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

For correspondence bental@tauex.tau.ac.il, halilogt@boun.edu.tr, lewinson@technion.ac.il

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: CFTR drugs bind at the periphery of the protein yet affect the gating residues that lie at the center of it. Herein, we combined two computational approaches; Anisotropic Normal Mode-Langevin dynamics (ANM-LD) and Transfer Entropy (TE) and investigated the allosteric communications network of CFTR. The results are in excellent agreement with experimental observations 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 observe that the degenerate and catalytic ATP sites asymmetrically contribute to the allosteric communication, and that the catalytic site provides the greater allosteric input. 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 identify a hitherto unknown allosteric hotspot near the docking site of the phosphorylated R domain, providing a molecular basis for its phosphorylation dependent excitatory role. This study uncovers the molecular basis of allosteric

connectivity in CFTR and reveals a novel allosteric hotspot that can serve as a target for the development of novel therapeutics.

Introduction. Cystic fibrosis (CF) is a pleiotropic disease that affects the epithelia of organs such as the lungs, intestinal tract, and pancreas. It is one of the most common hereditary disorders, especially in Caucasian populations, affecting about one out of every 3,000 newborns (1). Despite extensive progress in therapeutic approaches CF patients still suffer of acute morbidity and shortened 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 effects the functional properties of another (often distant) site (20–23). 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, 13). 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 (24). 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 (25). 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 (26). 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 (27, 28) yet increase the open probability (Po) of the channel (29, 30). These and other studies (31–33) 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 applied two complementary computational approaches; ANM-LD simulations which combine Anisotropic Network Model (ANM) (34–37) and Langevin Dynamics (LD) (38), and an information theoretic based method; Gaussian network model-Transfer Entropy (GNM-TE (39, 40)). The results align

very well with experimental data, identify novel allosteric hotspots in CFTR, and provide novel insight into the allosteric interactions in CFTR.

Results

Transfer Entropy calculations identify allosteric hotspots in CFTR. The concept of Transfer Entropy (TE) is derived from information theory and is used to measure the amount of information transferred between two processes (39). In the context of proteins' conformational changes, TE can be used to determine whether two residues are allosterically connected (40). 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 an excellent explanation about the correlation between TE and allostery see (41). In the present work, we performed TE calculations based on Gaussian Network Model (GNM), which considers harmonic interactions between residue pairs (39, 42). 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.

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 most often coincide with functional epitopes, *i.e.*, residues that are involved in ligand binding, catalysis, or serve as conformational hinges (21, 43). 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 identified the residues that are the main allosteric

sources in non-phosphorylated ATP-free human CFTR (PDB ID 5UAK, (8)). We then compared the locations of the GNM-TE identified allosteric sources to the positions of the 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, 25, 31). 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: Using the ten most collective GNM modes we 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 S1A, 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 determined peaks and that of the functionally essential residues (Figure S1B-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 (13). 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 S2A). This co-segregation was observed whether we considered only exact matches (Figure S2A), or whether we included also first- or second- coordination sphere interactions (Figures S2B and S2C, respectively). Next, we considered a more challenging reference set of residues, one which includes the 93 CFTR miss-sense 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 (44). 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 only the cluster of N-terminal residues (S13, L15, W60, E56, E60, P67, G85) that are known to be involved in interactions with cellular filamins/ER retention (45). As shown in Figure S3A we observed considerable association between the positions of the 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 S3B, the random distribution is roughly normal with a mean of 4.5 incidents of exact matches. As shown in Figure S3B 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 S1 and S2) 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 S3C and S3D, respectively). In further testimony 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 (24). 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, S1, S2, S3, and 2A) 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 obtained 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 (see shortly below). 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 shortly below).

As explained in the methods section and elsewhere (42, 46) GNM decomposes a complex motion (such as a conformational change, or a person walking) to discrete dynamic building blocks, 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 "collective modes"), and other modes that involve local conformational adjustments and a smaller number of residues (the less collective modes). The results shown in Figure 2A were obtained using the 10 most collective GNM modes. However, during the calculations the entropic contribution of the most collective modes may overshadow that of the less collective ones, which may nevertheless have important functional roles (for more details see (40)). To circumvent this problem, one may repeat the calculations while omitting the most collective GNM modes, enabling the detection of latent allosteric interactions that may otherwise be hidden in the global fluctuations. This approach (*i.e.*, breaking a complex dynamic behavior to different subsets of its components) improves the identification of residues which serve as entropy sources. Such residues are of great interest as they are often of functional importance (see shortly below). Therefore, to reveal latent allosteric interactions that may be hidden in the global fluctuations we repeated the TE calculations after removing one or two of the most collective modes. Several interesting features were revealed by these calculations. Upon removal of the most collective GNM mode, a new cluster of residues that serve as allosteric sources is 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 (27) were also revealed as entropy sources by removal of the most collective GNM mode (Figure 2B), demonstrating the

utility of this approach. Removal of the next most collective 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 ((47, 48) and see also later).

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 (see shortly below).

ATP binding and phosphorylation strengthens 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, (13)). We observed that phosphorylation and binding of ATP greatly rewires the overall allosteric connectivity in CFTR (Figure S4 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.

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 most collective 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 (47, 48). 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 (49, 50).

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 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 (34, 38). We simulated the transition between non-phosphorylated ATP-free and phosphorylated ATP-bound human CFTR (8, 13). 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 (42), a lipid-linked oligosaccharide flippase that adopts the canonical ABC exporter 3-D fold (51). These observations suggest that, as interfered from the cryo-EM structures of CFTR (8, 13), 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 (Figure 4C) which is known to stabilize the conducting state of channel (52).

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 impose 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.

Free energy calculations (25) and single molecule smFRET (24) suggest that the allosteric signal and conformational change starts at the NBDs. Along these lines, 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 S4A) 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 S5 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 (24).

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 S5 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 S5 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 S5 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 concomitant with a greatly increased life expectancy (53). 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 (54, 55). The potentiator ivacaftor (VX-770) enhances the activity of conductance-defective mutants by increasing the open probability (Po) of the channel (24, 27, 29), and the dual-function type III corrector elexacaftor (VX-445) both stabilizes unstable variants and increases channel open probability (30, 56). The recently published structures of WT and ΔF508CFTR bound to various drugs (28) enabled to us to investigate the allosteric effects of the drugs. First, we conducted GNM-based Transfer Entropy analysis on ΔF508CFTR bound to Trikafta (PDB ID 8EIQ, (28)), 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 (Figure 6A). 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 (55). 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) (27, 29, 57, 58). 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 (27). In addition, unlike the folding correcting drugs, ivacaftor does not induce a significant conformational change in ΔF508 CFTR (28). 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 6F, 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 0.953). These observations suggest that the drug 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 6C-E. The gating region (*e.g.*, gating residue F337) (13)) 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 6C). 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 6D). Superimposition of these effects, which represents the combined allosteric signaling of the ion-coordinating and ATP-binding residues (shown in Figure 6E) 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

In CFTR, the conformational changes at the cytosolic NBDs are transduced across the entire protein to the transmembrane gating residues. The molecular network that enables such long-range allosteric communication is not fully understood, and herein we sought to understand this phenomenon.

The Transfer Entropy (TE) analysis used in this work is based on GNM, which reduces a highly complex protein polymer such as CFTR to a network of balls and springs and considers only the harmonic interactions between the α-carbons (39). Despite this oversimplification, and without any input of functional information, GNM-Transfer Entropy calculations clearly distinguish 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 (Figure 2). We also find a remarkable correlation between the residues identified as main allosteric sources and the positions of functionally essential residues (Figure S1), ATP-binding residues (Figure S2), and disease-causing mutations (Figure S3). 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. At first glance, the allosteric connectivity networks of the ATP-free and -bound conformations seem very different (Figure S4). However, a closer examination reveals that binding of ATP mainly 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 (Figure 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 is to modulate conformational dynamics that are present to different extents in the different conformations. This differs from the role of ATP in *bona-fide* ATP transporters where the energy of ATP binding and hydrolysis is required to stabilize high-energy intermediates to enable bypassing of large energetic barriers.

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 one once phosphorylated (49, 59, 60). 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 (Gaussian Network Model) (46, 61) based Transfer Entropy (TE) formulation (Hacisuleyman and Erman, 2017) 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 (39)

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

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 (39, 46).

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.

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 (62), collectivity (K) values of residues in information transfer are calculated based on the positive net TE values (40) as

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

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

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

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 (40).

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

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

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) (13) human CFTR was simulated using the ANM-LD approach (42, 63). 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 (34). 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 (64).

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

$$Z = \frac{X - \mu}{\sigma} \quad (10)$$

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.

Visualization

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

Acknowledgements

The authors declare no conflicts of interest.

TH; The Scientific and Technological Research Council of Turkey (TUBITAK) with grant numbers 119F392

TH, NBT, OL; NATO Science for Peace Program (G5685)

OL; Israel Science Foundation (1006/18)

References

1. P. M. Farrell, The prevalence of cystic fibrosis in the European Union. *J. Cyst. Fibros.* **7**, 450–453 (2008).
2. K. A. McBennett, P. B. Davis, M. W. Konstan, Increasing life expectancy in cystic fibrosis: Advances and challenges. *Pediatr. Pulmonol.* **57**, S5 (2022).
3. Cftr2.org, The Clinical and Functional TRanslation of CFTR (CFTR2) (2011) (December 6, 2021).
4. J. V. Fahy, B. F. Dickey, Airway Mucus Function and Dysfunction. *N. Engl. J. Med.* **363**, 2233–2247 (2010).
5. M. Dean, A. Rzhetsky, R. Allikmets, The Human ATP-Binding Cassette (ABC) Transporter Superfamily. *Genome Res.* **11**, 1156–1166 (2001).

6. D. C. Rees, E. Johnson, O. Lewinson, ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* **10**, 218–227 (2009).
7. D. C. Gadsby, P. Vergani, L. Csanády, The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* **440**, 477–483 (2006).
8. F. Liu, Z. Zhang, L. Csanády, D. C. Gadsby, J. Chen, Molecular Structure of the Human CFTR Ion Channel. *Cell* **169**, 85-95.e8 (2017).
9. Z. Zhang, J. Chen, Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator. *Cell* **167**, 1586-1597.e9 (2016).
10. M. R. Picciotto, J. A. Cohn, G. Bertuzzi, P. Greengard, A. C. Nairn, Phosphorylation of the cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **267**, 12742–12752 (1992).
11. J. A. Tabcharani, X. B. Chang, J. R. Riordan, J. W. Hanrahan, Phosphorylation-regulated Cl- channel in CHO cells stably expressing the cystic fibrosis gene. *Nature* **352**, 628–631 (1991).
12. M. Mense, *et al.*, In vivo phosphorylation of CFTR promotes formation of a nucleotide-binding domain heterodimer. *EMBO J.* **25**, 4728–4739 (2006).
13. Z. Zhang, F. Liu, J. Chen, Molecular structure of the ATP-bound, phosphorylated human CFTR. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12757–12762 (2018).
14. M. Hohl, C. Briand, M. G. Grütter, M. A. Seeger, Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. *Nat. Struct. Mol. Biol.* **19**, 395–402 (2012).
15. L. Csanády, P. Vergani, D. C. Gadsby, Structure, gating, and regulation of the CFTR anion channel. *Physiol. Rev.* **99**, 707–738 (2019).
16. T. Stockner, R. Gradisch, L. Schmitt, The role of the degenerate nucleotide binding site in type I ABC exporters. *FEBS Lett.* **594**, 3815–3838 (2020).
17. Z. L. Johnson, J. Chen, ATP Binding Enables Substrate Release from Multidrug Resistance Protein 1. *Cell* **172**, 81-89.e10 (2018).
18. Y. Kim, J. Chen, Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. *Science* **359**, 915–919 (2018).
19. P. T. Nguyen, J. Y. Lai, A. T. Lee, J. T. Kaiser, D. C. Rees, Noncanonical role for the binding protein in substrate uptake by the MetNI methionine ATP Binding Cassette

(ABC) transporter. *Proc. Natl. Acad. Sci.* **115**, E10596–E10604 (2018).

20. A. Kessel, N. Ben-Tal, *Introduction to Proteins: Structure, Function, and Motion*, 2018th Ed. (CRC Press, 2018).
21. J. Liu, R. Nussinov, Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLoS Comput. Biol.* **12** (2016).
22. Z. N. Gerek, S. B. Ozkan, Change in allosteric network affects binding affinities of PDZ domains: Analysis through perturbation response scanning. *PLoS Comput. Biol.* **7** (2011).
23. P. Campitelli, T. Modi, S. Kumar, S. Banu Ozkan, The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys.* **49**, 267–288 (2020).
24. J. Levring, *et al.*, CFTR function, pathology and pharmacology at single-molecule resolution. *Nature* (2023) https:/doi.org/10.1038/s41586-023-05854-7 (March 26, 2023).
25. B. Sorum, D. Czégé, L. Csanády, Timing of CFTR Pore Opening and Structure of Its Transition State. *Cell* **163**, 724–733 (2015).
26. W. Wang, *et al.*, ATP-independent CFTR channel gating and allosteric modulation by phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3888–3893 (2010).
27. F. Liu, *et al.*, Structural identification of a hotspot on CFTR for potentiation. *Science (80-.).* **364**, 1184–1188 (2019).
28. K. Fiedorczuk, J. Chen, Molecular structures reveal synergistic rescue of D508 CFTR by Trikafta modulators. *Science (80-.).* **378**, 284–290 (2022).
29. P. D. W. Eckford, C. Li, M. Ramjeesingh, C. E. Bear, Cystic fibrosis transmembrane conductance regulator (CFTR) potentiator VX-770 (ivacaftor) opens the defective channel gate of mutant CFTR in a phosphorylationdependent but ATP-independent manner. *J. Biol. Chem.* **287**, 36639–36649 (2012).
30. C. A. Shaughnessy, P. L. Zeitlin, P. E. Bratcher, Elexacaftor is a CFTR potentiator and acts synergistically with ivacaftor during acute and chronic treatment. *Sci. Rep.* **11** (2021).
31. B. Sorum, B. Töröcsik, L. Csanády, Asymmetry of movements in CFTR's two ATP sites during pore opening serves their distinct functions. *Elife* **6** (2017).
32. D. Scholl, *et al.*, A topological switch in CFTR modulates channel activity and sensitivity to unfolding. *Nat. Chem. Biol.* **17**, 989–997 (2021).
33. S. Wei, *et al.*, Conserved allosteric hot spots in the transmembrane domains of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) channels and Multidrug

Resistance Protein (MRP) pumps. *J. Biol. Chem.* **289**, 19942–19957 (2014).

34. A. R. Atilgan, *et al.*, Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **80**, 505–515 (2001).
35. Y. Wang, A. J. Rader, I. Bahar, R. L. Jernigan, Global ribosome motions revealed with elastic network model. *J. Struct. Biol.* **147**, 302–314 (2004).
36. O. Keskin, S. R. Durell, I. Bahar, R. L. Jernigan, D. G. Covell, Relating molecular flexibility to function: A case study of tubulin. *Biophys. J.* **83**, 663–680 (2002).
37. O. Keskin, Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: A case study of antibodies. *BMC Struct. Biol.* **7** (2007).
38. A. Brünger, C. L. Brooks, M. Karplus, Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105**, 495–500 (1984).
39. A. Hacisuleyman, B. Erman, Causality, transfer entropy, and allosteric communication landscapes in proteins with harmonic interactions. *Proteins Struct. Funct. Bioinforma.* **85**, 1056–1064 (2017).
40. B. Altintel, B. Acar, B. Erman, T. Haliloglu, Subsets of Slow Dynamic Modes Reveal Global Information Sources as Allosteric Sites. *J. Mol. Biol.* **434** (2022).
41. A. Hacisuleyman, B. Erman, Entropy Transfer between Residue Pairs and Allostery in Proteins: Quantifying Allosteric Communication in Ubiquitin. *PLoS Comput. Biol.* **13** (2017).
42. B. Acar, *et al.*, Distinct Allosteric Networks Underlie Mechanistic Speciation of ABC Transporters. *Structure* **28**, 651-663.e5 (2020).
43. S. J. Wodak, *et al.*, Allostery in Its Many Disguises: From Theory to Applications. *Structure* **27**, 566–578 (2019).
44. Q. Chen, Y. Shen, J. Zheng, A review of cystic fibrosis: Basic and clinical aspects. *Anim. Model. Exp. Med.* **4**, 220–232 (2021).
45. W. R. Thelin, *et al.*, Direct interaction with filamins modulates the stability and plasma membrane expression of CFTR. *J. Clin. Invest.* **117**, 364–374 (2007).
46. T. Haliloglu, I. Bahar, B. Erman, Gaussian Dynamics of Folded Proteins. *Phys. Rev. Lett.* **79**, 3090–3093 (1997).
47. Z. Bozoky, *et al.*, Regulatory R region of the CFTR chloride channel is a dynamic integrator of phospho-dependent intra- and intermolecular interactions. *Proc. Natl. Acad.*

Sci. U. S. A. **110** (2013).

48. V. Chappe, T. Irvine, J. Liao, A. Evagelidis, J. W. Hanrahan, Phosphorylation of CFTR by PKA promotes binding of the regulatory domain. *EMBO J.* **24**, 2730–2740 (2005).
49. M. C. Winter, M. J. Welsh, Stimulation of CFTR activity by its phosphorylated R domain. *Nature* **389**, 294–296 (1997).
50. J. Ma, J. Zhao, M. L. Drumm, J. Xie, P. B. Davis, Function of the R domain in the cystic fibrosis transmembrane conductance regulator chloride channel. *J. Biol. Chem.* **272**, 28133–28141 (1997).
51. C. Perez, *et al.*, Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature* **524**, 433–438 (2015).
52. D. T. Infield, G. Cui, C. Kuang, N. A. McCarty, Positioning of extracellular loop 1 affects pore gating of the cystic fibrosis transmembrane conductance regulator. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **310**, L403–L414 (2016).
53. J. P. Clancy, Rapid therapeutic advances in CFTR modulator science. *Pediatr. Pulmonol.* **53**, S4–S11 (2018).
54. F. Van Goor, *et al.*, Correction of the F508del-CFTR protein processing defect in vitro by the investigational drug VX-809. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18843–18848 (2011).
55. K. Fiedorczuk, J. Chen, Mechanism of CFTR correction by type I folding correctors. *Cell* **185**, 158-168.e11 (2022).
56. G. Veit, C. Vaccarin, G. L. Lukacs, Elexacaftor co-potentiates the activity of F508del and gating mutants of CFTR. *J. Cyst. Fibros.* **20**, 895–898 (2021).
57. H. Yu, *et al.*, Ivacaftor potentiation of multiple CFTR channels with gating mutations. *J. Cyst. Fibros.* **11**, 237–245 (2012).
58. S. Hadida, *et al.*, Discovery of N -(2,4-Di- tert -butyl-5-hydroxyphenyl)-4-oxo-1,4-dihydroquinoline-3-carboxamide (VX-770, Ivacaftor), a potent and orally bioavailable CFTR potentiator. *J. Med. Chem.* **57**, 9776–9795 (2014).
59. L. S. Ostedgaard, O. Baldursson, M. J. Welsh, Regulation of the Cystic Fibrosis Transmembrane Conductance Regulator Cl- Channel by Its R Domain. *J. Biol. Chem.* **276**, 7689–7692 (2001).
60. L. Csanády, *et al.*, Severed channels probe regulation of gating of cystic fibrosis

transmembrane conductance regulator by its cytoplasmic domains. *J. Gen. Physiol.* **116**, 477–500 (2000).

61. I. Bahar, A. R. Atilgan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **2**, 173–81 (1997).
62. R. Brüschweiler, Collective protein dynamics and nuclear spin relaxation. *J. Chem. Phys.* **102**, 3396–3403 (1995).
63. M. Yang, *et al.*, Single-molecule probing of the conformational homogeneity of the ABC transporter BtuCD. *Nat. Chem. Biol.* **14**, 715–722 (2018).
64. and P. A. K. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, AMBER 11 (2010).

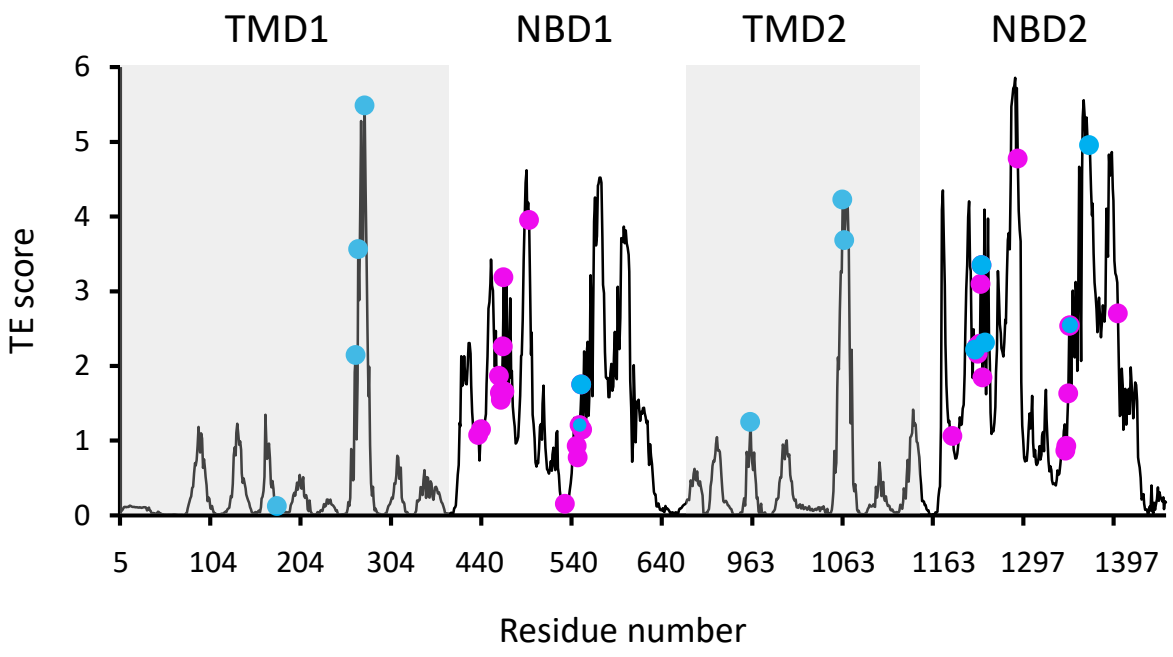
TMD1
NBD1
TMD2
NBD2
TE score
0
1
2
3
4
5
6
5
104
204
304
440
540
640
963
1063
1163
1297
1397
Residue number

Figure 1

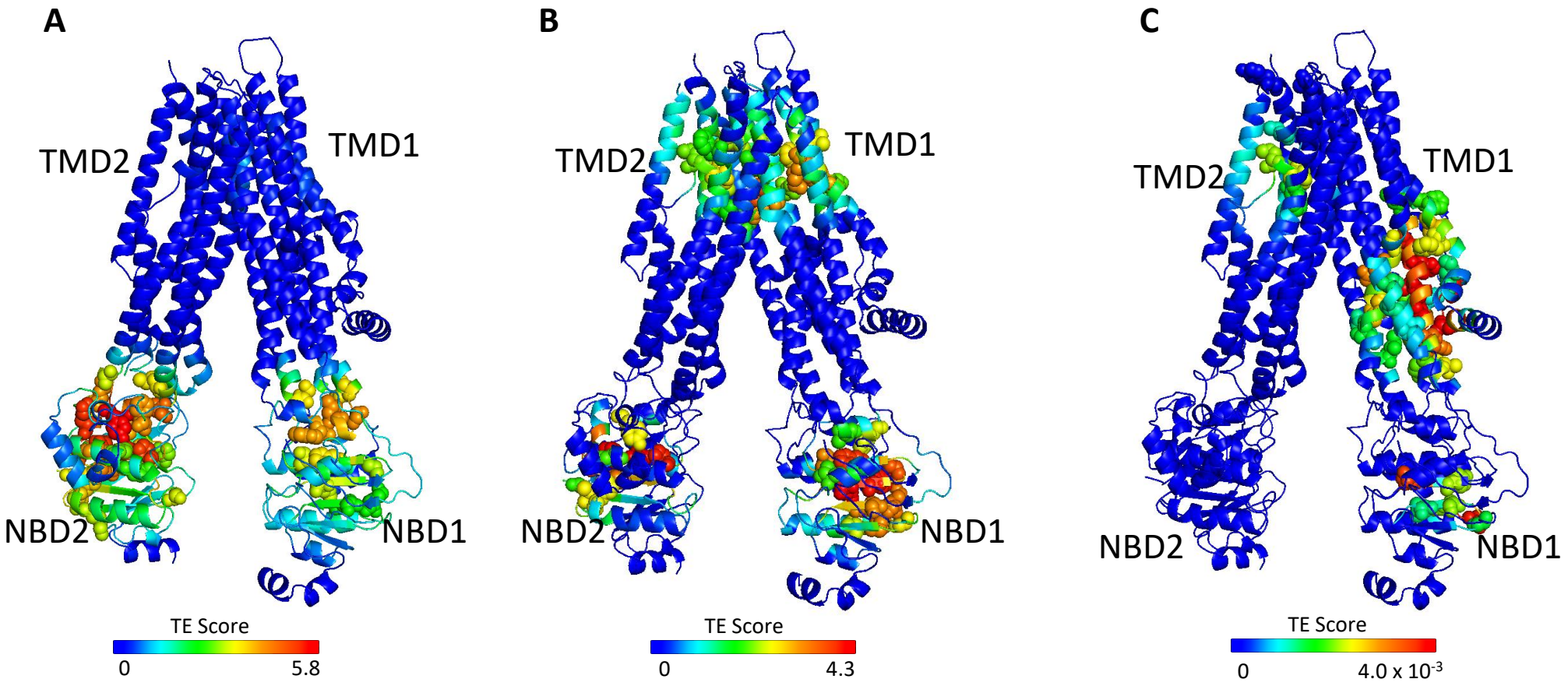
A
TMD2
TMD1
NBD2
NBD1
TE Score
0
5.8
B
TMD2
TMD1
NBD2
NBD1
TE Score
0
4.3
C
TMD2
TMD1
NBD2
NBD1
TE Score
0
4.0 x 10-3

Figure 2

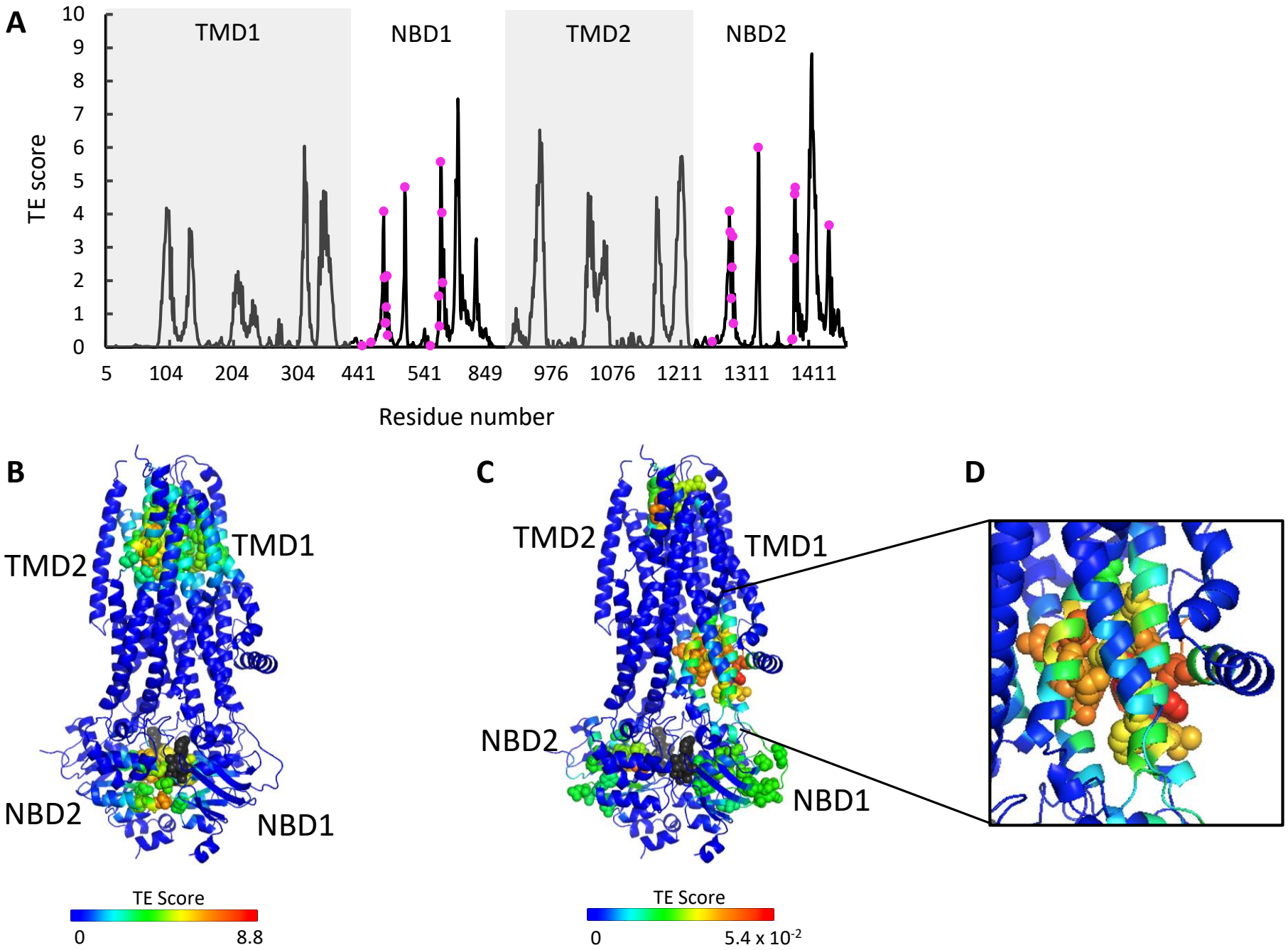
A
TMD1
NBD1
TMD2
NBD2
TE score
10
9
8
7
6
5
4
3
2
1
0
5
104
204
304
441
541
849
976
1076
1211
1311
1411
Residue number
B
TMD2
TMD1
NBD2
NBD1
TE Score
0
8.8
C
TMD2
TMD1
NBD2
NBD1
TE Score
0
5.4 x 10-2
D

Figure 3

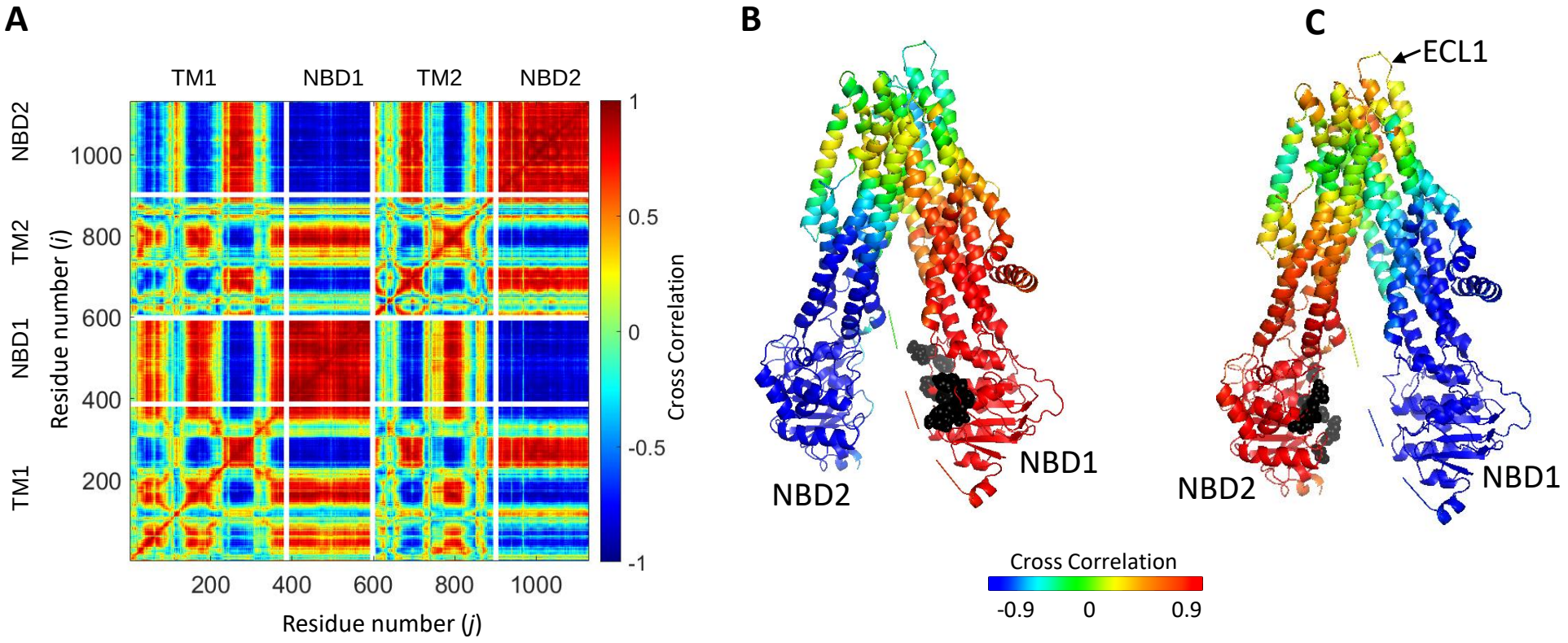
A
TM1
NBD1
TM2
NBD2
NBD2
TM2
NBD1
TM1
Residue number (i)
1000
800
600
400
200
200
400
600
800
1000
Residue number (j)
1
0.5
0
-0.5
-1
Cross Correlation
B
NBD2
NBD1
C
ECL1
NBD2
NBD1
Cross Correlation
-0.9
0
0.9

Figure 4

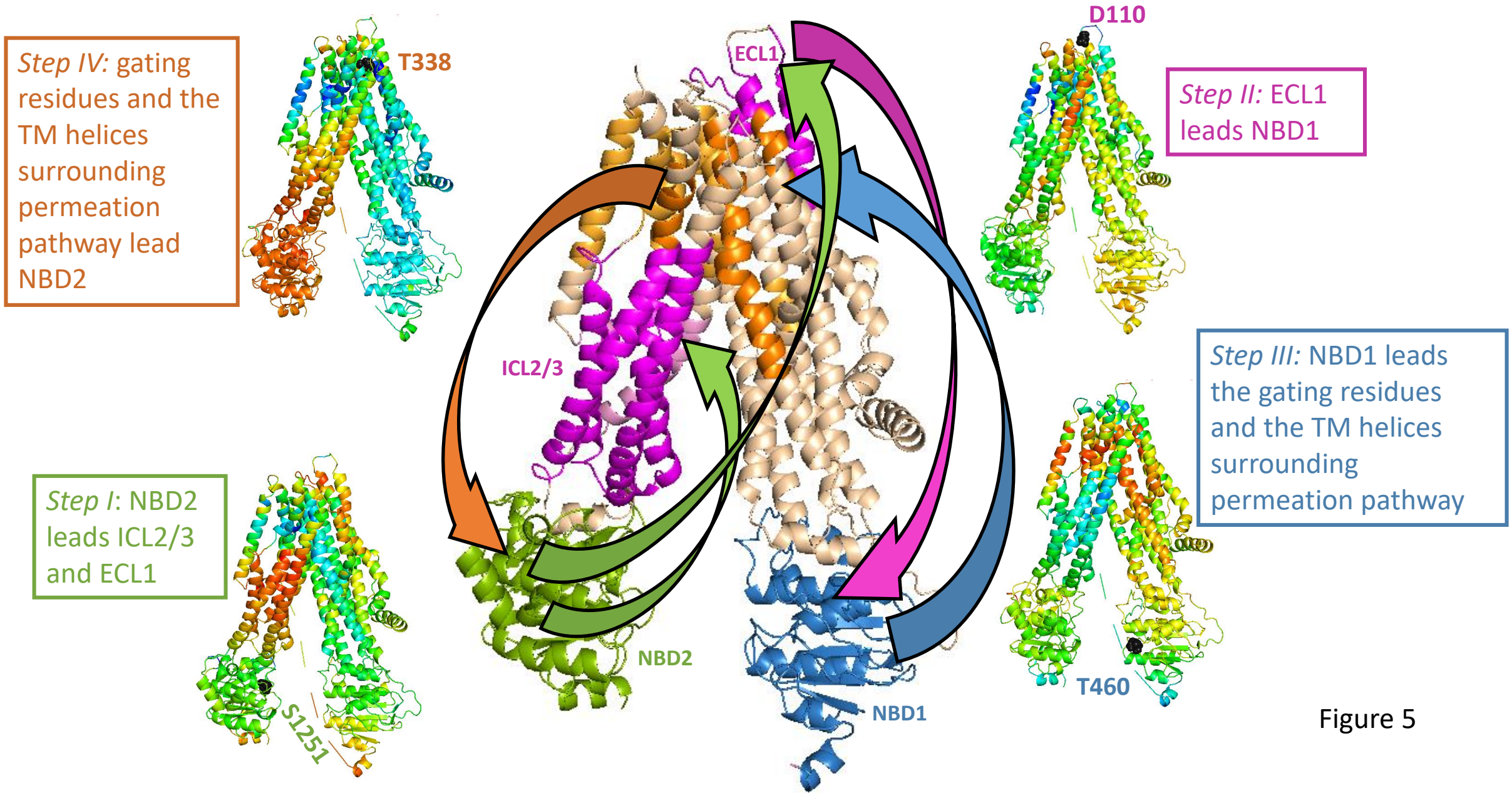
Step IV: gating residues and the TM helices surrounding permeation pathway lead NBD2
T338
ECL1
D110
Step II: ECL1 leads NBD1
ICL2/3
Step III: NBD1 leads the gating residues and the TM helices surrounding permeation pathway
Step I: NBD2 leads ICL2/3 and ECL1
S1251
NBD2
NBD1
T460

Figure 5

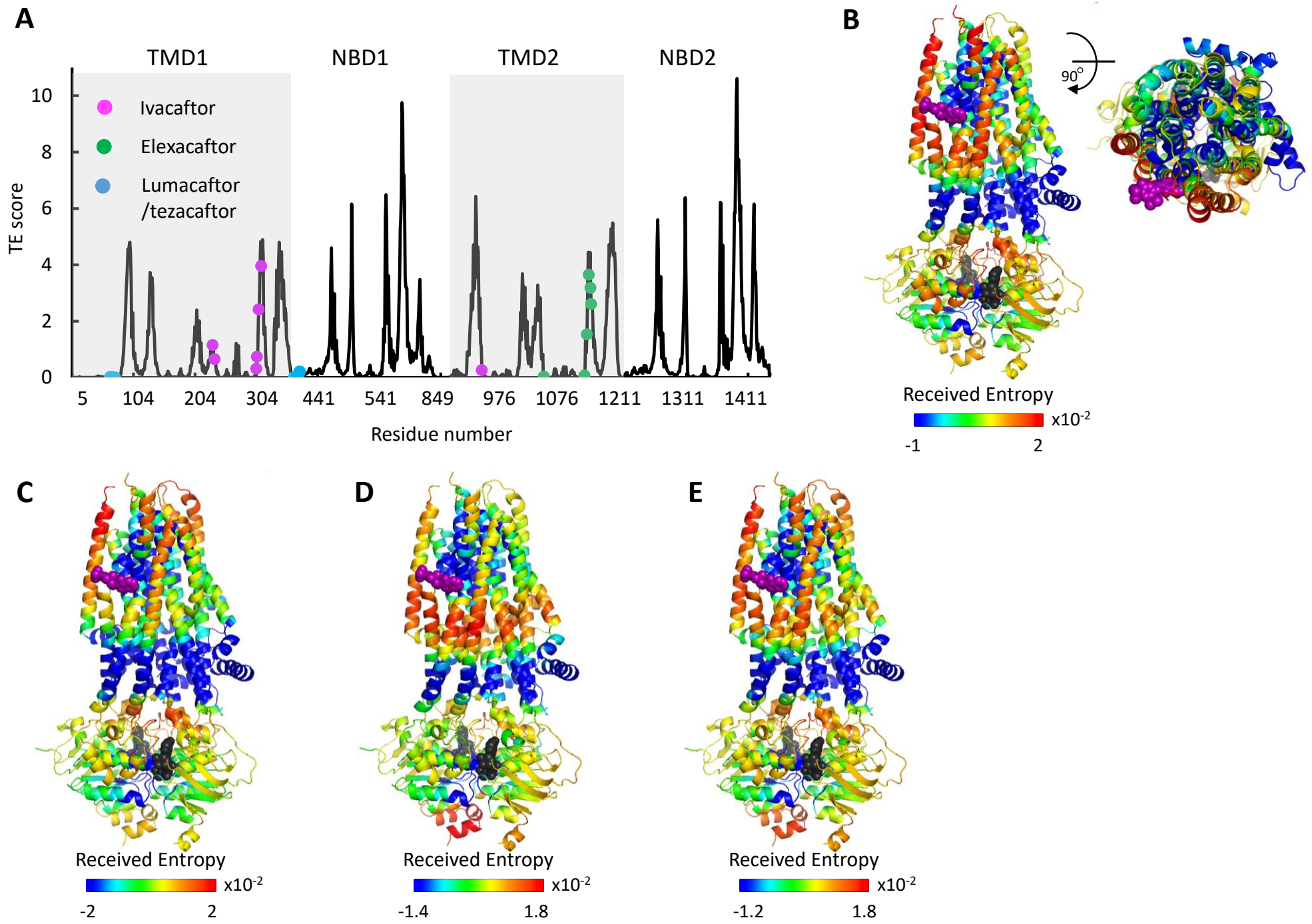
A
TMD1
NBD1
TMD2
NBD2
Ivacaftor
Elexacaftor
Lumacaftor
/tezacaftor
TE score
0
2
4
6
8
10
5
104
204
304
441
541
849
976
1076
1211
1311
1411
Residue number
B
90°
Received Entropy
x10-2
-1
2
C
Received Entropy
x10-2
-2
2
D
Received Entropy
x10-2
-1.4
1.8
E
Received Entropy
x10-2
-1.2
1.8

Figure 6

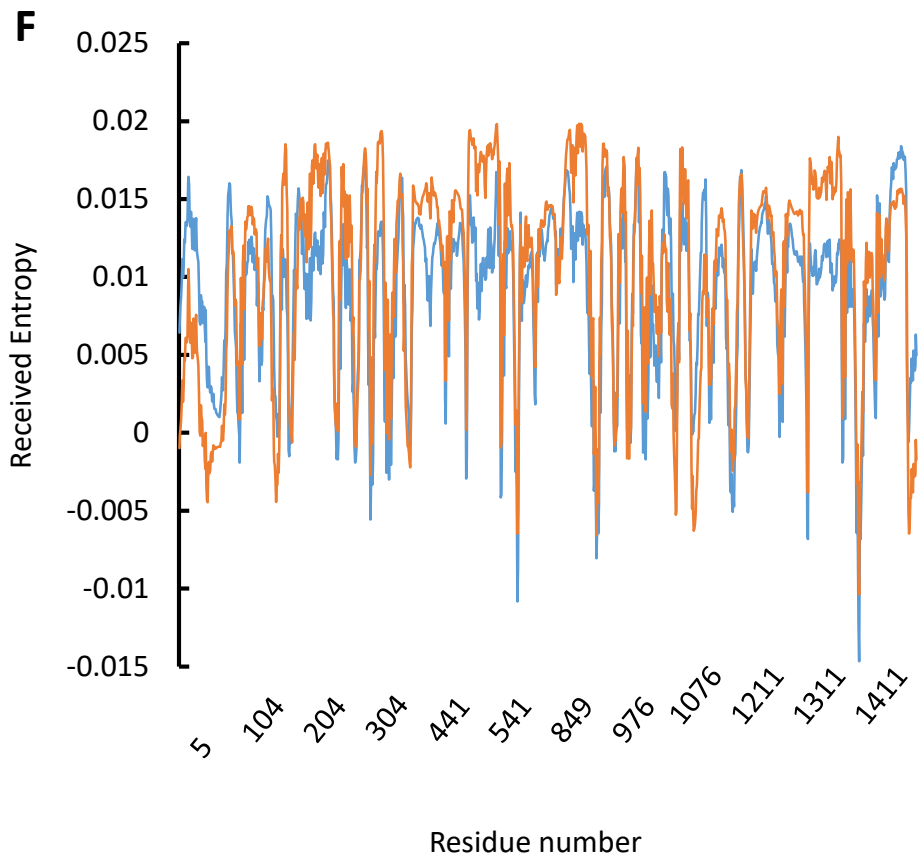
F
0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Received Entropy
5
104
204
304
441
541
849
976
1076
1211
1311
1411
Residue number

Figure 6

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-E) 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 (B), the gating residue F337 (C), the ATP binding residue G551 (D), and the sum of the entropy transferred from both G551 and F337 (E). In (F), 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).