Comprehensive mapping of Cystic Fibrosis mutations to CFTR protein identifies mutation clusters and molecular docking predicts corrector binding site

Running Title: Mutation mapping and corrector docking to CFTR protein

Authors:

Steven V. Molinski[1], Vijay M. Shahani[1], Adithya S. Subramanian[1], Stephen S. MacKinnon[1], Geoffrey Woollard[1], Marcon Laforet[1], Onofrio Laselva[2], Leonard D. Morayniss[1], Christine E. Bear[2,3,4] and Andreas Windemuth[1,#]

Affiliations:

[1]Cyclica Inc., Toronto, Ontario M5C 1C4, Canada

[2]Programme in Molecular Structure & Function, Research Institute, Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada

[3]Department of Physiology, University of Toronto, Toronto, Ontario M5S 1A8, Canada

[4]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada

[#]To whom correspondence should be addressed:
Andreas Windemuth, PhD
Cyclica Inc., 810-18 King St. East
Toronto, Ontario, Canada, M5C 1C4
Tel: 857-626-4888 | Email: andreas.windemuth@cyclicarx.com

Key Words: Cystic Fibrosis; CFTR; *in silico*; structural pharmacogenomics; molecular docking; correctors

Abstract

Background: Cystic Fibrosis (CF) is caused by mutations in the *CFTR* gene, of which over 2000 have been reported to date. Mutations have yet to be analyzed in aggregate to assess their distribution across the tertiary structure of the CFTR protein, an approach that could provide valuable insights into the structure-function relationship of CFTR. In addition, the binding site of Class I correctors (VX-809, VX-661, C18) is not well understood.

Methods: Exonic *CFTR* mutations and mutant allele frequencies described in three curated databases (ABCMdb, CFTR1 and CFTR2, comprising >130,000 data points) were mapped to two different structural models: a homology model of full-length CFTR protein in the open-channel state, and a cryo-electron microscopy core-structure of CFTR in the closed-channel state. Immunoblotting confirmed the approximate binding site of Class I correctors, and molecular docking generated binding poses for their complex with the cryo-electron microscopy structure.

Results: Residue positions of six high-frequency mutant *CFTR* alleles were found to spatially co-localize in CFTR protein, and a significant cluster was identified at the NBD1:ICL4 interdomain interface. Further, Class I correctors VX-809, VX-661 and C18 were shown to act via a similar mechanism *in vitro*, and a putative multi-domain corrector binding site near residues F374-L375 was predicted *in silico*.

Conclusions: Our results confirm the significance of interdomain interfaces as susceptible to disruptive mutation, and identify a putative corrector binding site. The structural pharmacogenomics approach of mapping mutation databases to protein models shows promise for facilitating drug discovery and personalized medicine for monogenetic diseases.

Introduction

Cystic Fibrosis (CF) is caused by mutations in the *cystic fibrosis transmembrane conductance regulator* (*CFTR/ABCC7*) gene which lead to defective biosynthesis, trafficking and/or activity of the CFTR protein[1,2]. CFTR is a chloride channel, comprised of five structural domains: membrane-spanning domain 1 (MSD1), nucleotide-binding domain 1 (NBD1), a regulatory (R) domain, MSD2 and NBD2. Together these domains cooperate to facilitate fluid transport and surface hydration across epithelial cells of the body's tubular organs (e.g. lungs, intestines)[2,3]. Loss of CFTR expression and/or activity via mutation leads to a severe disease phenotype, where inflammation and infection within the lungs and intestines run rampant, and thus these organs are unable to sufficiently exchange gases and absorb nutrients, respectively[2–4]. To date, over 2000 disease-causing *CFTR* mutations have been reported, most of which are rare. The most common, F508del, is present in approximately 70% of the worldwide CF population (estimated to be 100,000); ~5% have G551D, ~5% G542X, and the remaining ~20% have one of the ~2000 other mutations[5–8]. These mutations are catalogued across three databases (ABCMdb[25], abcmutations.hegelab.org; CFTR1, genet.sickkids.on.ca; CFTR2[7], cftr2.org) and provide a comprehensive resource supporting the effort to further our understanding of the complex structure-function relationship of CFTR.

In recent years, much effort has been directed towards discovery and development of novel small molecule therapies for CF patients bearing the F508del and G551D mutations, since these mutations capture a large fraction of the patient population[9–13]. Accordingly, two small molecule therapies (Kalydeco®, comprised of the potentiator ivacaftor or VX-770, and Orkambi®, comprised of ivacaftor and the corrector lumacaftor or VX-809) have been developed for individuals with these mutations (G551D and F508del, respectively), as well as for those having a few other mutations with similar CFTR protein defects[14,15]. However, these discoveries as well as most other investigations typically explore *CFTR* mutations on a case-by-case basis, and do not consider the global mutation landscape as a whole. This is partly due to the abundance of the deposited data, which currently comprises a total count of >200,000 unique data points. Therefore, there is a need for studies which probe the landscape of known *CFTR* mutations in

order to uncover salient trends and provide additional insight into the biochemical, biophysical and/or evolutionary nature of this fatal genetic disease.

Accordingly, we employed a structural pharmacogenomics approach[16,17] to map all exonic mutations and mutant allele frequencies found in these databases of *CFTR* mutations onto three-dimensional structural models of the CFTR protein in both the open-channel[18] and closed-channel[19] states. The spatial distribution of the mapped mutations was analyzed to uncover salient trends that may inform on their mechanistic underpinnings. In addition, structural information from the experimentally determined closed-channel conformation[19] was combined with computational docking calculations to predict CFTR residues that contribute to the binding sites of current small molecule therapeutics, mainly Class I correctors: VX-809 (lumacaftor, a component of Orkambi® co-therapy), VX-661 (tezacaftor, the putative successor of lumacaftor) and C18 (an investigational compound).

The proposed structural pharmacogenomics approach is shown to yield useful insights into structure-function relationships that underlie CF, and to predict a potential drug binding pocket which may facilitate further stratification of patients based on their mutation-sensitive 'theratype' or responsiveness to certain small molecule therapies[13,20].

Materials and Methods

CFTR Mutation Databases

The three CFTR mutation databases used in this study are: ABCMdb[21] (abcmutations.hegelab.org), CFTR1 (www.genet.sickkids.on.ca) and CFTR2[7] (www.cftr2.org). Databases include mutations reported in the scientific literature (i.e. ABCMdb: containing validated disease-causing mutations, unvalidated disease-associated mutations, as well as experimental mutations used to study CFTR), those catalogued by physicians and scientists (i.e. CFTR1: validated and unvalidated mutations; CFTR2: validated mutations) as well as allele frequencies for reported mutations (CFTR2); databases were accessed between August 1st and August 31st, 2016.

CFTR Protein Tertiary Structures

We used a previously described and biochemically validated structural model of full-length CFTR in the open-channel state[18] (i.e. residues 1-1480, including the extensively modeled R-domain), as well as the recent cryo-electron microscopy (cryo-EM)-derived partial/core-structure of CFTR at 3.87 Å resolution in the closed-channel state[19] (lacking ~297 CFTR residues, including amino acids 1-4, 403-438, 646-843, 884-908, 1173-1206 and 1437-1480; PDB ID: 5UAK), as templates to map exonic *CFTR* mutations and allele frequencies. The homology model of CFTR in the open-channel state by Mornon *et al.* (2009)[18] was used, rather than Mornon *et al.* (2015)[20], as it is the only structural model of full-length CFTR (i.e. containing all 1480 residues); the latter model lacks the R-domain (defined as residues 650-843 in their study; 194 amino acids or ~13% of the protein) and subsequently would omit a substantial proportion of *CFTR* mutations described in the three CFTR mutation databases (i.e. ABCMdb, CFTR1, CFTR2). Further, use of a homology model was necessary since no experimental structure of CFTR in the open-channel state has been solved at sufficient resolution (i.e. the available low-resolution, ~9 Å structure[21] does not have residue-level detail required for this study). In addition, although of lower resolution than desired for molecular docking approaches, the experimental cryo-EM structure of CFTR in the closed (inactive) conformation was used as a template for molecular docking studies, since it is well-established that 'activating' ligands (e.g. correctors) bind to *at least* this conformational state[5,24,25].

Mapping CFTR Mutations onto the Protein Tertiary Structures

CFTR mutations from the above listed databases were mapped onto the open- and closed-channel structures of CFTR. The ABCMdb, CFTR1 and CFTR2 mutation databases provided a detailed list of 1,955, 1,308 and 235 exonic mutations (missense and insertions/deletions, i.e. non-intronic), respectively, which allowed for mutation counts (i.e. number of unique missense or nonsense mutations at a given amino acid position) and allele frequencies to be tabulated and further highlighted on the CFTR protein model using an automated 'paint-by-number' approach. PyMOL (version 1.7) was used to visualize CFTR protein structures.

Mutation Density using Average Interatomic Distance Measurements

To quantify the probability that the positions of the highest-frequency mutations (i.e. F508del, G542X, G551D, R553X W1282X, N1303K), known to be detrimental to structure and/or function, cluster within the CFTR tertiary structure, we computed the average interatomic distance (AID; mean value of distances between the C_α atoms of groups of amino acids) of residues F508, G542, G551, R553 W1282 and N1303, using the atomic coordinates of the full-length CFTR model of the active, open-channel state[18]. This distance was compared to the AID values of 1,000,000 randomly-selected six-residue subsets within the CFTR structural model to determine the statistical significance of the spatial clustering observed for the considered mutated positions.

Identification of Interdomain Interface Residues using Proximity Measurements

PyMOL (version 1.7) was used to visualize interdomain interfaces in the full-length homology-based structural model of CFTR, mainly NBD1 (residues 381-630), NBD2 (residues 1171-1480) and MSD2 (i.e. intracellular loop 4 or ICL4, residues 1039-1093). Residues at domain-domain interfaces were defined as residues of one domain with at least one atom within a 4 Å distance of any atom in the neighboring domain. Interdomain interface residues for NBD1:NBD2 and NBD1:ICL4 were further annotated by mutation frequency, and residues contained within interdomain interfaces were coloured according to mutation counts.

Molecular Docking of Pharmacological Correctors to CFTR

VX-809, VX-661 and C18 were docked to the apo form of human CFTR in the closed-channel state[19] (PDB ID: 5UAK) using the CROSSx molecular docking tool from the Ligand Express™ computational suite (Cyclica Inc., Toronto, Canada)[26]. The lowest-energy conformers of the three correctors considered in isolation were modeled, and found to possess up to 20 Å of maximum possible length. Accordingly, a 30x30x30-Å cube centered around residues F374-L375 (amino acids shown to directly interact with the Class I corrector VX-809[27]) was used to define potential binding site region in the cryo-EM structure of CFTR. The small molecules were then docked to this region using a protocol, which allowed for flexibility of the ligand but kept the side-chains of target

protein fixed (i.e. 'rigid' docking). A diverse ensemble of top-scoring sites centered around residues F374-L375 was visually inspected for chemical complementarity between the ligand and target moieties, and a reduced set of structures was docked allowing for flexible side-chains (i.e. 'flexible' docking). This protocol produced a total of 9 best docking poses for each of the corrector molecules. Assuming a common recognition mode for the three ligands, poses that align the pharmacophore features shared by the three corrector ligands, denoted here as the common scaffold (**Figure S4A**), were selected as the final docking solutions. PyMOL (version 1.7) was used to visualize molecular docking results.

Quantifying CFTR MSD1 Protein Abundance

Human embryonic kidney (HEK)-293 cells were transfected with CFTR MSD1 constructs (i.e. full-length: K381X, and the C-terminus truncation: D373X) using PolyFect Transfection Reagent according to the manufacturer's protocol (Qiagen) and as previously described[28]. HEK-293 cells expressing MSD1 constructs (~35 kDa) were maintained in DMEM (Wisent) supplemented with non-essential amino acids (Life Technologies) and 10% FBS (Wisent) at 37 ^{o}C with 5% CO_2 (HEPA incubator, Thermo Electron Corporation)[6,9]. These cells were grown at 37 ^{o}C (24 h) ± VX-809 (3 µM, Selleck Chemicals), VX-661 (1 µM, Selleck Chemicals) or C18 (6 µM, Cystic Fibrosis Foundation Therapeutics), lysed in modified radioimmunoprecipitation assay buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 7.4, 0.2% (v/v) SDS, and 0.1% (v/v) Triton X-100) containing a protease inhibitor cocktail (Roche) for 10 min, and soluble fractions were analyzed by SDS-PAGE on 4-12% gels. After electrophoresis, proteins were transferred to nitrocellulose membranes and incubated in 5% (w/v) milk, and MSD1 constructs were detected using the human CFTR-MSD1-specific (amino acids 27-34)[29] murine mAb MM13-4 (1:10000, University of North Carolina Chapel Hill, NC), horseradish peroxidase-conjugated goat anti-mouse IgG secondary antibody (1:5000) and exposure to film for 1-5 min. Calnexin was used as a loading control, and detected using a Calnexin-specific rabbit Ab (1:5000, Sigma-Aldrich), horseradish peroxidase-conjugated goat anti-rabbit IgG secondary antibody (1: 5000) and exposure to film for 1-5 min. Expression levels of

MSD1 constructs were quantitated by densitometry using ImageJ software (version 1.42Q, National Institutes of Health) and reported values are normalized to Calnexin expression levels.

Statistical Analysis

All data are represented as mean ± S.E.M. unless otherwise noted. Non-paired Student's t-tests, one-way analysis of variance (ANOVA) and Fisher's exact test were used as appropriate, and *P* values less than 0.05 were considered significant.

Results

CFTR Mutation Landscape and Structural Context

Three *CFTR* mutation databases: ABCMdb, CFTR1 and CFTR2, were explored to identify *CFTR* mutations for which the context in the CFTR tertiary structure could be investigated. To this end, only exonic CFTR mutations and their allele frequencies were retrieved from these databases and mapped to their corresponding residue positions in the protein structure. The CFTR structures used for this analysis were a previously described and biochemically validated homology-based model of full-length CFTR in the open-channel state[18], and a recently reported cryo-EM partial structure of CFTR in the closed-channel state[19]. Boundaries of the CFTR structural domains were defined as follows: MSD1 (1-380), NBD1 (381-630), R-domain (631-849), MSD2 (850-1170), NBD2 (1171-1480), with the numbers under parentheses indicating the first and last residues of the corresponding domain (**Figure S1A and S1B**).

ABCMdb curates the scientific literature and aggregates all reports for a given user-defined search regarding mutations in ATP-Binding Cassette (ABC) proteins, CFTR included. Importantly, this database contains validated disease-causing mutations, unvalidated disease-associated mutations, as well as those used to experimentally investigate CFTR structure and function. For example, data for F508 includes the disease mutation: F508del, as well as all possible amino acid substitutions, and therefore contains the greatest number of 'counts' (i.e. twenty) for a single residue position within the CFTR amino acid sequence. Accordingly, this database contains many

non-disease-causing mutations. However, such mutations still provide useful insight into the location within the CFTR sequence, which have been biochemically investigated in greater detail.

In total, 1,955 exonic *CFTR* mutations from the ABCMdb were mapped to the CFTR structures (**Figure S2**). Analysis of the structurally mapped mutations reveals that the R-domain, MSD2 and NBD2 are less-well studied compared with MSD1 and NBD1 of CFTR. A somewhat smaller number of mutations were mapped onto the CFTR protein structures from the other two databases: 1,308 mutations from CFTR1 (**Figure 1A and 1B**), and 235 mutations from CFTR2 (**Figure 2A and 2B**). The CFTR1 database represents physician and scientist catalogued mutations, and likely contains the oldest and most comprehensive collection of both validated disease-causing and unvalidated disease-associated *CFTR* mutations. The CFTR2 database contains a smaller collection of *CFTR* mutations, but the corresponding data are enriched for validated disease-causing variants, as they correlate genetic information to biochemical (*in vitro*) and clinical (*in vivo*) parameters and benchmarks (e.g. lung function and pancreatic sufficiency)[7]. Furthermore, the CFTR2 database reports allele frequencies for CF patients mainly in North America and Europe, and thus represents a substantial subset of the worldwide patient population. This dataset is useful for epidemiological studies, and provides a holistic view of the large-scale *CFTR* mutation landscape.

Importantly, the spatial distribution of mapped mutation counts differed between the ABCMdb, CFTR1 and CFTR2 mutation databases. Many mutations from the CFTR1 database mapped to all of the domains (except for the R-domain for which mutations were under-represented) (**Figure 1C**). A smaller collection of mapped mutations was contributed by the CFTR2 database, which had at least one dominant mutation in each of the first three domains (i.e. MSD1, NBD1 and the R-domain) (**Figure 2C**). Overall however, a greater proportion of known mutations from all three databases were located in the first two N-terminal domains of CFTR (MSD1 and NBD1), suggesting that these two domains play an important role in CFTR structure and function.

Mutant Allele Frequencies and Disease-Causing Mutations Cluster at Distinct CFTR Interdomain Interfaces

In total, 128,354 mutant alleles from the CFTR2 database were mapped onto the tertiary structure of CFTR (**Figure 3A**). Six amino acid positions (F508, G542, G551, R553, W1282, N1303) comprised a majority of the mutant alleles (110,585, ~86% of the total), including 98,735 alleles (~77%) for the F508del mutation alone, and upon further inspection we found that these six positions were spatially co-localized in the tertiary structure of CFTR (**Figure 3B and 3C**). This co-localization (AID of 18.75 Å) was found to be statistically significant (p-value ~0.0003) (**Figure 3D**; see Methods for details). More specifically, these six residues are part of two distinct interdomain interfaces (NBD1:NBD2 and NBD1:ICL4) which have been previously described as being integral to CFTR structure and function, and accordingly, represents a region of mutation clustering that is over-represented in the CF patient population when compared to all of the alleles reported in the CFTR2 database[30–33].

To assess the frequencies of disease-causing mutations at residue positions of the NBD1:NBD2 and NBD1:ICL4 interdomain interfaces, we first identified amino acids that contribute to these interfaces (see **Figure S3** and Methods for details). This revealed that the NBD1:NBD2 interface was more extensive (94 residues) than the NBD1:ICL4 interface (29 residues) as illustrated in **Figure 4 and Table 1**. We then computed the fraction of residues (i.e. ratio) with disease-causing mutations for the full-length CFTR protein to be 0.88 (i.e. 1,308 exonic *CFTR* mutations among 1,480 total residues, as reported in the CFTR1 database). The corresponding fractions for the NBD1:NBD2 and NBD1:ICL4 interfaces were 1.06 (i.e. 88 mutations among 94 residues, $p>0.05$) and 1.60 (i.e. 41 mutations among 29 residues, $p<0.05$), respectively (**Table 1**). These data indicate that these two interdomain interfaces have, on average, a larger fraction of mutated residues than full-length CFTR. Importantly, NBD1:ICL4 was found to harbour a statistically significant cluster of mutations, in agreement with the current paradigm that this region plays a crucial role for competent CFTR biosynthesis, stability and activity, as demonstrated by the major mutation (F508del) which is located at this interface[32,33].

Prediction of Corrector-sensitive CFTR Residues

The recently determined Cryo-EM structure of CFTR represents an unprecedented opportunity for identifying the binding sites of corrector ligands, as this structure represents the closed-channel conformation of the protein to which such ligands are believed to bind. Identifying the long sought binding site, or sites, of such ligands should help gain insight into the mechanism-of-action of these molecules and allow for patient stratification based on their mutation-sensitive 'theratype'.

To this end, we consider the Class I corrector ligands VX-809, VX-661 and C18, which share a common scaffold (pharmacophore), but also display some differences (**Figure S4A**). Previous studies have shown that VX-809 stabilizes the MSD1 domain of CFTR[27,28,34]. This stabilizing effect was abolished in MSD1 constructs that lacked the C-terminal segment comprising residues 374-380[27]. In addition, the deletion of only residues 371-375 from full-length CFTR produced a severe folding defect that could not be corrected by VX-809[27]. It therefore seems of interest to explore this region of CFTR for corrector binding sites. This choice was further confirmed by immunoblotting experiments carried out in this study. These experiments show that deletion of residues 373-380 abrogates corrector-mediated increase in MSD1 abundance, a common measure of the presence of folded protein, for all three Class I correctors (VX-809, VX-661 and C18) (**Figure 5A and 5B**). This observation suggests that all three ligands interacted with the C-terminal region of MSD1, in agreement with previous findings on VX-809 alone.

Docking calculations for all three ligands were hence carried out in the vicinity of residues F374 and L375 using the bounded volume described in the Methods and illustrated in **Figure 5C**. Assuming a similar mode of action for the three ligands, we favored docking poses in which the common scaffold of these molecules was closely aligned. The docking results, consisting of a top-scoring binding pocket, and the associated corrector conformations are shown in **Figures 5D and 5E**. All three modeled CFTR-ligand complexes featured comparable docking scores and involve very similar interactions with CFTR residues lining the pocket (**Table S1**). The majority of the residues making contact with the ligands are from MSD1 (i.e. lasso motif, ICL1 and C-term), a few are from the N-terminus of NBD1, and only a couple of residues at most from MSD2 (i.e. ICL4). The side-chain of one residue in particular, K166 (ICL1), appears to play an important role in corrector recognition. It was positioned closest (~4.0 Å) to the plane of the common scaffold in the top-

scoring CFTR-corrector complexes, allowing the ε-amino group of the lysine side-chain to possibly coordinate with the aromatic ring or carbonyl oxygen of this scaffold of each corrector so as to form cation-π interactions or H-bonds, respectively (**Figure S4B**). It is noteworthy that the ligands make no contact with the side-chain of F374 and minimal contact with the side-chain of L375, consistent with the finding that mutations of these residues to Alanine has little effect on the potency of VX-809[27].

Discussion

In this study, the known repertoire of exonic mutations and mutant allele frequencies in *CFTR*, compiled in three public databases, was mapped onto the three-dimensional structural models of the CFTR protein. Analysis of the spatial distribution of the mapped mutations revealed that disease-causing mutations tend to occur more frequently in certain regions of the CFTR structure, which likely play a strategic role in conferring the thermodynamic and dynamic properties required for protein function.

Mapping *CFTR* mutations from the ABCMdb, which includes validated disease-causing mutations, unvalidated disease-associated mutations, and those of experimental origin, showed that the N-terminal domains of CFTR (MSD1 and NBD1) have been more extensively probed by mutation studies. The greater focus on the N-terminal domains may be due to the occurrence in these domains of two relatively more common disease-causing mutations, R117H and F508del, and perhaps also to the increased influence that residues in the N-terminal half of this large polypeptide may have on proper folding of the full CFTR protein.

In contrast, mapping disease-causing mutations only from the CFTR1 and CFTR2 databases to the CFTR protein structures revealed that, for CFTR1, such mutations are found in all domains across the structure, while for CFTR2, where the mutations data is more limited but enriched for validated variants, MSD1 and NBD1 contain the greatest abundance of mutated residues. Further, analysis of the spatial distribution of amino acid positions of the most frequent mutant alleles derived from the CFTR2 database, revealed a statistically significant cluster in the tertiary protein

structure containing six of the most common disease-causing mutations; this spatial distribution was not completely apparent in the primary sequence. We speculate that selection pressures (e.g. environmental stresses) played a role in giving rise to this cluster, possibly supporting the *heterozygote advantage* hypothesis[35–38]. The latter suggests that carriers (i.e. heterozygotes) of a mutant *CFTR* allele are thought to have had a survival advantage compared to individuals with wild-type, and thus pressure-susceptible alleles; the leading hypotheses, although controversial among CF researchers, propose a survival advantage against cholera, typhoid fever and/or tuberculosis[35–38].

In addition, we found that disease-causing mutations were clustered at two distinct interdomain interfaces, NBD1:NBD2 and NBD1:ICL4, with the cluster at the latter interface being statistically significant ($p<0.05$). This observation confirms that these regions are fragile 'linchpins', essential for intramolecular communication and facilitating conformational changes required for CFTR function, and further, that mutation of individual residues within these interfaces is sufficient to destabilize and/or hamper proper tertiary assembly of functionally competent states of the protein[29,30]. This is in agreement with previous studies ,which showed that residues buried within domain-domain interfaces are commonly associated with disease-causing mutations of many proteins in the human proteome[39].

Two models for the three-dimensional structure of the CFTR protein were used for the mapping analysis, the full-length CFTR homology model representing the open-channel, active state of the protein and the cryo-EM partial structure of CFTR in the closed-channel, inactive state. Neither model is of high accuracy by current standards of protein crystal structures. But this may only marginally affect the observations made from the mutation mapping analysis, as the latter should not critically depend on model accuracy. On the other hand, the cryo-EM structure has the advantage of being the first experimentally-derived CFTR structure with amino acid side-chain resolution. The asset of the homology model is that it is truly full-length. It includes built atomic coordinates for the structurally flexible R-domain (residues 631-849), which contains approximately 10% of all *CFTR* mutations. In addition, the model for the MSD-NBD dimer portion is based on the canonical scaffold, which is well-represented in the structurally-characterized

Protein Data Bank (PDB). This model represents the ATP-bound state of CFTR, which correlates with a large set of structural homologues in the PDB (i.e. ATP-bound, non-mammalian ABC transporters) and is in agreement with several biochemical studies, further supporting its validity[18].

In the last part of our study, we investigated the putative binding site of three Class I correctors molecules, VX-809, VX-661 and C18. We confirmed experimentally that all three ligands stabilized MSD1 of CFTR, requiring the C-terminal segment of this domain to elicit their corrector action, as reported for VX-809 in other studies[27,28,34]. The three ligands were then computationally docked onto a 30-$Å^3$ region centered at the C-terminal region of MSD1 (i.e. residues F374-L375) in the cryo-EM structure of CFTR, and poses which demonstrated alignment of the pharmacophore (i.e. common scaffold) of these three small molecules suggested the most plausible binding pocket. This enabled the identification of a sizable binding pocket with least 12 residues of the CFTR protein (a sufficient number for ligands of these molecular weights: i.e. ~450-520 Da), many of which were in MSD1, engaging in interactions with the bound ligands (**Table S1**). Our findings that the identified corrector binding pocket comprises residues from different domains (i.e. lasso motif, ICL1 and C-term of MSD1, N-terminus of NBD1, and ICL4 of MSD2; **Table S1**) further elaborate on previous *in silico* docking studies, which propose that a composite, multi-domain pocket comprised of residues within the NBD1:ICL4 interface (near the major mutation: F508del) is likely to accommodate VX-809[24,25].

More specifically, we found that the three non-pharmacophore hydroxyl groups of VX-661, which are not featured in the investigational and first-generation corrector molecules: C18 and VX-809, respectively (**Figure S4A**), formed H-bonds with residues E54, W57 and K163 in the predicted pocket (**Figure S4B**), possibly explaining the *in vitro* effectiveness of VX-661 over these other correctors. Furthermore, binding poses for all three correctors position the pharmacophore in a groove comprising residues K166, Y380, T382, R1066 and Q1071, and interacting more closely with K166. In the complex with VX-661, the positively charged epsilon amino group of K166 was appropriately positioned towards the aromatic benzodioxole group to form a potential cation-π interaction[40-42]. C18 featured a slight rotation of the K166 side-chain that places its charged group

in an appropriate geometry for an H-bond with the corrector's carbonyl oxygen of the pharmacophore (data not shown). These different options for the K166 side-chain to form stabilizing interactions with different ligands or possibly with the same ligand, support the predicted involvement of this residue in corrector binding. Furthermore, our future studies will investigate the role of K166 in coordination and subsequent binding of these Class I correctors using a site-directed mutagenesis approach.

Finally, it is noteworthy that disease-causing mutations have been reported for several residues in the predicted binding pocket (e.g. R1066C/S/H/L; see **Table S1**). It is therefore possible that the Class I correctors analyzed here may repair structural defects caused by these variants (locally in this pocket and globally throughout CFTR protein), and potentially become therapeutically relevant for CF patients bearing these mutations. Taken together, these data suggest that the current CF co-therapy (i.e. Orkambi®, but mainly lumacaftor) may benefit this patient subpopulation. Furthermore, it is clear that additional investigations using mutational and biophysical analyses are needed to confirm the predicted binding pocket and the proposed contributions of various CFTR residues therein to the binding of the three Class I correctors.

Acknowledgments

Author contributions: The overall design of the study was by S.V.M.; S.V.M., V.M.S., A.S.S., S.S.M., G.W., M.L., O.L., L.D.M., C.E.B. and A.W. performed experiments, analyzed and/or interpreted results. The manuscript was primarily written by S.V.M. with input from all authors. **Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. **Competing interests**: All authors affiliated with Cyclica are employees or advisors to the company. All other authors declare no conflicts of interest.

References

1. Riordan, J. R. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245,** 1066–73 (1989).
2. Riordan, J. R. CFTR function and prospects for therapy. *Annu. Rev. Biochem.* **77,** 701–26 (2008).
3. Rowe, S. M., Miller, S. & Sorscher, E. J. Cystic fibrosis. *N. Engl. J. Med.* **352,** 1992–2001 (2005).
4. Boyle, M. P. & De Boeck, K. A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect. *Lancet. Respir. Med.* **1,** 158–63 (2013).
5. Eckford, P. D. W. *et al.* VX-809 and related corrector compounds exhibit secondary activity stabilizing active F508del-CFTR after its partial rescue to the cell surface. *Chem. Biol.* **21,** 666–678 (2014).
6. Molinski, S. V. *et al.* Genetic, cell biological, and clinical interrogation of the CFTR mutation c.3700 A>G (p.Ile1234Val) informs strategies for future medical intervention. *Genet. Med.* **16,** 625–632 (2014).
7. Sosnay, P. R. *et al.* Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat. Genet.* **45,** 1160–7 (2013).
8. Veit, G. *et al.* From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. *Mol. Biol. Cell* **27,** 424–33 (2016).
9. Pasyk, S. *et al.* The major cystic fibrosis causing mutation exhibits defective propensity for phosphorylation. *Proteomics* **15,** 447–461 (2015).
10. Malik, F. A. *et al.* Sphingosine-1-Phosphate Is a Novel Regulator of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Activity. *PLoS One* **10,** e0130313 (2015).
11. Molinski, S. *et al.* Functional rescue of F508del-CFTR using small molecule correctors. *Front. Pharmacol.* **3 SEP,** 1–18 (2012).
12. Pasyk, S., Molinski, S., Yu, W., Eckford, P. D. W. & Bear, C. E. Identification and validation of hits from high throughput screens for CFTR modulators. *Curr. Pharm. Des.* **18,** 628–41 (2012).

13. Molinski, S. V, Ahmadi, S., Hung, M. & Bear, C. E. Facilitating Structure-Function Studies of CFTR Modulator Sites with Efficiencies in Mutagenesis and Functional Screening. *J. Biomol. Screen.* (2015). doi:10.1177/1087057115605834

14. Van Goor, F. *et al.* Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 18825–30 (2009).

15. Van Goor, F. *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,** 18843–8 (2011).

16. Maggio, E. T., Shenderovich, M., Kagan, R., Goddette, D. & Ramnarayan, K. Structural pharmacogenomics, drug resistance and the design of anti-infective super-drugs. *Drug Discov. Today* **7,** 1214–20 (2002).

17. Fülöp, K., Barna, L., Symmons, O., Závodszky, P. & Váradi, A. Clustering of disease-causing mutations on the domain-domain interfaces of ABCC6. *Biochem. Biophys. Res. Commun.* **379,** 706–9 (2009).

18. Mornon, J.-P., Lehn, P. & Callebaut, I. Molecular models of the open and closed states of the whole human CFTR protein. *Cell. Mol. Life Sci.* **66,** 3469–86 (2009).

19. Liu, F., Zhang, Z., Csanády, L., Gadsby, D.C. & Chen, J. Molecular Structure of the Human CFTR Ion Channel. *Cell* **169(1),** 85–95 (2017).

20. Mornon, J.-P., Hoffmann, B., Jonic, S., Lehn, P., & Callebaut, I. Full-open and closed CFTR channels, with lateral tunnels from the cytoplasm and an alternative position of the F508 region, as revealed by molecular dynamics. *Cell Mol Life Sci.* **72(7),** 1377–403 (2015).

21. Rosenberg, M.F. *et al*. The cystic fibrosis transmembrane conductance regulator (CFTR): three-dimensional structure and localization of a channel gate. *J Biol Chem*. **286(49),** 42647–54 (2011).

22. Cutting, G. R. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat. Rev. Genet.* **16,** 45–56 (2015).

23. Gyimesi, G. *et al.* ABCMdb: a database for the comparative analysis of protein mutations in ABC transporters, and a potential framework for a general application. *Hum. Mutat.* **33,** 1547–56 (2012).

24. He, L., *et al*. Correctors of ΔF508 CFTR restore global conformational maturation without thermally stabilizing the mutant protein. *FASEB J* **27(2)**, 536–45 (2013).
25. Okiyoneda, T., *et al*. Mechanism-based corrector combination restores ΔF508-CFTR folding and function. *Nat Chem Biol* **9(7)**, 444–54 (2013).
26. Molinski, S.V. *et al.* Computational proteome-wide screening predicts neurotoxic drug-protein interactome for the investigational analgesic BIA 10-2474. *Biochem Biophys Res Commun* **483(1),** 502–508 (2017).
27. Ren, H. Y. *et al.* VX-809 corrects folding defects in cystic fibrosis transmembrane conductance regulator protein through action on membrane-spanning domain 1. *Mol. Biol. Cell* **24,** 3016–24 (2013).
28. Laselva, O., Molinski, S., Casavola, V. & Bear, C. E. The investigational Cystic Fibrosis drug Trimethylangelicin directly modulates CFTR by stabilizing the first membrane-spanning domain. *Biochem. Pharmacol.* **119,** 85–92 (2016).
29. Cui, L. *et al.* Domain interdependence in the biosynthetic assembly of CFTR. *J. Mol. Biol.* **365,** 981–94 (2007).
30. Vergani, P., Lockless, S. W., Nairn, A. C. & Gadsby, D. C. CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. *Nature* **433,** 876–80 (2005).
31. Mense, M. *et al.* In vivo phosphorylation of CFTR promotes formation of a nucleotide-binding domain heterodimer. *EMBO J.* **25,** 4728–39 (2006).
32. Rabeh, W. M. *et al.* Correction of both NBD1 energetics and domain interface is required to restore ΔF508 CFTR folding and function. *Cell* **148,** 150–63 (2012).
33. Mendoza, J. L. *et al.* Requirements for efficient correction of ΔF508 CFTR revealed by analyses of evolved sequences. *Cell* **148,** 164–74 (2012).
34. Loo, T. W., Bartlett, M. C. & Clarke, D. M. Corrector VX-809 stabilizes the first transmembrane domain of CFTR. *Biochem. Pharmacol.* **86,** 612–9 (2013).
35. Gabriel, S. E., Brigman, K. N., Koller, B. H., Boucher, R. C. & Stutts, M. J. Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science* **266,** 107–9 (1994).

36. Poolman, E. M. & Galvani, A. P. Evaluating candidate agents of selective pressure for cystic fibrosis. *J. R. Soc. Interface* **4,** 91–8 (2007).

37. Lubinsky, M. Hypothesis: Cystic fibrosis carrier geography reflects interactions of tuberculosis and hypertension with vitamin D deficiency, altitude and temperature. Vitamin D deficiency effects and CF carrier advantage. *J. Cyst. Fibros.* **11,** 68–70 (2012).

38. Azimi, A. 'Cystic fibrotics could survive cholera, choleraics could survive cystic fibrosis'; hypothesis that explores new horizons in treatment of cystic fibrosis. *Med. Hypotheses* **85,** 715–7 (2015).

39. Gao, M., Zhou, H. Skolnick J. Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis. *Structure* **23(7),** 1362–9 (2015).

40. Burley, S. K. & Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **203(2),** 139–43 (1986).

41. Schärer, K, *et al.* Quantification of cation-pi interactions in protein-ligand complexes: crystal-structure analysis of Factor Xa bound to a quaternary ammonium ion ligand. *Angew Chem Int Ed Engl*. **44(28)**, 4400–4 (2005).

42. Gallivan, J.P. & Dougherty, D.A. Cation-pi interactions in structural biology. *Proc Natl Acad Sci USA* **96(17)**, 9459–64 (1999).

Figure Legends

Figure 1: Computational mapping of exonic *CFTR* mutation frequencies from the CFTR1 mutation database to CFTR protein structures.

(**A**) Side and top-down views of the full-length CFTR homology model (surface representation) in the open-channel (active) state, and (**B**) side and bottom-up views of the cryo-EM-derived partial structure of CFTR (surface representation) in the closed-channel (inactive) state coloured according to exonic *CFTR* mutation frequencies (counts) reported in the CFTR1 mutation database. (**C**) Histogram of exonic *CFTR* mutation frequencies (counts) as reported in the CFTR1 mutation database for each residue in the protein sequence. Highly mutated residue positions are indicated.

Figure 2: Computational mapping of exonic *CFTR* mutation frequencies from the CFTR2 mutation database to CFTR protein structures.

(**A**) Side and top-down views of the full-length CFTR homology model (surface representation) in the open-channel (active) state, and (**B**) side and bottom-up views of the cryo-EM-derived partial structure of CFTR (surface representation) in the closed-channel (inactive) state coloured according to exonic *CFTR* mutation frequencies (counts) reported in the CFTR2 mutation database. (**C**) Histogram of exonic *CFTR* mutation frequencies (counts) as reported in the CFTR2 mutation database for each residue in the protein sequence. Highly mutated residue positions are indicated.

Figure 3: Computational mapping of exonic *CFTR* mutant allele frequencies from the CFTR2 mutation database to a structural model of CFTR.

(**A**) Side and top-down views of the full-length CFTR homology model (surface representation) in the open-channel (active) state coloured according to exonic *CFTR* mutant allele frequencies ($\log_{10}$counts) reported in the CFTR2 mutation database. The boxed (blue) region of CFTR indicates a mutation cluster. (**B**) Close-up view of mutation cluster. Six residues with top-ranking mutant allele frequencies ($\log_{10}$counts) are labelled, and include: F508, G542, G551, R553, W1282 and N1303. (**C**) Histogram of exonic *CFTR* mutant allele frequencies ($\log_{10}$counts) as reported in the

CFTR2 mutation database for each residue in the protein sequence. Highly mutated residue positions are indicated. (**D**) Distribution of AID values for randomly selected six-residue sets relative to the AID of the six-residue cluster (F508, G542, G551, R553, W1282, N1303).

Figure 4: Identification of exonic *CFTR* mutations at NBD1:NBD2 and NBD1:ICL4 interdomain interfaces in the structural model of CFTR.

(**A**) NBD1:NBD2 and (**B**) NBD1:ICL4 interdomain interfaces in the full-length CFTR homology model (surface representation) the open-channel (active) state. Interface residues (i.e. those < 4 Å from the opposing domain) are coloured white, and exonic *CFTR* mutations (counts) found within interface residues are coloured according to values reported in the CFTR1 mutation database. Cyan, NBD1 (residues 381-630); purple, NBD2 (1171-1480); pink, ICL4 (1039-1093).

Figure 5: Identification of the putative Class I corrector binding pocket using molecular docking.

(**A**) Immunoblots and (**B**) quantitative analysis of HEK-293 cells transiently expressing CFTR MSD1 fragments (i.e. full-length: K381X, and the C-terminus truncation: D373X) following chronic treatment (24 h) with correctors VX-809 (3 μM), VX-661 (1 μM) and C18 (6 μM) (n=3). Calnexin was used as a loading control. **, $p<0.01$; *, $p<0.05$; n.s., not significant. (**C**) The cryo-EM-derived partial/core-structure of CFTR in the closed-channel (inactive) state served as a template for molecular docking studies. The 30x30x30-Å cube representing the docking boundary, localized around residues F374-L375 (red spheres), is shown (dashed, dark blue lines); CFTR is shown in surface representation and coloured by domain as described in the legend. (**D**) Close-up view of binding modes for VX-809 (green), VX-661 (magenta) and C18 (orange), demonstrating alignment of the common scaffold, following molecular docking into the top-ranking and putative corrector binding site; CFTR is shown in surface representation and domains are coloured as in panel **C**. (**E**) Mean docking scores for top-scoring binding modes of Class I correctors following rigid (white bars) or flexible (gray bars) docking to the cryo-EM-derived structure of CFTR in the closed-channel state (n=9).

Table Captions

Table 1: Clustering of exonic *CFTR* mutations at the NBD1:ICL4 interdomain interface of CFTR. The CFTR1 mutation database was used to map exonic *CFTR* mutation frequencies (counts) onto interdomain interface residues of the full-length CFTR homology model in the open-channel (active) state using a proximity-measure (i.e. residues < 4 Å from an opposing surface or 'face' were deemed interdomain positions). Relative mutation frequencies at interdomain interfaces (i.e. NBD1:NBD2 and NBD1:ICL4) were compared to the relative mutation frequency of full-length CFTR to assess significance. *, $p<0.05$.

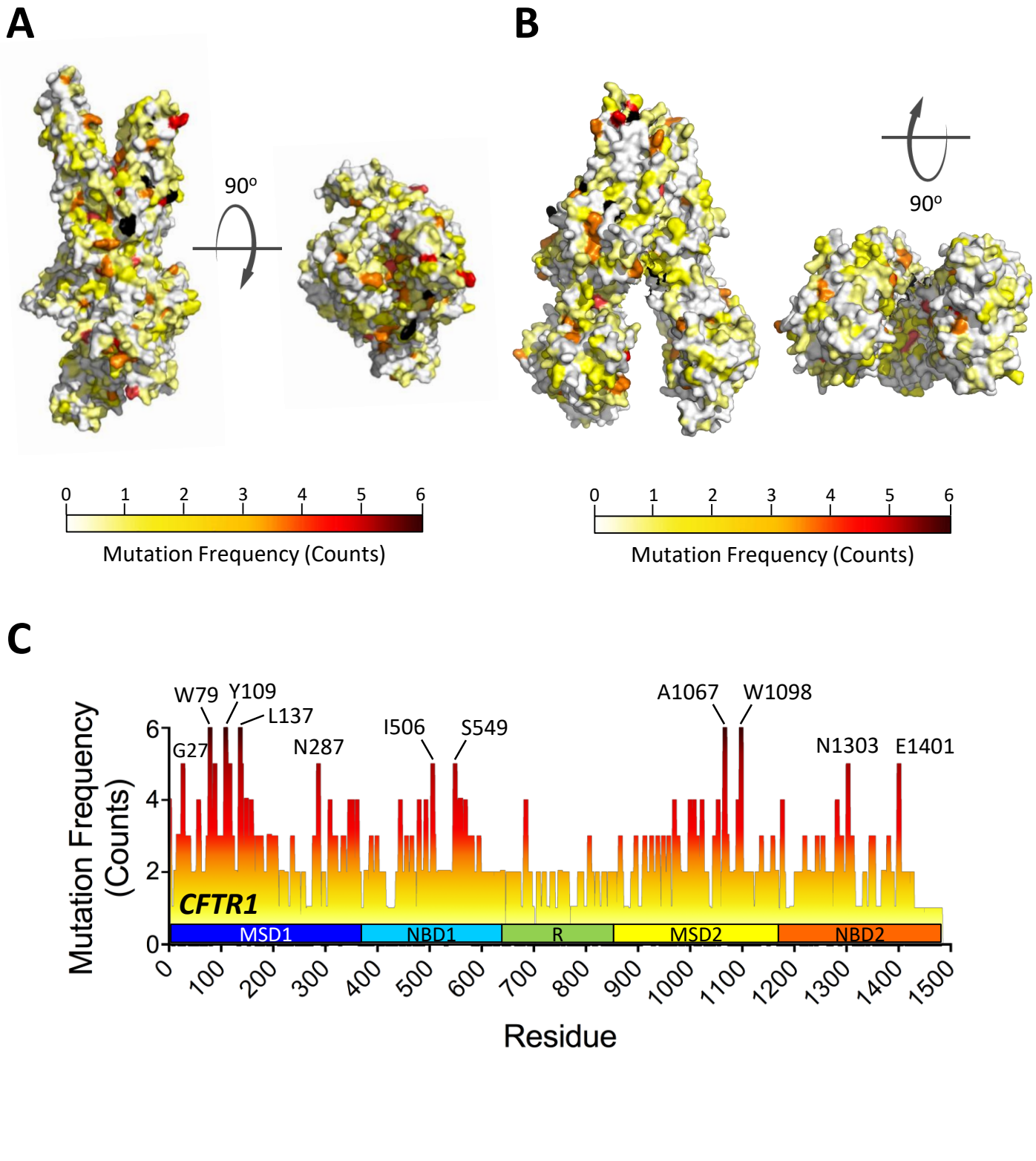
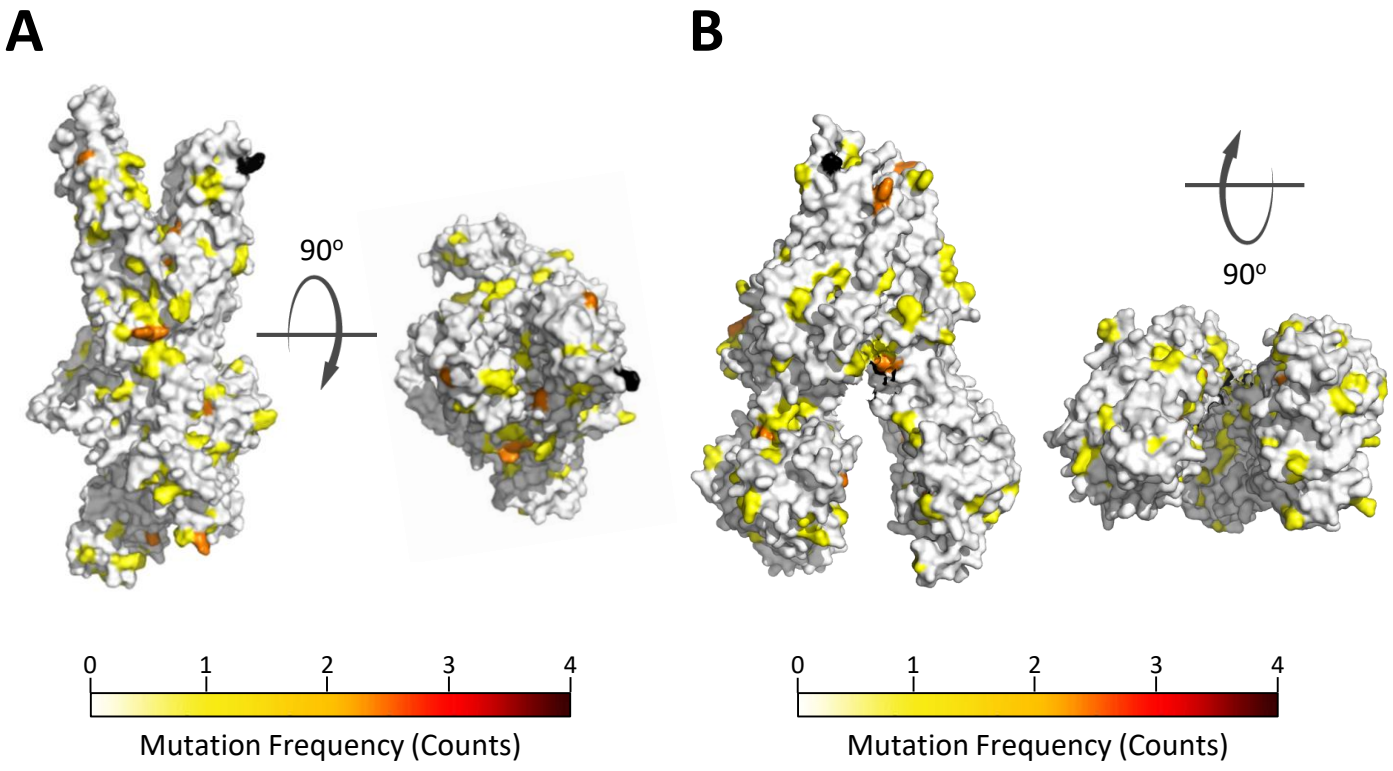

A
90°
0 1 2 3 4 5 6
Mutation Frequency (Counts)
B
90°
0 1 2 3 4 5 6
Mutation Frequency (Counts)
C
W79
Y109
L137
G27
N287
I506
S549
A1067
W1098
N1303
E1401
Mutation Frequency (Counts)
CFTR1
MSD1
NBD1
R
MSD2
NBD2
Residue

Fig. 1


A
B
90°
90°
0
1
2
3
4
Mutation Frequency (Counts)
0
1
2
3
4
Mutation Frequency (Counts)

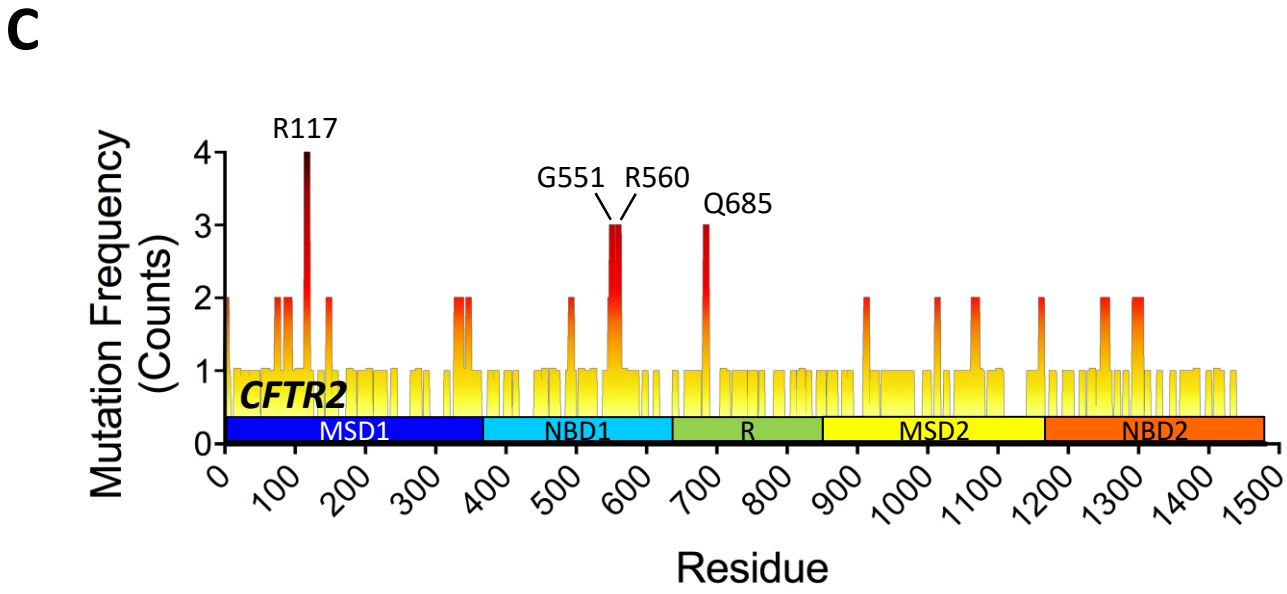

C
R117
G551
R560
Q685
CFTR2
MSD1
NBD1
R
MSD2
NBD2
Mutation Frequency (Counts)
Residue
0
1
2
3
4
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500

Fig. 2

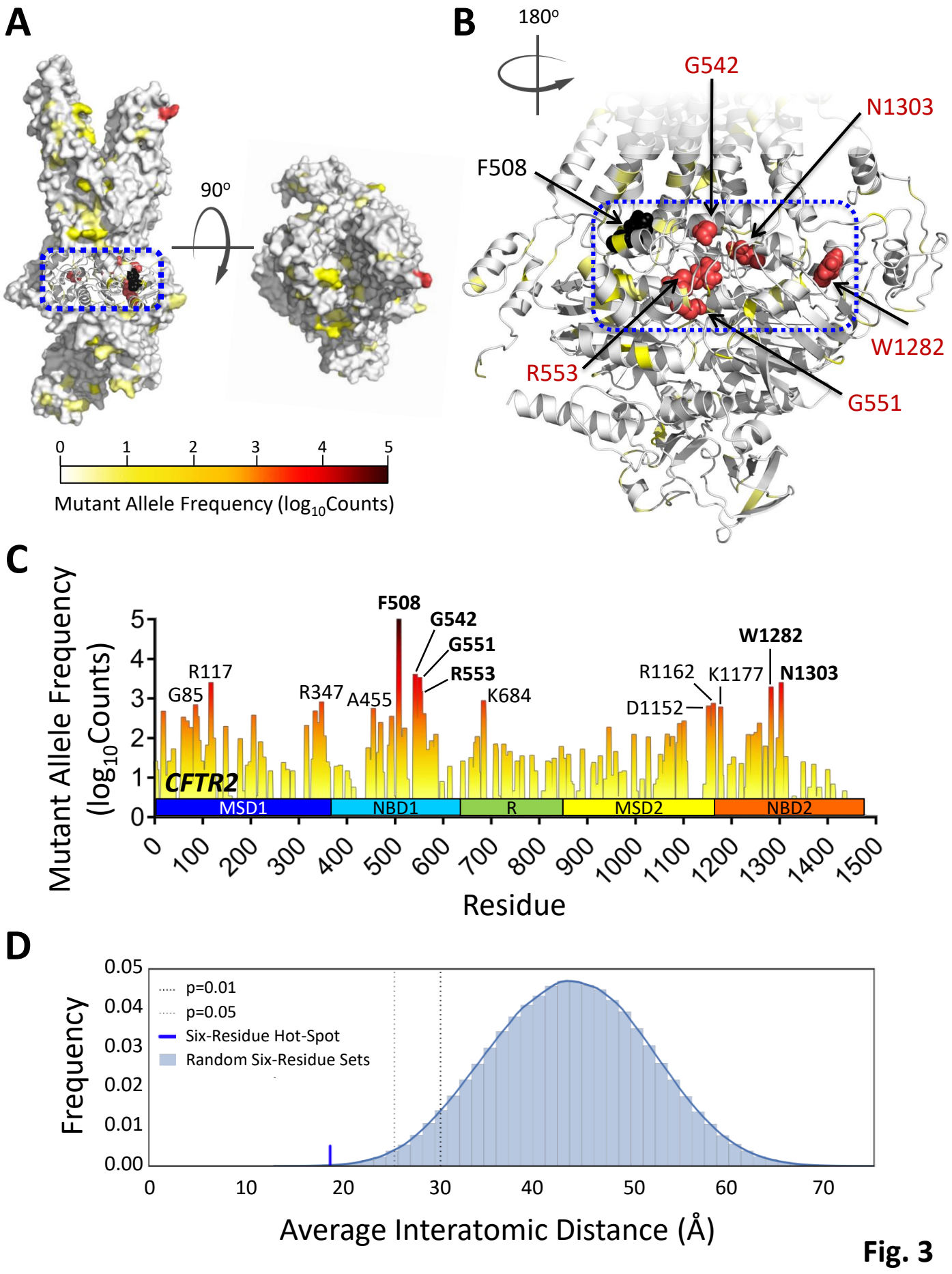

A
90°
0
1
2
3
4
5
Mutant Allele Frequency (log10Counts)
B
180°
G542
N1303
F508
W1282
R553
G551
C
Mutant Allele Frequency (log10Counts)
F508
G542
G551
R553
W1282
R117
G85
R347
A455
K684
R1162
K1177
N1303
D1152
CFTR2
MSD1
NBD1
R
MSD2
NBD2
Residue
D
Frequency
p=0.01
p=0.05
Six-Residue Hot-Spot
Random Six-Residue Sets
Average Interatomic Distance (Å)

Fig. 3

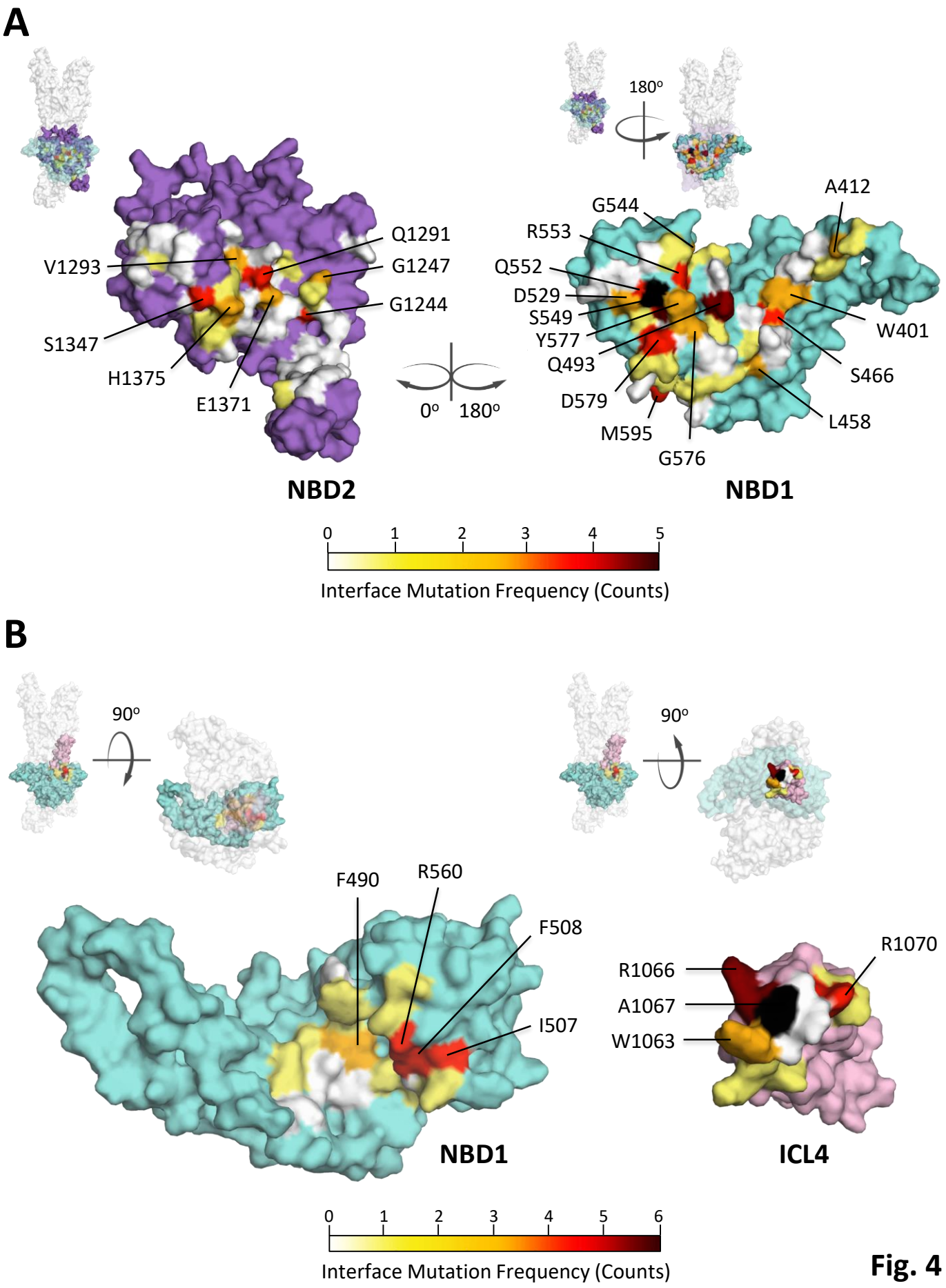

A
180°
Q1291
V1293
G1247
G1244
S1347
H1375
E1371
0°
180°
NBD2
G544
R553
A412
Q552
D529
S549
Y577
W401
Q493
S466
D579
L458
M595
G576
NBD1
0
1
2
3
4
5
Interface Mutation Frequency (Counts)
B
90°
90°
F490
R560
F508
I507
NBD1
R1070
R1066
A1067
W1063
ICL4
0
1
2
3
4
5
6
Interface Mutation Frequency (Counts)

Fig. 4

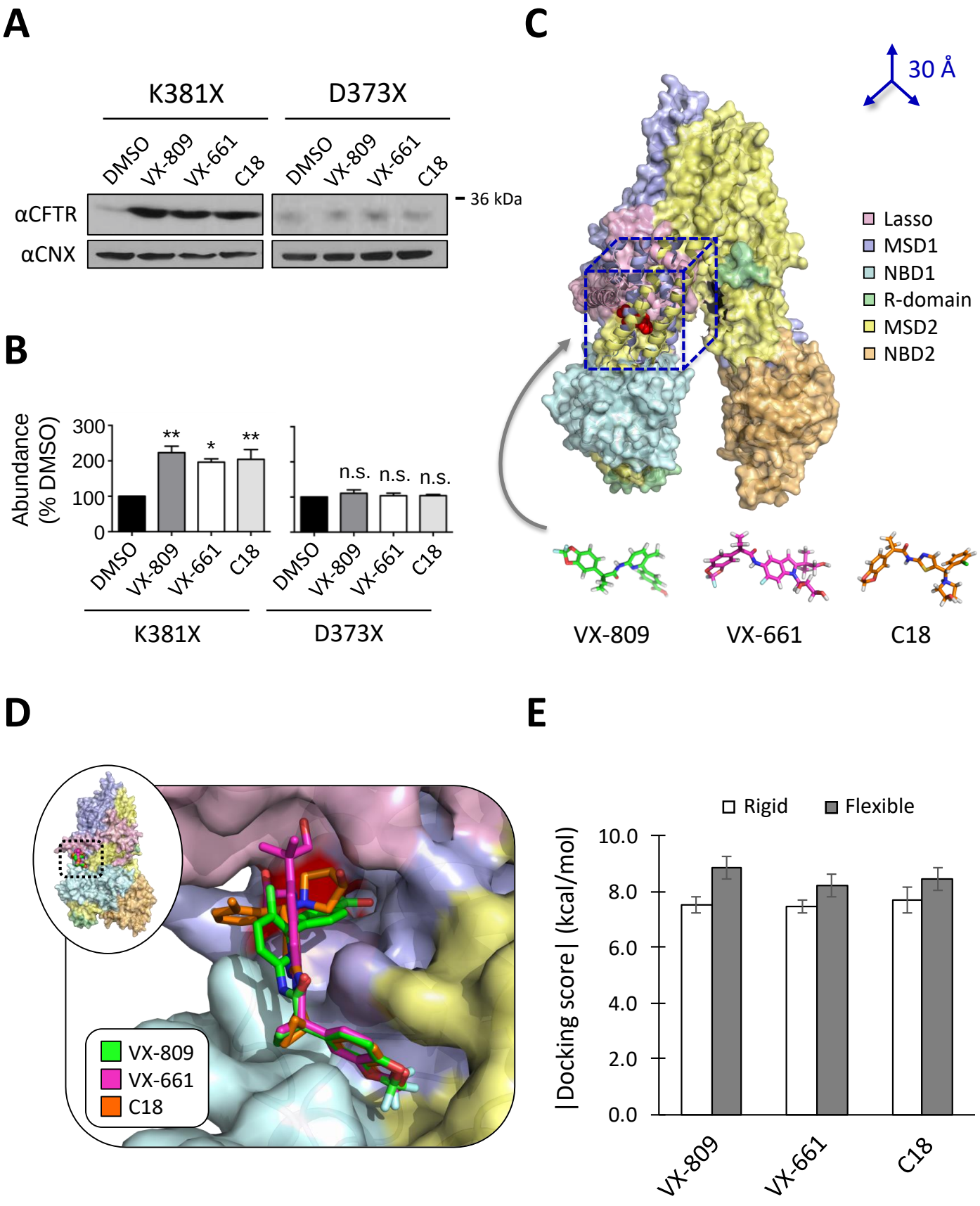

A
K381X
D373X
DMSO
VX-809
VX-661
C18
DMSO
VX-809
VX-661
C18
36 kDa
αCFTR
αCNX
B
Abundance (% DMSO)
300
200
100
0
**
*
**
n.s.
n.s.
n.s.
DMSO
VX-809
VX-661
C18
DMSO
VX-809
VX-661
C18
K381X
D373X
C
30 Å
Lasso
MSD1
NBD1
R-domain
MSD2
NBD2
VX-809
VX-661
C18
D
VX-809
VX-661
C18
E
Rigid
Flexible
|Docking score| (kcal/mol)
10.0
8.0
6.0
4.0
2.0
0.0
VX-809
VX-661
C18

Fig. 5

Supplementary Figure Legends

Figure S1: Topological model and domain-coloured tertiary structures of CFTR.
(**A**) A topology model of CFTR showing MSD1 (blue, residues 1-380), NBD1 (cyan, residues 381-630), R-domain (green, residues 631-849), MSD2 (yellow, residues 850-1170) and NBD2 (orange, residues 1171-1480). The 'Y-shaped' symbols represent N-glycosylation sites. TM, transmembrane α-helix. The relative position of the plasma membrane is shown in gray. (**B**) Side and top-down views of the full-length CFTR homology model (surface representation) in the open-channel (active) state, shown in surface representation. (**C**) Side and bottom-up views of the cryo-EM-derived structure of CFTR (lacking a majority of the R-domain, i.e. residues 646-843) in the closed-channel (inactive) state, shown in surface representation. Domains are coloured as in panel **A**, and the relative position of the plasma membrane is shown with a dotted line.

Figure S2: Computational mapping of exonic *CFTR* mutation frequencies from the ABC mutation database to CFTR protein structures.
(**A**) Side and top-down views of the full-length CFTR homology model (surface representation) in the open-channel (active) state, and (**B**) side and bottom-up views of the cryo-EM-derived partial structure of CFTR (surface representation) in the closed-channel (inactive) state coloured according to exonic *CFTR* mutation frequencies (counts) reported in the ABC mutation database. (**C**) Histogram of exonic *CFTR* mutation frequencies (counts) as reported in the ABC mutation database for each residue in the protein sequence. Highly mutated residue positions are indicated.

Figure S3: List of CFTR residues located at NBD1:NBD2 and NBD1:ICL4 interdomain interfaces.
(**A**) NBD1 and NBD2 residues found within 4 Å of the opposing face of the NBD1:NBD2 interdomain interface, as generated from a PyMOL-based proximity-measure script. Syntax includes 'X', 'Y' and 'Z' coordinates as: position, native residue, and number of mutations (in bold italics), respectively. (**B**) NBD1 and ICL4 residues found within 4 Å of the opposing face of the NBD1:ICL4 interdomain interface, as generated from a PyMOL-based proximity-measure script.

Syntax includes 'X', 'Y' and 'Z' coordinates as: position, native residue, and number of mutations (in bold italics), respectively.

Figure S4: Chemical structures of VX-809, VX-661, C18, and the predicted binding mode of VX-661 in the putative corrector binding pocket.

(**A**) Chemical structures of each Class I corrector. The common scaffold or 'pharmacophore' of each small molecule is highlighted with a dashed box (orange). (**B**) As an example, the predicted binding mode of VX-661, captured in the putative corrector binding pocket of the cryo-EM-derived partial structure of CFTR in the closed-channel (inactive) state, is shown in the inset (pink, lasso motif; yellow, ICL4 of MSD2; cyan, NBD1; light blue, ICL1 of MSD1; dark gray, K166; magenta, VX-661; dashed yellow line, predicted cation-π interaction). A two-dimensional representation of the predicted binding mode of the example Class I corrector, VX-661, in the putative corrector binding pocket (corresponding to the inset) is also shown, as generated using the PoseView tool from the ZBH Proteins*Plus* webserver (proteinsplus.zbh.uni-hamburg.de).

Supplementary Table Captions

Table S1: CFTR residues predicted to interact with Class I correctors VX-809, VX-661 and C18.

List of residues (subdomain, domain) predicted to be within 5 Å of each respective small molecule corrector following molecular docking studies. Residues in italics are common among the three compounds, and the residue in bold italics (i.e. K166) is found ~4.0 Å of the common scaffold or 'pharmacophore' of each corrector. Residues bearing disease-associated missense mutations, as reported in the CFTR1 database, are labelled with an asterisk.

VX-809	VX-661	C18
E51 (lasso, MSD1)	*E54* (lasso, MSD1)	*E54* (lasso, MSD1)
E54 (lasso, MSD1)	W57 (lasso, MSD1)*	W57 (lasso, MSD1)*
L159 (ICL1, MSD1)*	D58 (lasso, MSD1)*	D58 (lasso, MSD1)*
K162 (ICL1, MSD1)*	*L159* (ICL1, MSD1)*	L61 (lasso, MSD1)*
K163 (ICL1, MSD1)*	I160 (ICL1, MSD1)	*L159* (ICL1, MSD1)*
K166 (ICL1, MSD1)*	Y161 (ICL1, MSD1)*	*K162* (ICL1, MSD1)*
E379 (C-term, MSD1)*	*K162* (ICL1, MSD1)*	*K163* (ICL1, MSD1)*
Y380 (C-term, MSD1)	*K163* (ICL1, MSD1)*	***K166*** (ICL1, MSD1)*
K381 (N-term, NBD1)	L165 (ICL1, MSD1)*	L375 (C-term, MSD1)*
T382 (N-term, NBD1)	***K166*** (ICL1, MSD1)*	*E379* (C-term, MSD1)*
R1066 (ICL4, MSD2)*	L375 (C-term, MSD1)*	*Y380* (C-term, MSD1)
Q1071 (ICL4, MSD2)*	*E379* (C-term, MSD1)*	*K381* (N-term, NBD1)
-	*Y380* (C-term, MSD1)	*T382* (N-term, NBD1)
-	*K381* (N-term, NBD1)	*R1066* (ICL4, MSD2)*
-	*T382* (N-term, NBD1)	*Q1071* (ICL4, MSD2)*
-	L1065 (ICL4, MSD2)*	-
-	*R1066* (ICL4, MSD2)*	-
-	*Q1071* (ICL4, MSD2)*	-

1

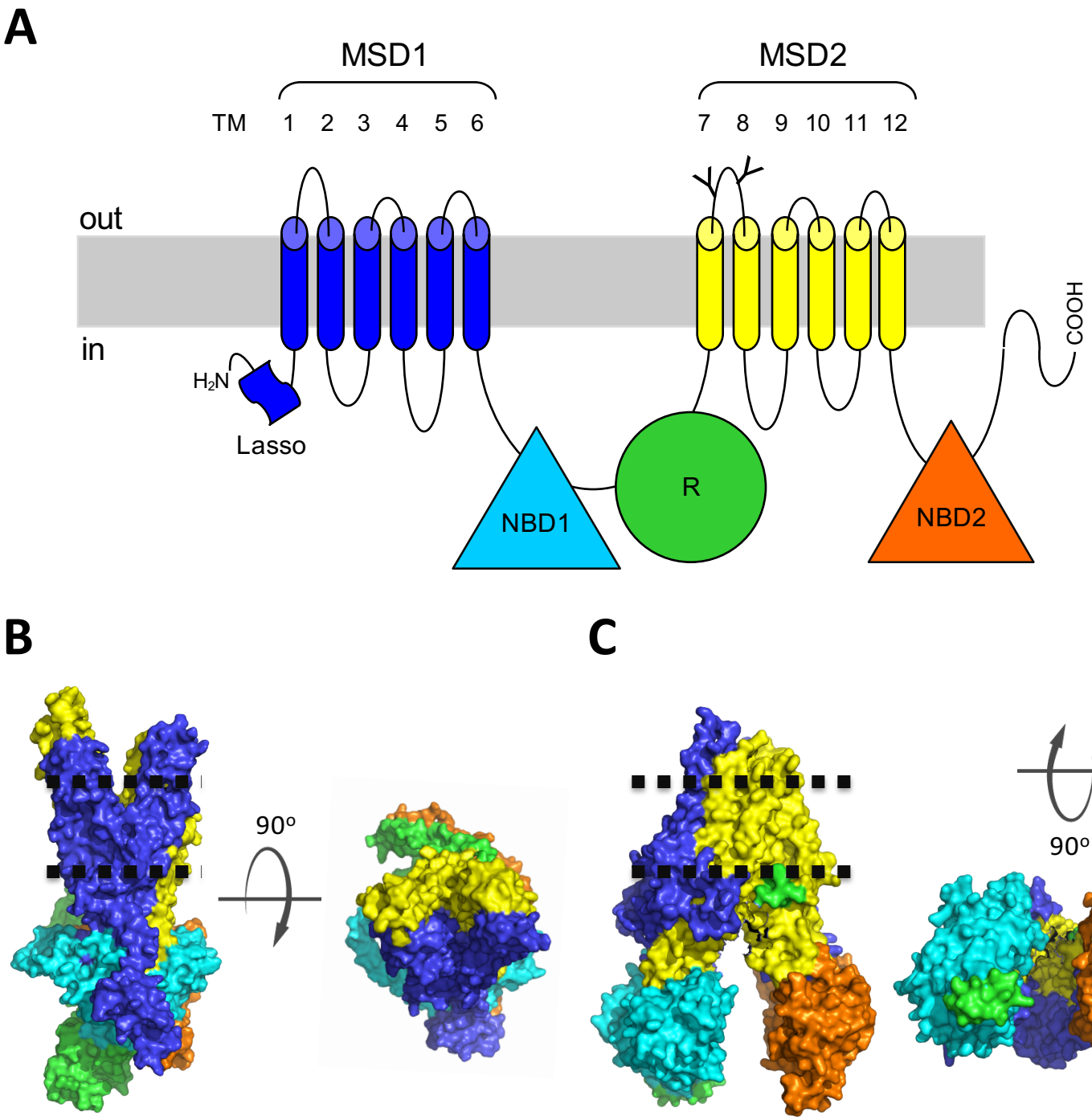
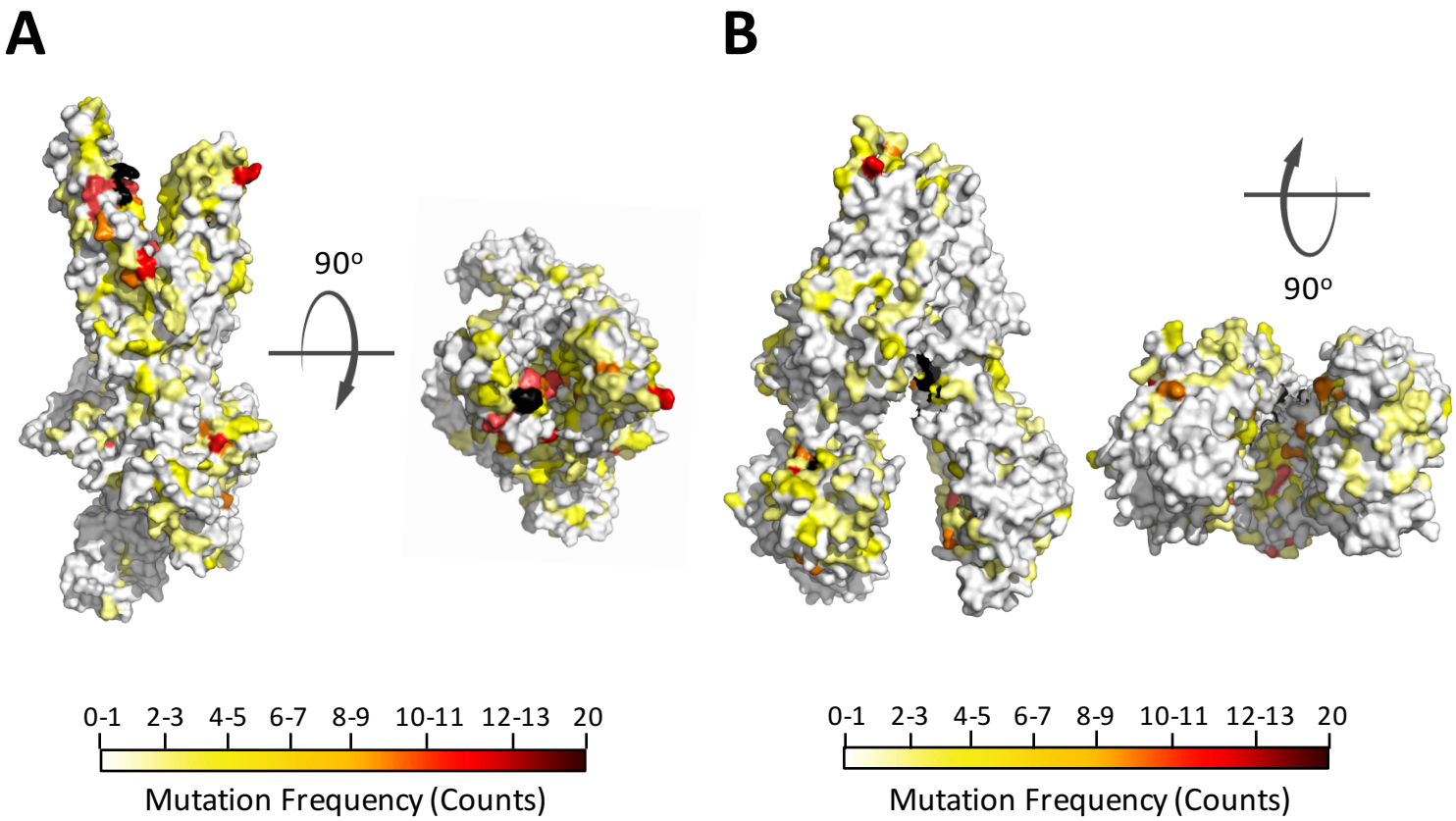

A
MSD1
MSD2
TM 1 2 3 4 5 6
7 8 9 10 11 12
out
in
H2N
Lasso
COOH
NBD1
R
NBD2
B
90°
C
90°

Fig. S1


A
90°
0-1 2-3 4-5 6-7 8-9 10-11 12-13 20
Mutation Frequency (Counts)
B
90°
0-1 2-3 4-5 6-7 8-9 10-11 12-13 20
Mutation Frequency (Counts)

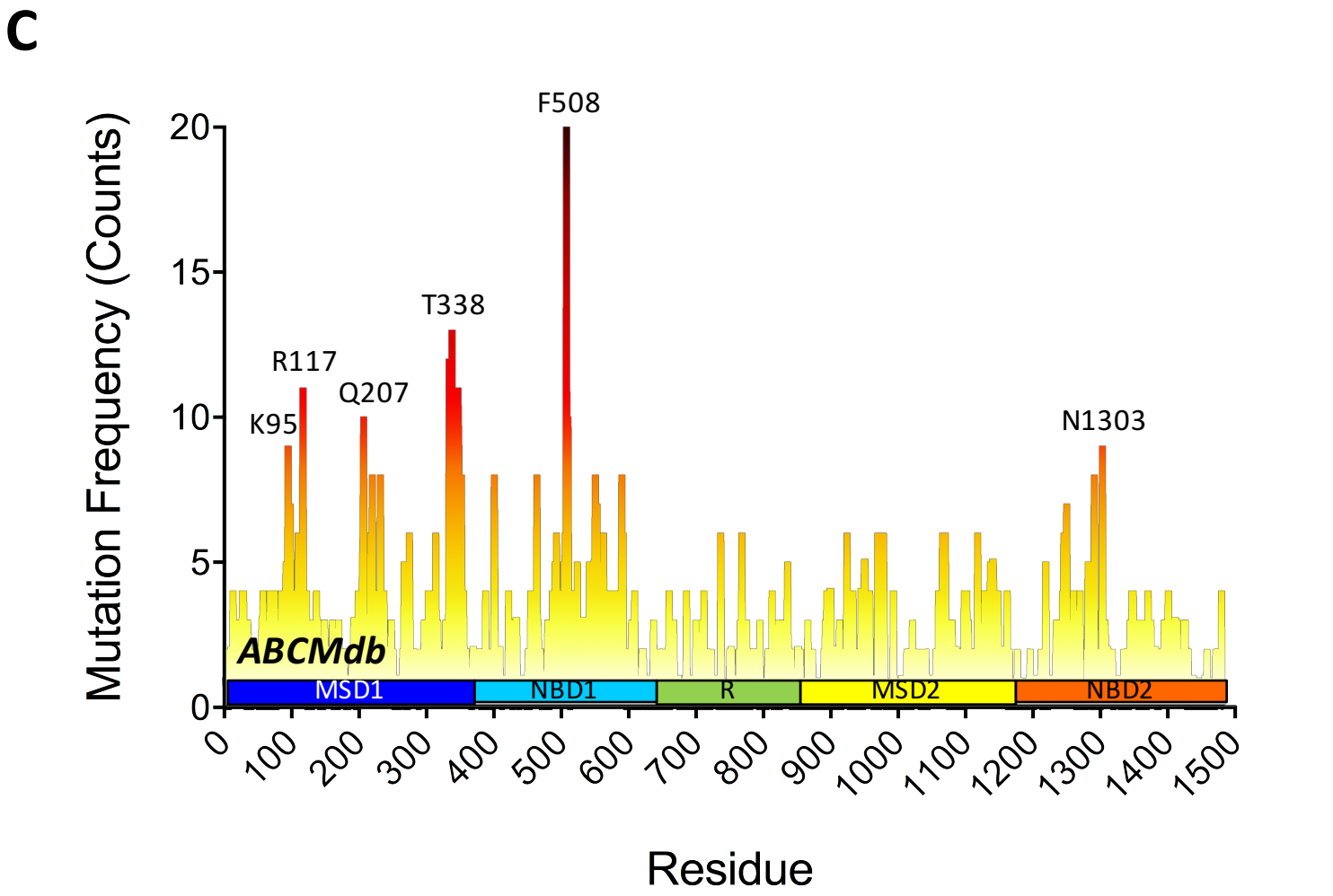

C
Mutation Frequency (Counts)
20
15
10
5
0
F508
T338
R117
Q207
K95
N1303
ABCMdb
MSD1
NBD1
R
MSD2
NBD2
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
Residue

Fig. S2

A

NBD1:NBD2 Interdomain Interface Residues (< 4 Å)

NBD1:
[('401', 'TRP', '***2***'), ('402', 'GLU', '***0***'), ('405', 'PHE', '***0***'), ('408', 'LEU', '***0***'), ('409', 'PHE', '***1***'), ('412', 'ALA', '***2***'), ('414', 'GLN', '***1***'), ('458', 'GLY', '***2***'), ('459', 'SER', '***0***'), ('460', 'THR', '***0***'), ('461', 'GLY', '***0***'), ('462', 'ALA', '***0***'), ('463', 'GLY', '***2***'), ('464', 'LYS', '***1***'), ('465', 'THR', '***0***'), ('466', 'SER', '***3***'), ('493', 'GLN', '***4***'), ('494', 'PHE', '***0***'), ('496', 'TRP', '***1***'), ('497', 'ILE', '***1***'), ('529', 'ASP', '***2***'), ('532', 'LYS', '***0***'), ('533', 'PHE', '***0***'), ('544', 'GLY', '***2***'), ('546', 'ILE', '***1***'), ('547', 'THR', '***0***'), ('549', 'SER', '***5***'), ('551', 'GLY', '***4***'), ('552', 'GLN', '***3***'), ('553', 'ARG', '***3***'), ('555', 'ARG', '***1***'), ('573', 'SER', '***1***'), ('576', 'GLY', '***2***'), ('577', 'TYR', '***2***'), ('578', 'LEU', '***0***'), ('579', 'ASP', '***3***'), ('580', 'VAL', '***1***'), ('581', 'LEU', '***1***'), ('582', 'THR', '***3***'), ('584', 'LYS', '***0***'), ('587', 'PHE', '***1***'), ('588', 'GLU', '***1***'), ('591', 'VAL', '***0***'), ('595', 'MET', '***3***'), ('605', 'SER', '***0***'), ('606', 'LYS', '***0***'), ('607', 'MET', '***1***'), ('608', 'GLU', '***1***'), ('611', 'LYS', '***0***'), ('612', 'LYS', '***0***')]

NBD2:
[('1224', 'ASN', '***0***'), ('1244', 'GLY', '***3***'), ('1245', 'ARG', '***0***'), ('1246', 'THR', '***1***'), ('1247', 'GLY', '***2***'), ('1251', 'SER', '***1***'), ('1291', 'GLN', '***3***'), ('1292', 'LYS', '***0***'), ('1293', 'VAL', '***2***'), ('1294', 'PHE', '***0***'), ('1301', 'ARG', '***0***'), ('1332', 'PRO', '***0***'), ('1336', 'ASP', '***0***'), ('1337', 'PHE', '***1***'), ('1338', 'VAL', '***0***'), ('1340', 'VAL', '***0***'), ('1345', 'VAL', '***0***'), ('1347', 'SER', '***3***'), ('1348', 'HIS', '***1***'), ('1349', 'GLY', '***2***'), ('1350', 'HIS', '***0***'), ('1370', 'ASP', '***0***'), ('1371', 'GLU', '***2***'), ('1374', 'ALA', '***0***'), ('1375', 'HIS', '***2***'), ('1376', 'LEU', '***1***'), ('1377', 'ASP', '***1***'), ('1378', 'PRO', '***0***'), ('1380', 'THR', '***1***'), ('1402', 'HIS', '***0***'), ('1403', 'ARG', '***0***'), ('1439', 'GLN', '***0***'), ('1440', 'ALA', '***0***'), ('1443', 'PRO', '***0***'), ('1444', 'SER', '***0***'), ('1446', 'ARG', '***0***'), ('1447', 'VAL', '***0***'), ('1448', 'LYS', '***0***'), ('1449', 'LEU', '***0***'), ('1450', 'PHE', '***0***'), ('1451', 'PRO', '***0***'), ('1464', 'ILE', '***0***'), ('1465', 'ALA', '***0***'), ('1468', 'LYS', '***1***')]

B

NBD1:ICL4 Interdomain Interface Residues (< 4 Å)

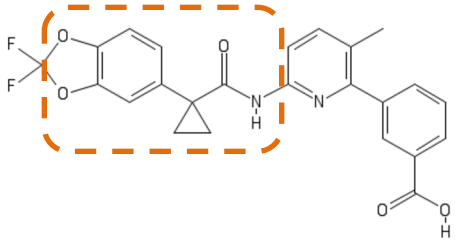
NBD1:
[('381', 'LYS', '***0***'), ('469', 'MET', '***1***'), ('472', 'MET', '***0***'), ('474', 'GLU', '***1***'), ('488', 'ILE', '***0***'), ('490', 'PHE', '***2***'), ('492', 'SER', '***1***'), ('494', 'PHE', '***0***'), ('496', 'TRP', '***1***'), ('498', 'MET', '***1***'), ('499', 'PRO', '***1***'), ('507', 'ILE', '***3***'), ('508', 'PHE', '***3***'), ('509', 'GLY', '***0***'), ('510', 'VAL', '***1***'), ('511', 'SER', '***1***'), ('560', 'ARG', '***4***'), ('564', 'LYS', '***1***')]

ICL4:
[('1057', 'THR', '***1***'), ('1060', 'LYS', '***1***'), ('1061', 'GLY', '***1***'), ('1063', 'TRP', '***2***'), ('1064', 'THR', '***0***'), ('1066', 'ARG', '***4***'), ('1067', 'ALA', '***6***'), ('1068', 'PHE', '***0***'), ('1070', 'ARG', '***3***'), ('1073', 'TYR', '***1***'), ('1074', 'PHE', '***1***')]

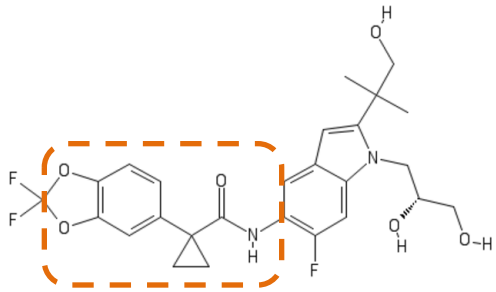
Fig. S3

A

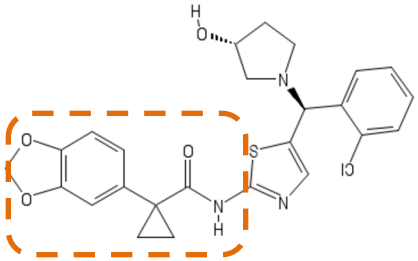
VX-809



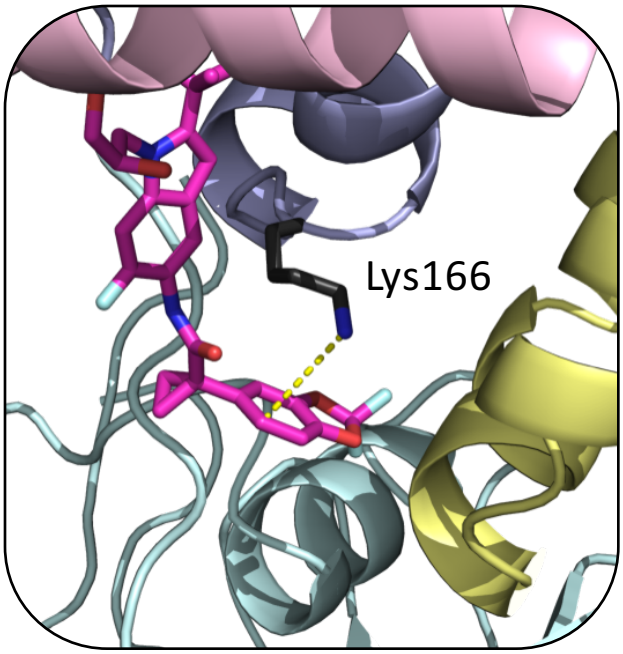
VX-661



C18



B



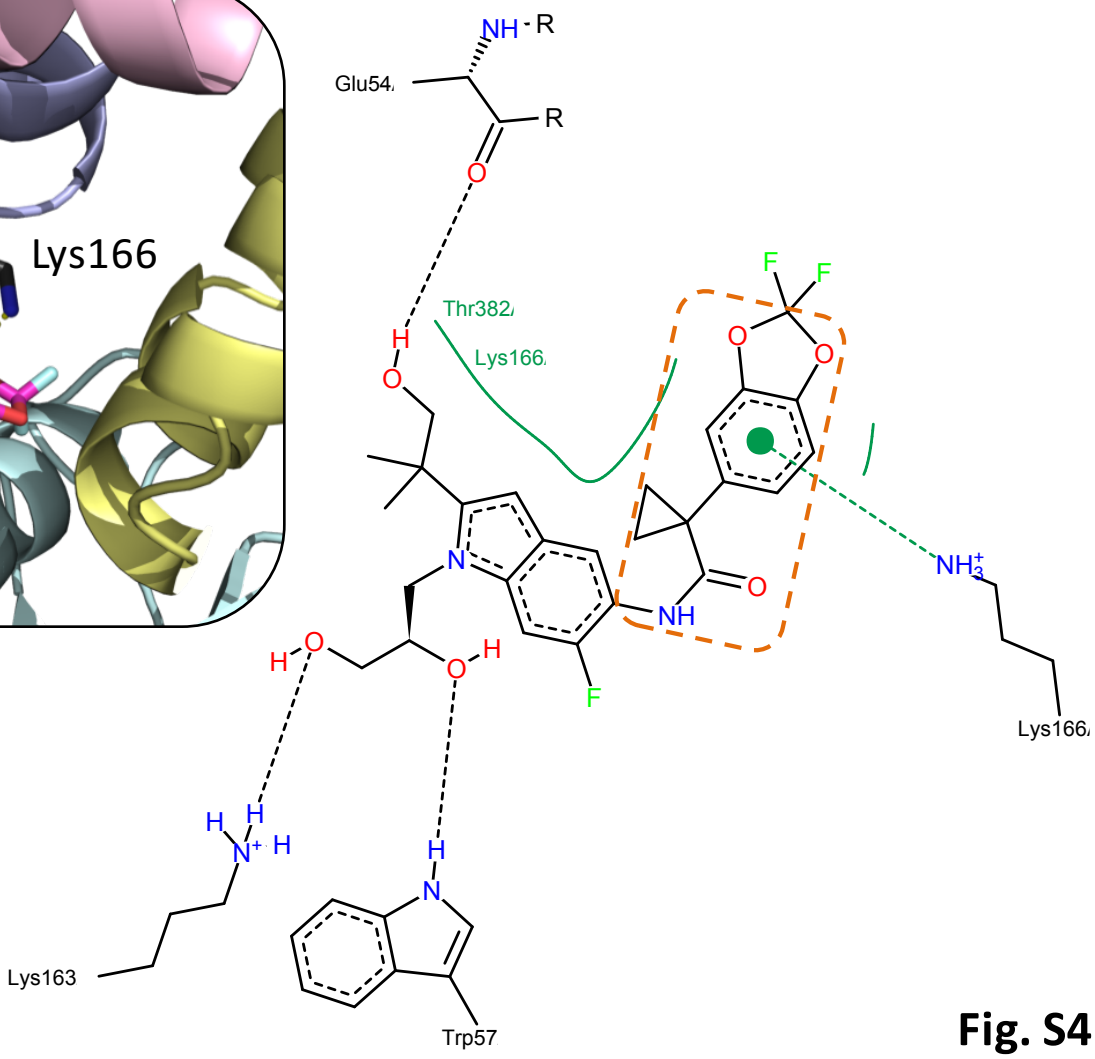


Fig. S4