability and plasma membrane expression of CFTR. J Clin
863 Invest. 2007 Feb 1;117(2):364–74.
864 63. Bozoky Z, Krzeminski M, Muhandiram R, Birtley JR, Al-Zahrani A, Thomas PJ, et al.
865 Regulatory R region of the CFTR chloride channel is a dynamic integrator of phospho-
866 dependent intra- and intermolecular interactions. Proc Natl Acad Sci U S A. 2013 Nov
867 19;110(47).
868 64. Chappe V, Irvine T, Liao J, Evagelidis A, Hanrahan JW. Phosphorylation of CFTR by PKA
869 promotes binding of the regulatory domain. EMBO Journal. 2005 Aug 3;24(15):2730–40.
870 65. Zhang Z, Liu F, Chen J. Conformational Changes of CFTR upon Phosphorylation and ATP
871 Binding. Cell [Internet]. 2017 Jul 27 [cited 2023 Sep 14];170(3):483-491.e8. Available
872 from: https://pubmed.ncbi.nlm.nih.gov/28735752/
873 66. Winter MC, Welsh MJ. Stimulation of CFTR activity by its phosphorylated R domain.
874 Nature. 1997;389(6648):294–6.
875 67. Ma J, Zhao J, Drumm ML, Xie J, Davis PB. Function of the R domain in the cystic fibrosis
876 transmembrane conductance regulator chloride channel. Journal of Biological Chemistry.
877 1997 Oct 31;272(44):28133–41.
878 68. Perez C, Gerber S, Boilevin J, Bucher M, Darbre T, Aebi M, et al. Structure and mechanism
879 of an active lipid-linked oligosaccharide flippase. Nature. 2015 Aug 12;524(7566):433–8.
880 69. Infield DT, Cui G, Kuang C, McCarty NA. Positioning of extracellular loop 1 affects pore
881 gating of the cystic fibrosis transmembrane conductance regulator. Am J Physiol Lung
882 Cell Mol Physiol. 2016 Mar 1;310(5):L403–14.
883 70. Levring J, Terry DS, Kilic Z, Fitzgerald G, Blanchard S, Chen J. CFTR function, pathology and
884 pharmacology at single-molecule resolution. Nature [Internet]. 2023 Apr 20 [cited 2023
885 Sep 14];616(7957):606–14. Available from: https://pubmed.ncbi.nlm.nih.gov/36949202/
886 71. Clancy JP. Rapid therapeutic advances in CFTR modulator science. Pediatr Pulmonol.
887 2018 Nov 5;53(S3):S4–11.
888 72. Van Goor F, Hadida S, Grootenhuis PDJ, Burton B, Stack JH, Straley KS, et al. Correction of
889 the F508del-CFTR protein processing defect in vitro by the investigational drug VX-809.
890 Proc Natl Acad Sci U S A. 2011 Nov 15;108(46):18843–8.
891 73. Fiedorczuk K, Chen J. Mechanism of CFTR correction by type I folding correctors. Cell.
892 2022 Jan 6;185(1):158-168.e11.
893 74. Veit G, Vaccarin C, Lukacs GL. Elexacaftor co-potentiates the activity of F508del and
894 gating mutants of CFTR. Journal of Cystic Fibrosis. 2021 Sep 1;20(5):895–8.
895 75. Hadida S, Van Goor F, Zhou J, Arumugam V, McCartney J, Hazlewood A, et al. Discovery
896 of N -(2,4-Di- tert -butyl-5-hydroxyphenyl)-4-oxo-1,4-dihydroquinoline-3-carboxamide
897 (VX-770, Ivacaftor), a potent and orally bioavailable CFTR potentiator. J Med Chem. 2014
898 Dec 11;57(23):9776–95.
899 76. Yu H, Burton B, Huang CJ, Worley J, Cao D, Johnson JP, et al. Ivacaftor potentiation of
900 multiple CFTR channels with gating mutations. J Cyst Fibros. 2012 May;11(3):237–45.
901 77. Ribeiro AAST, Ortiz V. A Chemical Perspective on Allostery. Chem Rev [Internet]. 2016 Jun
902 8 [cited 2023 Sep 14];116(11):6488–502. Available from:
903 https://pubmed.ncbi.nlm.nih.gov/26741913/

904 78. Gadsby DC, Nairn AC. Control of CFTR channel gating by phosphorylation and nucleotide
905 hydrolysis. Physiol Rev [Internet]. 1999 [cited 2023 Sep 14];79(1 Suppl). Available from:
906 https://pubmed.ncbi.nlm.nih.gov/9922377/
907 79. Zeng ZW, Linsdell P, Pomès R. Molecular dynamics study of Cl- permeation through cystic
908 fibrosis transmembrane conductance regulator (CFTR). Cell Mol Life Sci [Internet]. 2023
909 Feb 1 [cited 2023 Sep 14];80(2). Available from:
910 https://pubmed.ncbi.nlm.nih.gov/36694009/
911 80. Szollosi A, Vergani P, Csanády L. Involvement of F1296 and N1303 of CFTR in induced-fit
912 conformational change in response to ATP binding at NBD2. J Gen Physiol [Internet].
913 2010 Oct [cited 2023 Sep 14];136(4):407–23. Available from:
914 https://pubmed.ncbi.nlm.nih.gov/20876359/
915 81. Nussinov R, Ma B, Tsai CJ. Multiple conformational selection and induced fit events take
916 place in allosteric propagation. Biophys Chem [Internet]. 2014 [cited 2023 Sep
917 18];186:22. Available from: /pmc/articles/PMC6361548/
918 82. Liu J, Sankar K, Wang Y, Jia K, Jernigan RL. Directional Force Originating from ATP
919 Hydrolysis Drives the GroEL Conformational Change. Biophys J [Internet]. 2017 Apr 25
920 [cited 2023 Sep 14];112(8):1561–70. Available from:
921 https://pubmed.ncbi.nlm.nih.gov/28445748/
922 83. Ostedgaard LS, Baldursson O, Welsh MJ. Regulation of the Cystic Fibrosis
923 Transmembrane Conductance Regulator Cl- Channel by Its R Domain. Vol. 276, Journal of
924 Biological Chemistry. J Biol Chem; 2001. p. 7689–92.
925 84. Csanády L, Chan KW, Seto-Young D, Kopsco DC, Nairn AC, Gadsby DC. Severed channels
926 probe regulation of gating of cystic fibrosis transmembrane conductance regulator by its
927 cytoplasmic domains. Journal of General Physiology. 2000;116(3):477–500.
928 85. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a
929 single-parameter harmonic potential. Fold Des [Internet]. 1997 Jun [cited 2018 Dec
930 11];2(3):173–81. Available from:
931 http://linkinghub.elsevier.com/retrieve/pii/S1359027897000242
932 86. Brüschweiler R. Collective protein dynamics and nuclear spin relaxation. J Chem Phys
933 [Internet]. 1995 Feb 22 [cited 2023 Mar 29];102(8):3396–403. Available from:
934 http://aip.scitation.org/doi/10.1063/1.469213
935 87. Brüschweiler R. Collective protein dynamics and nuclear spin relaxation. J Chem Phys.
936 1995 Feb 22;102(8):3396–403.
937 88. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S.
938 Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G.
939 Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov and PAK.
940 AMBER 11 [Internet]. San Francisco.: University of California; 2010. Available from:
941 http://ambermd.org/
942 89. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S.
943 Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G.
944 Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov and PAK.
945 AMBER 11. San Francisco.: University of California; 2010.
946

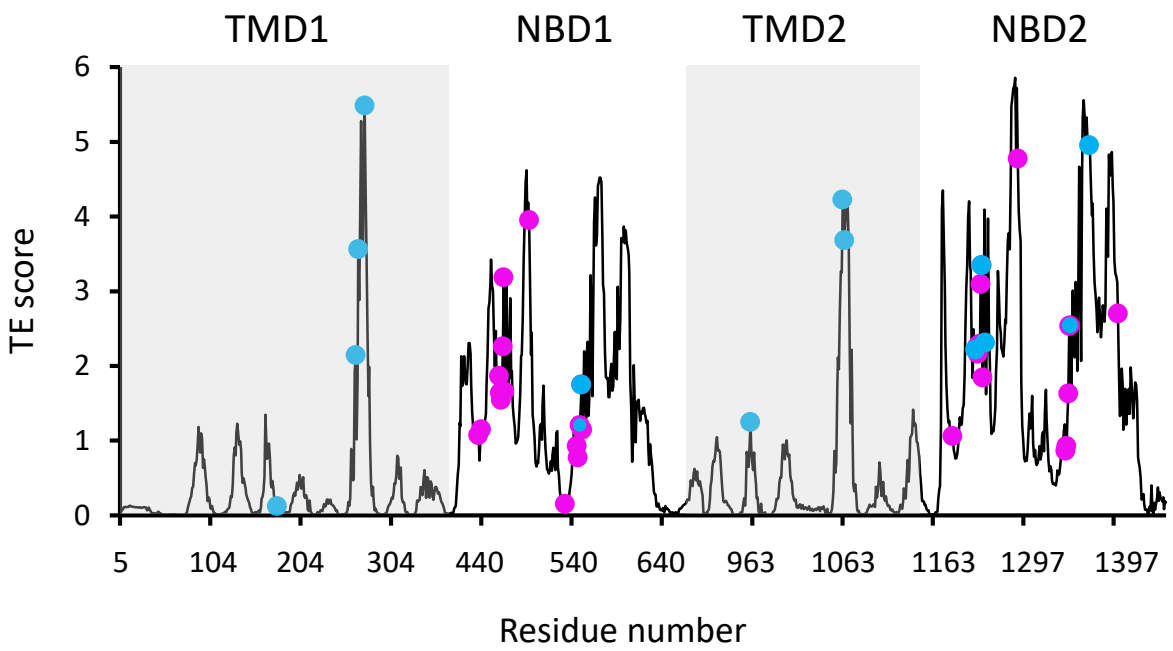

TMD1
NBD1
TMD2
NBD2
TE score
0
1
2
3
4
5
6
5
104
204
304
440
540
640
963
1063
1163
1297
1397
Residue number

Figure 1

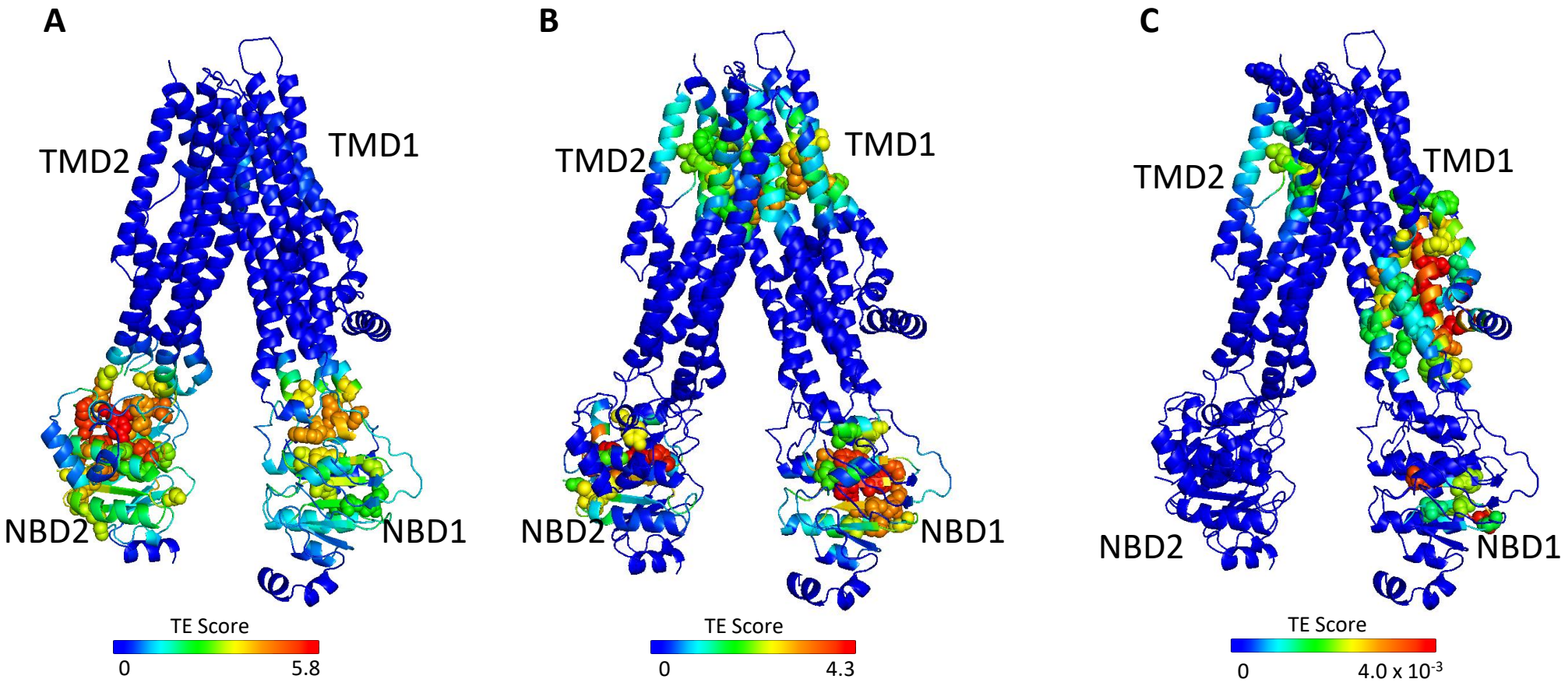

A
TMD2
TMD1
NBD2
NBD1
TE Score
0
5.8
B
TMD2
TMD1
NBD2
NBD1
TE Score
0
4.3
C
TMD2
TMD1
NBD2
NBD1
TE Score
0
4.0 x 10-3

Figure 2

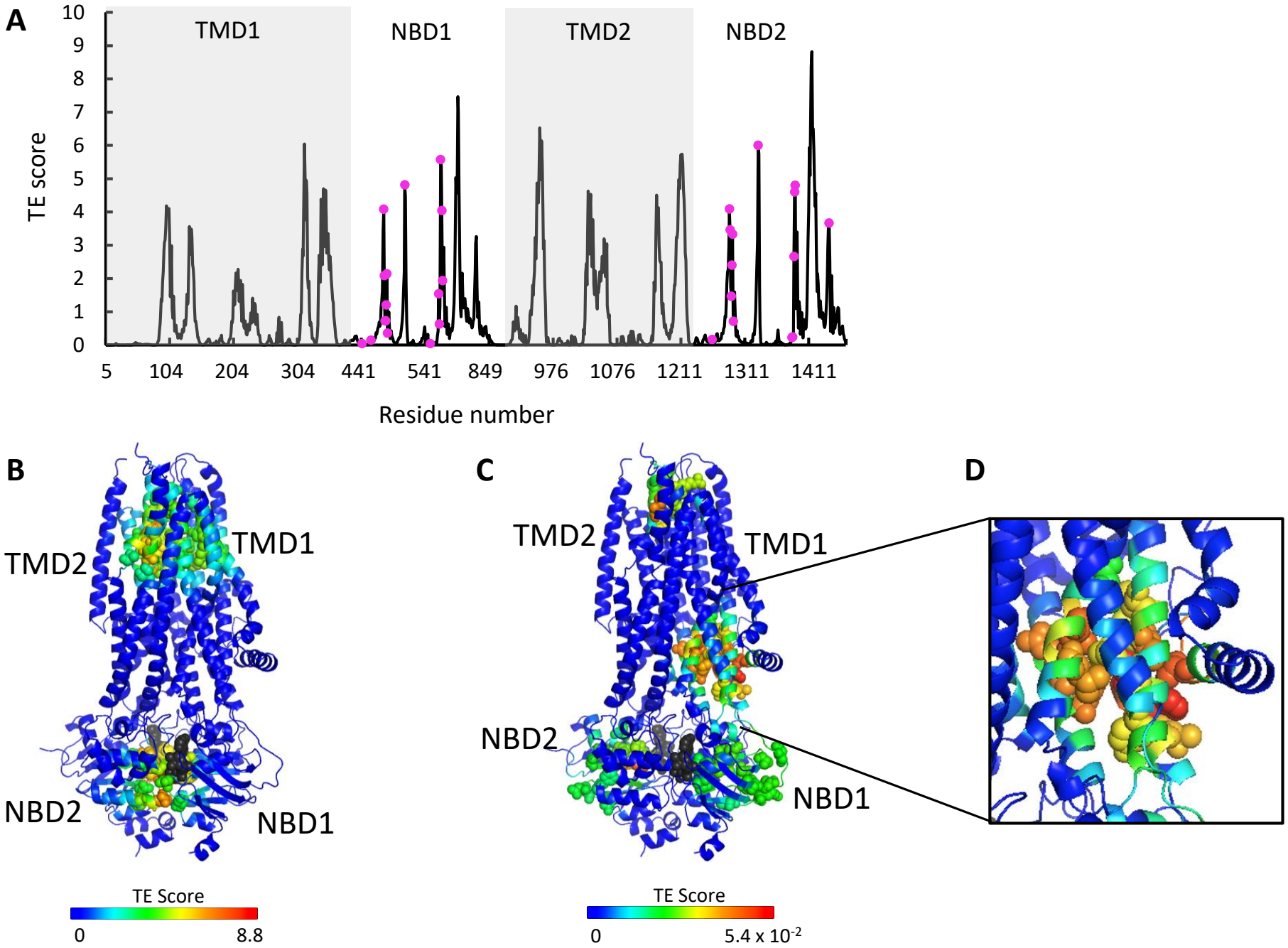

A
TMD1
NBD1
TMD2
NBD2
TE score
10
9
8
7
6
5
4
3
2
1
0
5
104
204
304
441
541
849
976
1076
1211
1311
1411
Residue number
B
TMD2
TMD1
NBD2
NBD1
TE Score
0
8.8
C
TMD2
TMD1
NBD2
NBD1
TE Score
0
5.4 x 10-2
D

Figure 3

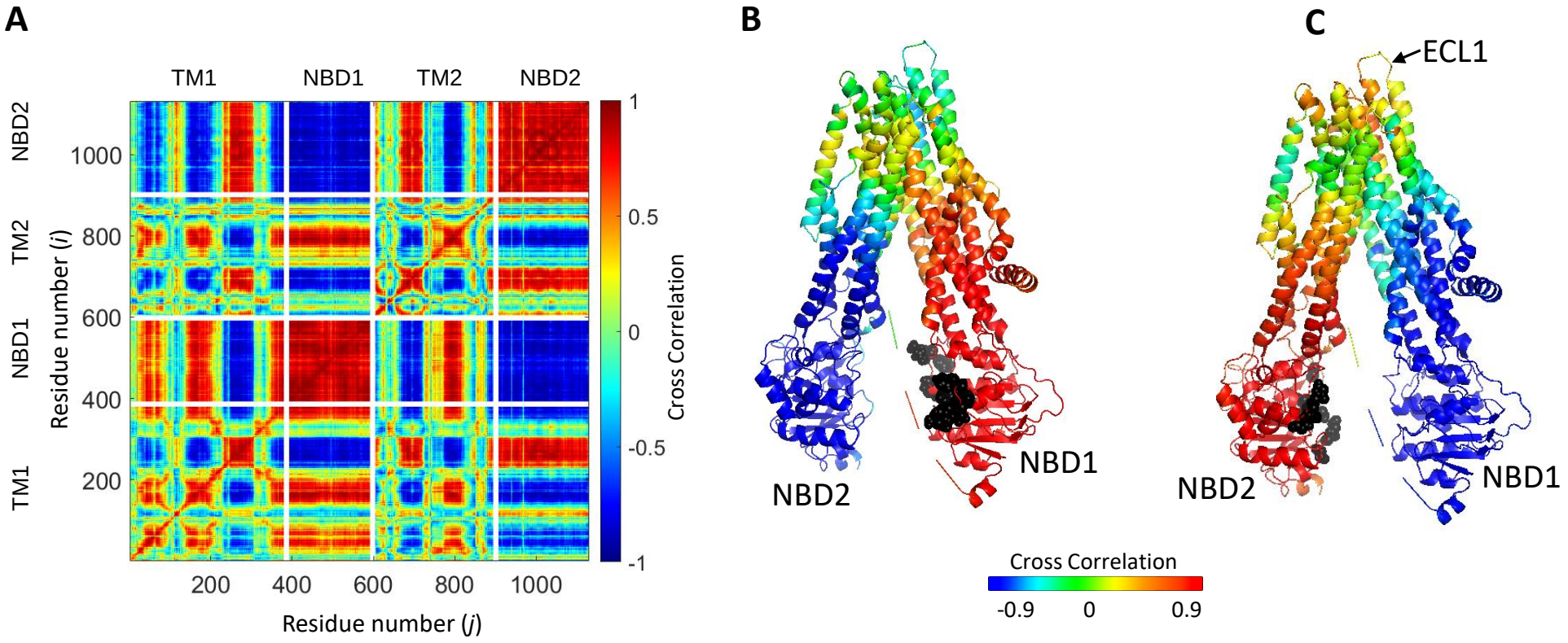

A
TM1
NBD1
TM2
NBD2
NBD2
TM2
NBD1
TM1
Residue number (i)
1000
800
600
400
200
200
400
600
800
1000
Residue number (j)
1
0.5
0
-0.5
-1
Cross Correlation
B
NBD2
NBD1
C
ECL1
NBD2
NBD1
Cross Correlation
-0.9
0
0.9

Figure 4

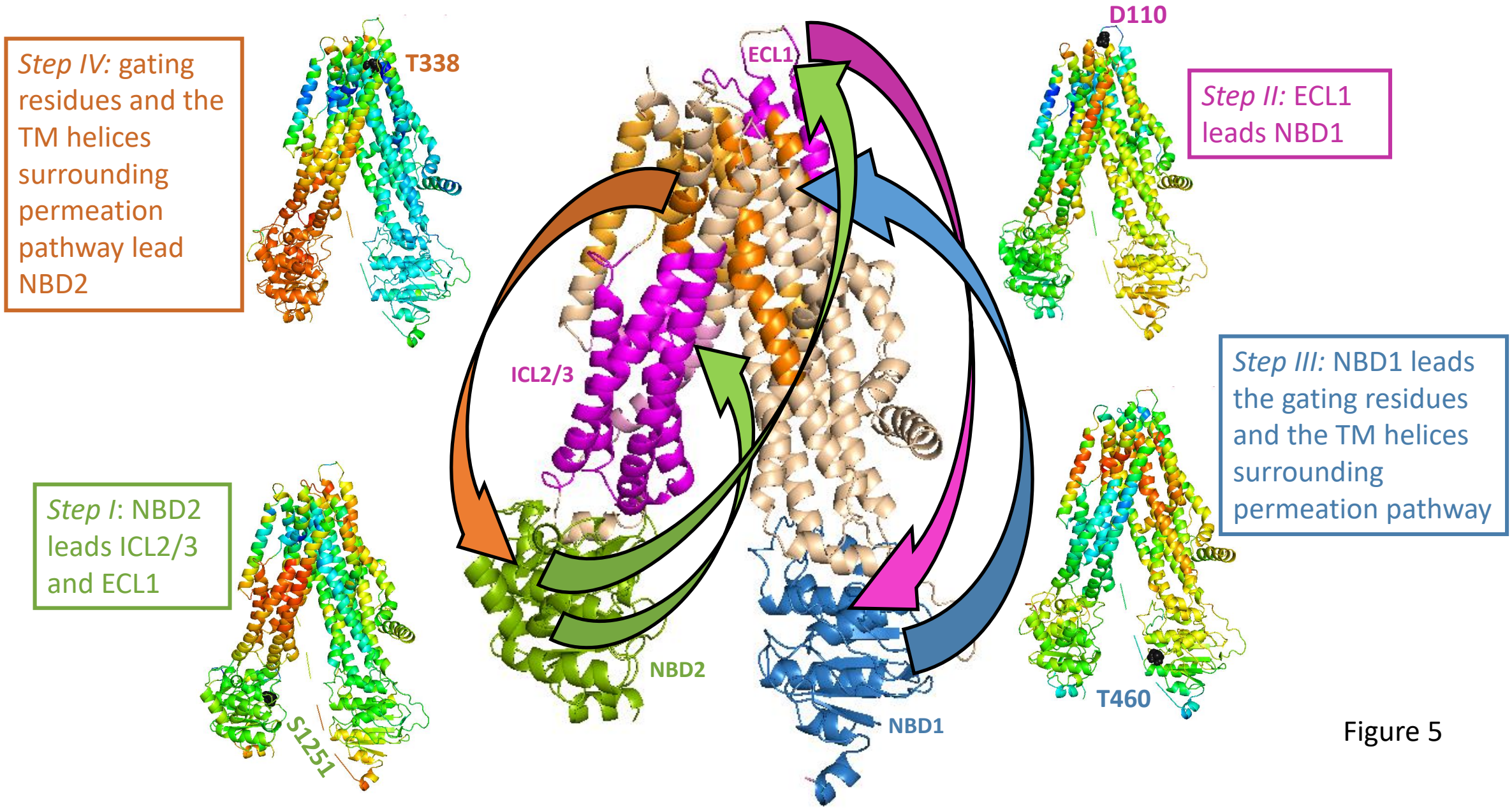
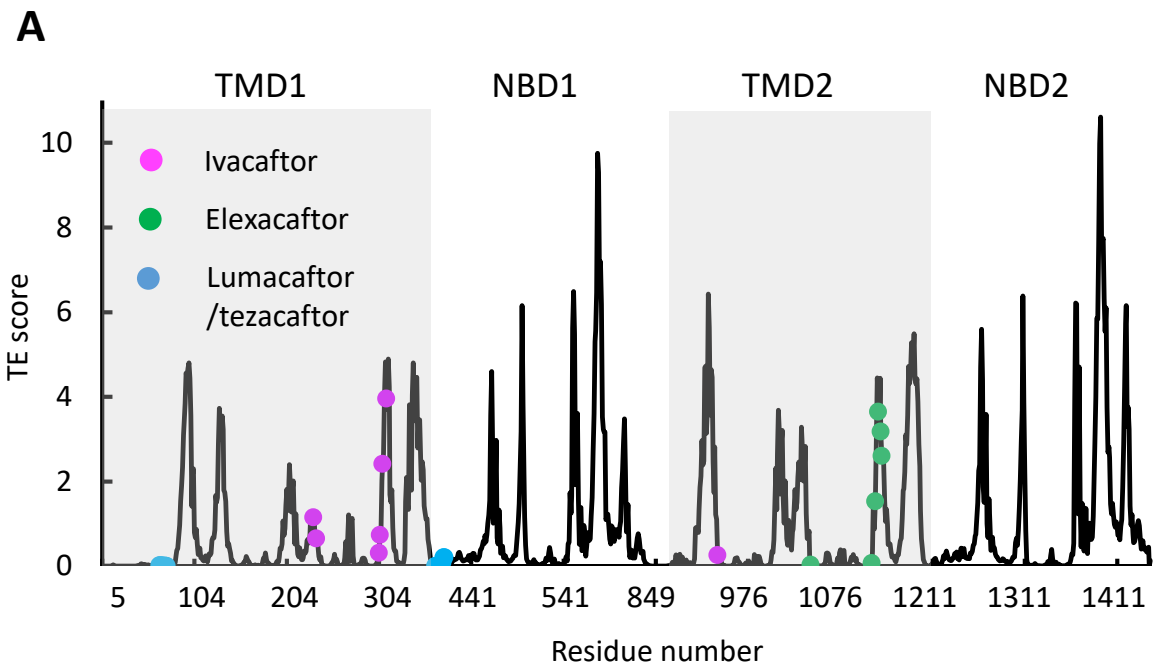
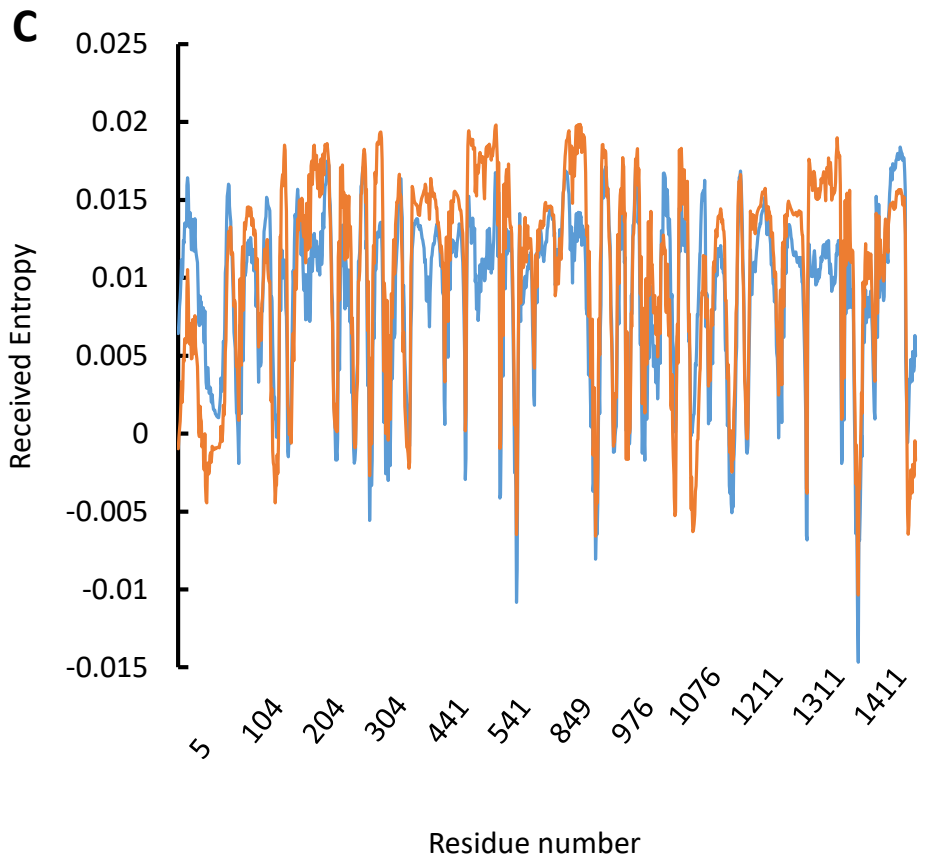

Step IV: gating residues and the TM helices surrounding permeation pathway lead NBD2
T338
ECL1
D110
Step II: ECL1 leads NBD1
ICL2/3
Step III: NBD1 leads the gating residues and the TM helices surrounding permeation pathway
Step I: NBD2 leads ICL2/3 and ECL1
S1251
NBD2
NBD1
T460

Figure 5


A
TMD1
NBD1
TMD2
NBD2
Ivacaftor
Elexacaftor
Lumacaftor
/tezacaftor
TE score
10
8
6
4
2
0
5
104
204
304
441
541
849
976
1076
1211
1311
1411
Residue number


C
Received Entropy
0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
5
104
204
304
441
541
849
976
1076
1211
1311
1411
Residue number

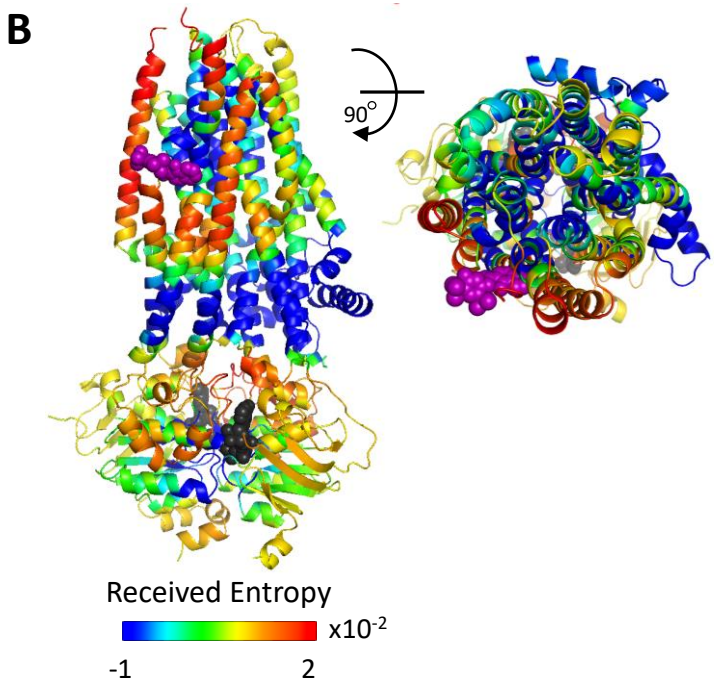

B
90°
Received Entropy
x10-2
-1
2

Figure 6

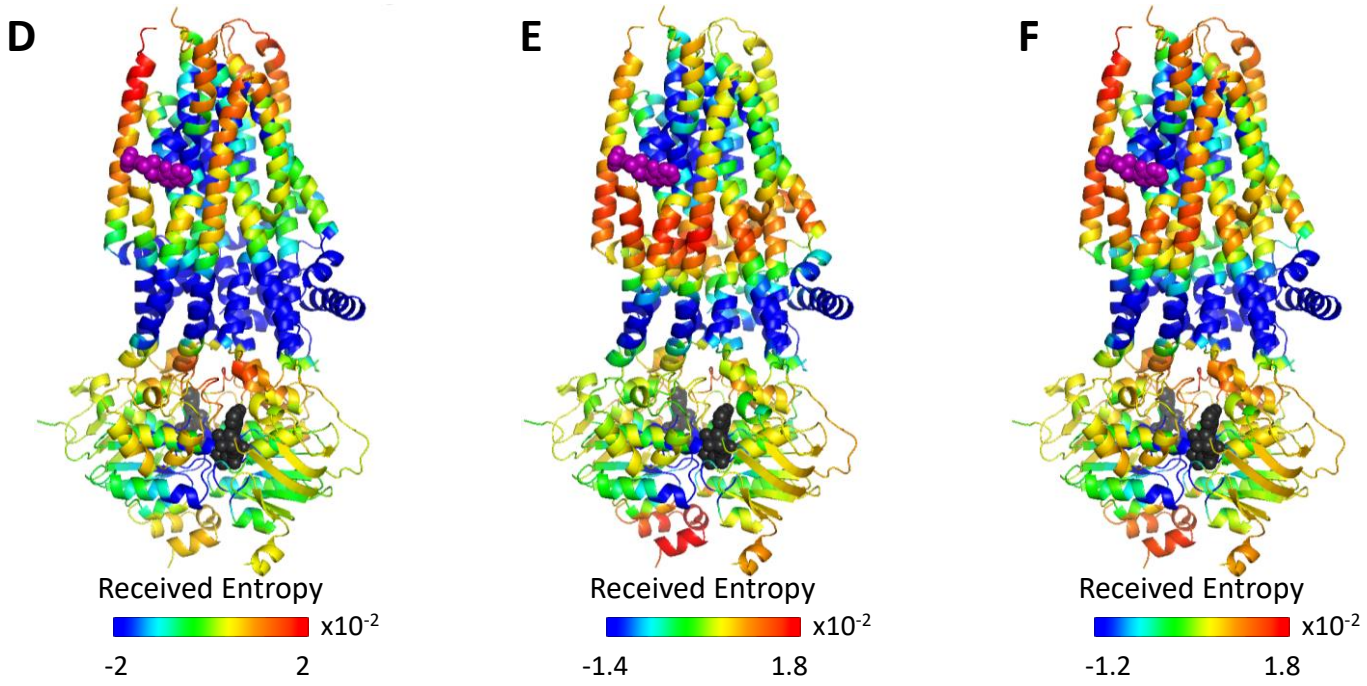


Figure 6 (cont.)

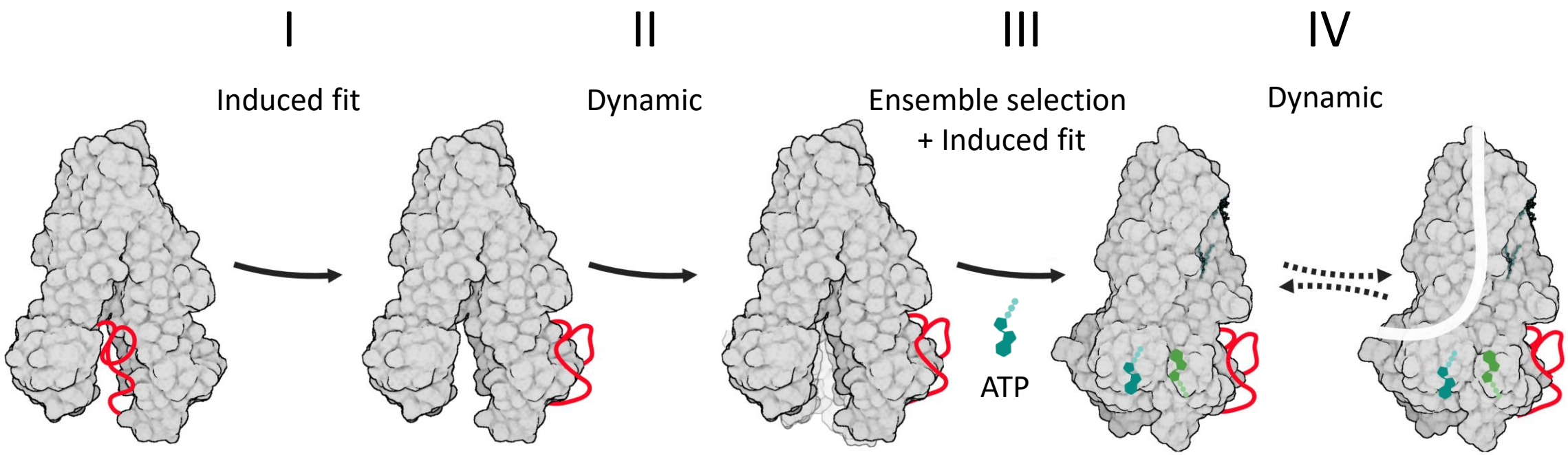

I
Induced fit
II
Dynamic
III
Ensemble selection
+ Induced fit
ATP
IV
Dynamic

Figure 7

Figure 1. Functional sites of CFTR segregate to allosteric hotspots. Shown is the amount of information (TE score) transmitted by each residue of dephosphorylated ATP-free human CFTR (PDB ID 5UAK) calculated using the ten most collective GNM modes (solid black trace). The positions of the 14 functionally important residues (see text for details) and of 30 ATP binding residues are shown as cyan and magenta spheres, respectively.

Figure 2. Identification of allosteric hotspots in dephosphorylated ATP-free CFTR.
A cartoon representation of dephosphorylated ATP-free human CFTR (PDB ID 5UAK), where each residue is colored according to the amount of information it transmits (TE score), expressed as Transfer Entropy x Collectivity (TECol) calculated using the ten most collective GNM modes (A), or upon removal of the first (B) or first and second (C) most collective GNM modes. Red and blue colors represent high and low levels of information output, respectively.

Figure 3. ATP binding and phosphorylation rewires and focuses the allosteric connectivity of CFTR. (A) Shown is the amount of information transmitted (TE score) by each residue of phosphorylated ATP-bound human CFTR (PDB ID 6MSM) calculated using the ten most collective GNM modes (solid black trace). The positions of the ATP binding residues are shown as magenta spheres. (B,C) A cartoon representation of phosphorylated ATP-bound human CFTR (PDB ID 6MSM), where each residue is colored according to the amount of information it transmits (TE score), calculated using the ten most collective GNM modes (B), or upon removal of the first most collective GNM mode (C). Red and blue colors represent high and low levels of information output, respectively, and the ATP molecules are shown as black/grey spheres. (D) a magnified view of the allosteric cluster that is adjacent to the docking site of the R domain.

Figure 4. Synchronized movements of residues and domains of CFTR.
(A) Shown is a 2-D map of the dynamic cross-correlations between all residues of CFTR during the conformational transiting from the dephosphorylated ATP-free state to the phosphorylated ATP-bound state (PDB ID 5UAK and 6MSM, respectively). Red colors indicate strong positive correlation, meaning that the residues move in parallel vectors in space and time, blue colors indicate negative correlations. Domain boundaries are indicated by white lines. (B, C) Cartoon representation of CFTR where each residue is colored according to its dynamic cross-correlations with the ATP binding residues (black spheres) of NBD1 (B) or NBD2 (C).

Figure 5. Trajectory and sequence of allosteric transduction in CFTR.
Dynamic cross-correlations were calculated using a time delay (τ) of 16 cycles out of 50 required to complete the conformational transition between the dephosphorylated ATP-free and phosphorylated ATP-bound states. Red and blue colors indicate high and low correlations, respectively.
Steps I-IV: Each residue is colored according to the degree of the correlation of its movement at time ($t + \tau$) relative to the movement at time (t) of S1251 of NBD2 (Step 1), D110 of ECL1 (Step 2), T460 of NBD1 (Step 3), and the gating residue T338 (Step 4).
Center (large) panel: shown is a summary of the allosteric trajectory, originating from NBD2 (green) to ICL2/3 and ECL1 (magenta), from ECL1 to NBD1 (blue), from NBD1 to the permeation pathway helices (orange), and finally back to NBD2.

Figure 6. Allosteric modulation of CFTR by drugs. (A) Shown is the amount of information TRANSMITTED (TE score) by each residue of CFTR ΔF508 bound to Trikafta (PDB ID 8EIQ) calculated using the ten most collective GNM modes (solid black trace). The binding sites of ivacaftor, elexacaftor, and lumacaftor/tezacaftor are shown as magenta, green, or cyan spheres, respectively. (B) Cartoon representations of ivacaftor bound CFTR (PDB ID 6O2P), with ivacaftor and ATP shown as magenta and black/grey spheres, respectively. Residues are colored according to the amount of entropy RECEIVED from the ivacaftor binding residue F312 (C) Shown is the amount of entropy RECEIVED by each residue of CFTR from the ATP binding residue G551 (blue trace) or the ivacaftor binding residue F312 (orange trace).
(D-F) Same as B, only the residues are colored according to the amount of entropy RECEIVED from the gating residue F337 (D), the ATP binding residue G551 (E), and the sum of the entropy received from both G551 and F337 (F).

Figure 7. Model of CFTR's conformational changes and their allosteric control. CFTR is shown in a grey space-filling model, the regulatory (R) domain as a red ribbon, and ATP in shades of green. Following its phosphorylation, the R domain dislodges from the NBDs dimer interface (Step I of the conformational change). This allows the ATP binding residues to fluctuate and sample conformers near the NBDs-open state (Step II), some of which provide an initial ATP binding site. In step III, this subset of conformers is selected by ATP to initiate binding (*i.e.,* ensemble selection), which then proceeds to complete closure of the NBDs by an induced fit mechanism. In step IV, the channel bursts to the conducting state by a dynamic mechanism, without a major conformational change.