Cell. Mol. Life Sci. (2009) 66:3469–3486
DOI 10.1007/s00018-009-0133-0

Cellular and Molecular Life Sciences

RESEARCH ARTICLE

# Molecular models of the open and closed states of the whole human CFTR protein

**Jean-Paul Mornon · Pierre Lehn · Isabelle Callebaut**

Received: 25 May 2009 / Revised: 17 July 2009 / Accepted: 12 August 2009 / Published online: 26 August 2009


**Abstract** Cystic fibrosis transmembrane conductance regulator (CFTR), involved in cystic fibrosis (CF), is a chloride channel belonging to the ATP-binding cassette (ABC) superfamily. Using the experimental structure of Sav1866 as template, we previously modeled the human CFTR structure, including membrane-spanning domains (MSD) and nucleotide-binding domains (NBD), in an outward-facing conformation (open channel state). Here, we constructed a model of the CFTR inward-facing conformation (closed channel) on the basis of the recent corrected structures of MsbA and compared the structural features of those two states of the channel. Interestingly, the MSD:NBD coupling interfaces including F508 (ΔF508 being the most common CF mutation) are mainly left unchanged. This prediction, completed by the modeling of the regulatory R domain, is supported by experimental data and provides a molecular basis for a better understanding of the functioning of CFTR, especially of the structural features that make CFTR the unique channel among the ABC transporters.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00018-009-0133-0) contains supplementary material, which is available to authorized users.

J.-P. Mornon (✉) · I. Callebaut (✉)
IMPMC, UMR7590, CNRS, Universités Pierre et Marie Curie-Paris 6 et Denis Diderot-Paris 7, 140 rue de Lourmel, Paris, France
e-mail: jean-paul.mornon@impmc.jussieu.fr

I. Callebaut
e-mail: isabelle.callebaut@impmc.jussieu.fr

P. Lehn
INSERM U613, IFR148 ScInBioS, Université de Bretagne Occidentale, Brest, France

**Keywords** Cystic fibrosis · Structure · Model · Channel · P-gp

## Introduction

Mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene cause cystic fibrosis (CF), the most common severe inherited genetic disease among the Caucasian population [1, 2]. The CFTR protein belongs to the large superfamily of ATP-binding cassette (ABC) transporters, which are integral membrane proteins that use the energy generated from ATP-binding and hydrolysis to translocate a wide variety of molecules across cellular membranes via an alternating access mechanism [3]. CFTR (ABCC7) belongs to the subfamily C of ABC transporters, which also includes other proteins involved in human diseases, such as the multidrug resistance-associated protein 1 (MRP1, ABCC1) and the sulfonyl urea receptor 1 (SUR1, ABCC8). CFTR is, however, the only one of the numerous members of the ABC superfamily known to function as an ion channel, as it is a phosphorylation-dependent chloride channel located at the apical membrane of epithelial cells [4, 5]. According to the general domain architecture of ABC transporters, CFTR has two membrane-spanning domains (MSDs) and two nucleotide-binding domains (NBDs). However, while prokaryotic ABC transporters have several subunits composed of identical MSDs and NBDs, the CFTR and other eukaryotic ABC proteins have distinct MSDs and NBDs, which belong to a single polypeptide encoded by a single gene. The two NBDs form tightly interacting dimers upon ATP-binding at their interface, which later on dissociate upon ATP hydrolysis [1]. These NBDs movements are transmitted to the MSDs to alternatively open and close the

gates to the substrate binding sites. In CFTR, ATP-binding to the NBDs is believed to open the ion pore, whereas ATP hydrolysis leads to the channel closure [6]. In addition, CFTR also has a large regulatory (R) domain, which links the two transporter halves and contains multiple phosphorylation sites, the channel activity being increased upon phosphorylation [7, 8].

Several CF-causing mutations, including the deletion of a phenylalanine at position 508 (ΔF508) which is the most prevalent one, prevent the protein from attaining its native global conformation. Such mutant proteins are consequently degraded by the proteasome, in contrast to the wild-type which is transported from the endoplasmic reticulum to the Golgi [9–11]. An important role of the F508 residue in the formation of inter-domain contacts was hypothesized by our modeling studies of the NBD1:NBD2 heterodimer [12, 13] and by crystal structures of the wild-type and mutated NBD1s [14, 15], which indicated that F508 was exposed to the solvent and that the ΔF508 mutation did not lead to a significant alteration of the overall NBD1 structure. This was further supported by experimental studies showing that interactions between MSDs and NBDs are disrupted by the ΔF508 mutation [16, 17].

The elucidation of the complete 3D structure of the multidrug transporter Sav1866 from *Staphylococcus aureus* [18] constituted a major advance in the overall understanding of the functions of ABC exporters and triggered studies of homology modeling of CFTR in an outward (or extracellular)-facing configuration, representing the ATP-bound state [19, 20]. Indeed, the bacterial and eukaryotic ABC exporters appear to form a relatively homogeneous family, as they share similar MSDs and NBDs, as well as long intracellular loops (ICLs) allowing the communication between their MSDs and NBDs. Even though these homology models have to be taken with caution, especially as regards some CFTR regions such as parts of the MSDs where the sequence identities are very low, they nevertheless provide a precious framework for further investigations of the CFTR functioning and for the design of pharmacotherapeutic approaches for CF. Along these lines, homology models of the CFTR MSD1:NBD1:MSD2:NBD2 assembly [19, 20] have revealed important features of the channel formed by the MSDs, as well as of the interfaces between the NBDs and MSDs. A groove on the surface of the NBD1:NBD2 heterodimer, especially rich in aromatic amino acids (among which F508), makes critical contacts with the intracellular loops 2 (ICL2 from MSD1) and 4 (ICL4 from MSD2), whereas the two other intracellular loops (ICL1 from MSD1 and ICL3 from MSD2) interact with the NBD ATP-binding sites. These studies highlighted in particular the critical position of F508 at the interface between NBD1 and ICL4. This was further supported by the observation that chemical cross-linkings could be established between non-native cysteine residues in the vicinity of F508 and ICL4 of wild-type CFTR, but that they were disrupted in the ΔF508 protein [20, 21]. It should be stressed here that, due to differences in the modeling strategies, the two models mentioned above [19, 20] do, however, differ significantly as regards the ICL4 residues involved. Most importantly, an arginine residue (R1070), located near the ICL4/NBD1 pivot, was observed in our model to establish particularly tight links with the main chain atoms of F508 and proximal residues, with almost all those bonds being lost in the ΔF508 protein [19]. These particular features of our atomic model are at present strongly supported by the recent crystal structure of mouse P-gp (MDR) at 3.8 Å resolution ([22], see "Discussion"). Of note, recent work has also shown that the R1070P and R1070W mutations lead, respectively, to the absence or to reduced levels of insertion of the protein into the apical membrane of polarized epithelial cells [23], suggesting that these mutations may, as ΔF508, affect the inter-domain interactions of CFTR. In contrast, the protein with the R1070Q mutation was found to be located at the apical membrane of the polarized epithelial cells [23]. In our model, this last mutation leads only to a relatively limited modification of the ICL4/NBD1 interface characteristics, a finding consistent with a moderate CFTR dysfunction [24]; however, a more severe phenotype has been observed when this mutation was associated with the nonsense mutation S466X [23].

Recently, several corrected X-ray structures of the ABC lipid flippase MsbA from three closely related bacterial orthologs were reported [25] (also see, for reviews, [26, 27]). The 3.7 Å resolution structure of *S. typhimurium* MsbA (StMsbA) in complex with adenylyl imidodiphosphate (AMPPNP) displayed an outward-facing conformation, which was similar to that observed with Sav1866 (<2.2 Å RMSD between the monomer Cα positions), this latter structure being identical to the crystal structure observed for StMsbA with bound ADP.Vi. Most interestingly, in addition to these two nucleotide-bound, outward-facing conformations, two inward (intracellular)-facing MsbA structures were obtained in the absence of nucleotide; these apo structures may represent the resting state of the exporter. The 5.3 Å resolution inward-facing *E. coli* MsbA (EcMsbA) structure showed fully dissociated, largely separated NBDs and was thus termed "open apo", whereas the inward-facing conformation of *V. cholerae* MsbA (VcMsbA) solved at 5.5 Å resolution was characterized by closer NBDs and thus termed "closed apo". It is noteworthy that, in the "closed apo" conformation, the NBDs, although closer than in the "open apo" conformation, display a smaller interface than in the outward-facing conformation; indeed, in the "open apo"

configuration, the P-loops of the opposing NBDs face each other, whereas the nucleotides are sandwiched between the P-loop of one NBD and the ABC signature sequence of the other NBD in the outward-facing configuration. These structures thus provide a framework to understand the main differences between the inward- and outward-facing conformations.

In the present work, we took into account the different MsbA structures and constructed a model of the human CFTR MSDs:NBDs assembly in an inward-facing conformation, which probably represents the CFTR closed channel state, and compared it to our previous model of the human CFTR MSDs:NBDs assembly in an outward-facing configuration, which is thought to correspond to the CFTR open channel state. The changes between these two physiological states of CFTR were analyzed and discussed. Moreover, the analysis of the CFTR outward-facing conformation allowed us to highlight structural features of the intracellular loops, which are lacking in prokaryotic ABC transporters and may thus account for the unique channel function of CFTR. Furthermore, we also tentatively proposed a two-domain model of the regulatory (R) domain and indicated how this model might accommodate the two states of CFTR. Finally, we also evaluated how our present model of the overall CFTR MSD1:NBD1:R:MSD2:NBD2 assembly fits with recent structural data obtained by electron microscopy [28] and cryo-electron microscopy [29], as well as with previous work with electron crystallography [30].

## Materials and methods

### Modeling of the MSD:NBD assembly

Homology modeling of the 3D structures was performed using Modeller (9v2 release) [31], according to the alignments shown in Fig. 1. Only the Cα-trace was reported for the VcMsbA closed-apo structure, which was solved at a resolution of only 5.5 Å [25]. Thus, the main chain and side chain atoms were reconstructed using Modeller [31] and the energy of the structure obtained was further refined using the Swiss-PDB-Viewer tools [32].

### Modeling of the R domain and of the N- and C-termini of CFTR

Hydrophobic cluster analysis (HCA) [33–35] was used in combination with other classical bioinformatics tools, in order to detect remote relationships between protein sequences, at high levels of divergence. HCA, which is based on a bidimensional representation of the protein sequence, combines the analysis of the sequence and of secondary structures, the regular secondary structures mainly corresponding to the hydrophobic clusters delineated by the method. The efficiency of this methodology has widely been demonstrated (see, for example, some references at http://www.impmc.jussieu.fr/∼callebau). Homology modeling was then performed on the basis of the remote templates detected, using Modeller (9v2 release) [31]. As no template could be detected for some sequence segments within the N-terminus, the C-terminus, and the R domain, those segments were modeled ab initio, taking into account the secondary structure predictions as well as well-defined neighboring segments.

### Availability of the models

The 3D coordinates of our models can be downloaded from the following URL (http://www.impmc.jussieu.fr/∼callebau/CFTR.html).

## Results

### Inward-facing conformation of the CFTR MSDs:NBDs assembly

#### *Modeling strategy*

For modeling the inward-facing conformation of the human CFTR MSDs:NBDs assembly on the basis of a MsbA 3D template, we used a global strategy very similar to that which we have previously followed for building its nucleotide-bound, outward-facing conformation, on the basis of the experimental 3D structure of Sav1866 [19]. Figure 1 shows the alignment of the MSDs and NBDs of *S. aureus* Sav1866 with those of human CFTR, as previously reported [19], to which we added the sequences of the three closely related bacterial MsbA orthologs already mentioned above: *E. coli* Msba (EcMsbA), *V. cholerae* MsbA (VcMsbA), and *S. typhimurium* MsbA (StMsbA). The Sav1866 and MsbA sequences share 22 and 51% identity within their MSDs and NBDs, respectively, with few insertions and deletions. Sequence identities between CFTR and Sav1866 are comparable to those between CFTR and MsbA (CFTR(A)/Sav1866: 14% (MSD) and 26% (NBD); CFTR(A)/VcMsbA: 10% (MSD) and 22% (NBD); CFTR(B)/Sav1866: 12% (MSD) and 27% (NBD); CFTR(B)/VcMsbA: 12% (MSD) and 27% (NBD)).

As already indicated above, MsbA can undergo a large range of motions between the apo and nucleotide-bound states [25]. We chose here to model a possible, inward (cytoplasmic)-facing conformation of CFTR, likely representing the closed state of the channel, on the basis of the closed-apo 3D structure of VcMsbA (pdb 3b5x, [25]), in

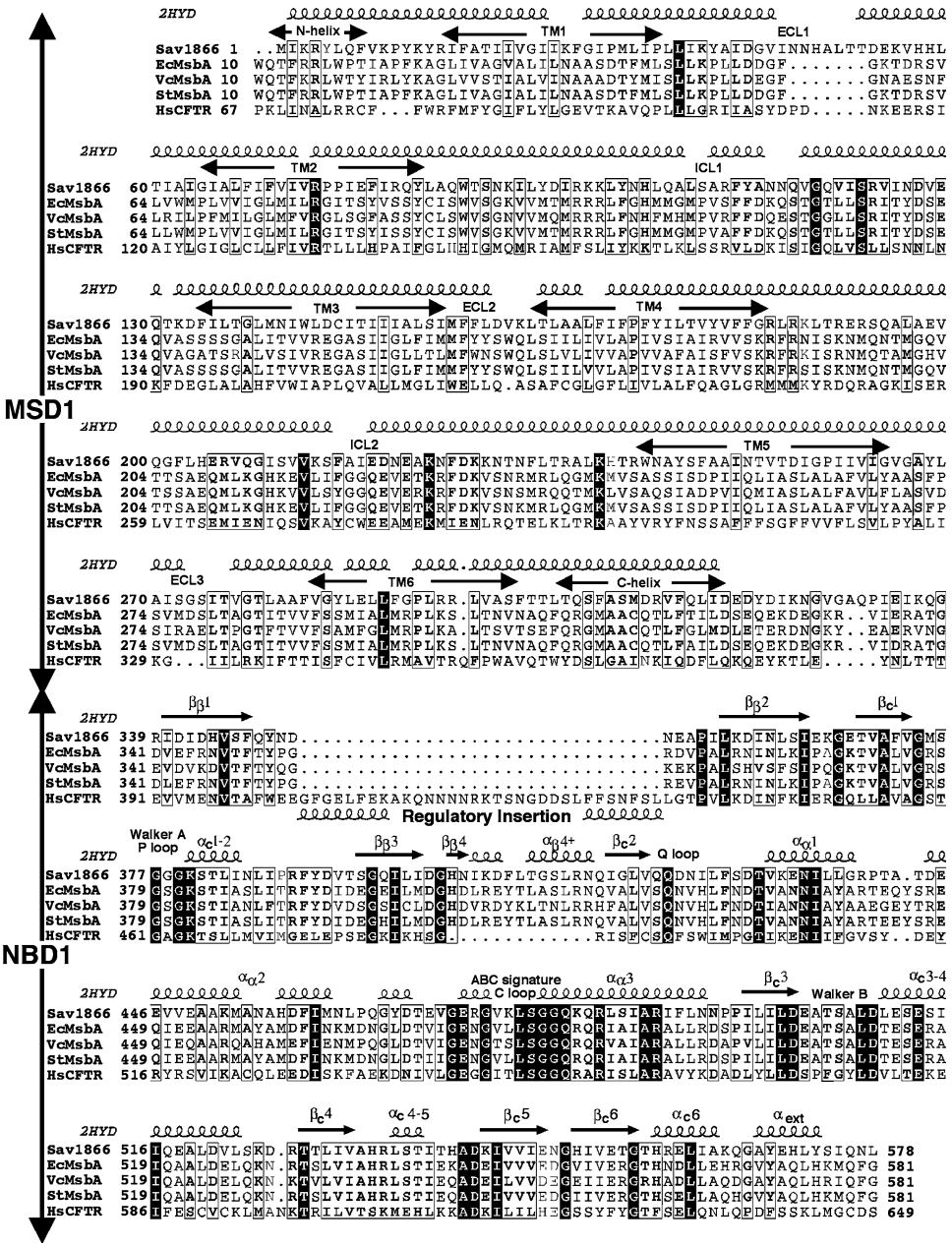


**Fig. 1** Alignment of the human CFTR MSDs and NBDs sequence with sequences of bacterial ABC exporters, whose 3D structures are known in different conformations [*S. aureus* Sav1866 (pdb 2hyd) and *Escherichia coli*, *Vibrio cholerae*, and *Salmonella typhimurium* MsbA (pdb 3b5w, 3b5x, and 3b60, respectively)]. **a** MSD1/NBD1; **b** MSD2/NBD2. The MsbA sequences were added to the Sav1866/CFTR alignment, which is identical to that previously reported [19]. The observed secondary structures of Sav1866 are indicated on top of its sequence. The MSD and NBD secondary structures are labeled according to the labeling previously reported for CFTR [12, 19]. The α helices of the CFTR NBD1-specific regulatory insertion (RI) are reported below its sequence. Identities are shown in *white* on a *black background*, whereas similarities are *boxed*. This figure was made using ESPript [65]

which the NBDs are not completely separated. Indeed, we thought that this conformation may be more relevant than any other mode of NBD separation, for several reasons. First, the CFTR-specific NBD1 regulatory insertion (RI) can indeed interact, in this closed-apo conformer, with the modified LSHGH signature sequence of NBD2 (see below). Moreover, it is likely that the R domain (which is C-terminal to NBD1) might prevent the complete dissociation of the CFTR NBDs, similar to what was observed in the structure of MalK NBDs in the absence of nucleotides [36]. Another factor which may hinder the complete dissociation of the NBDs is the very poorly hydrolysable ATP, which remains tightly bound to the NBD1 non-canonical binding site for several minutes [12]), a time

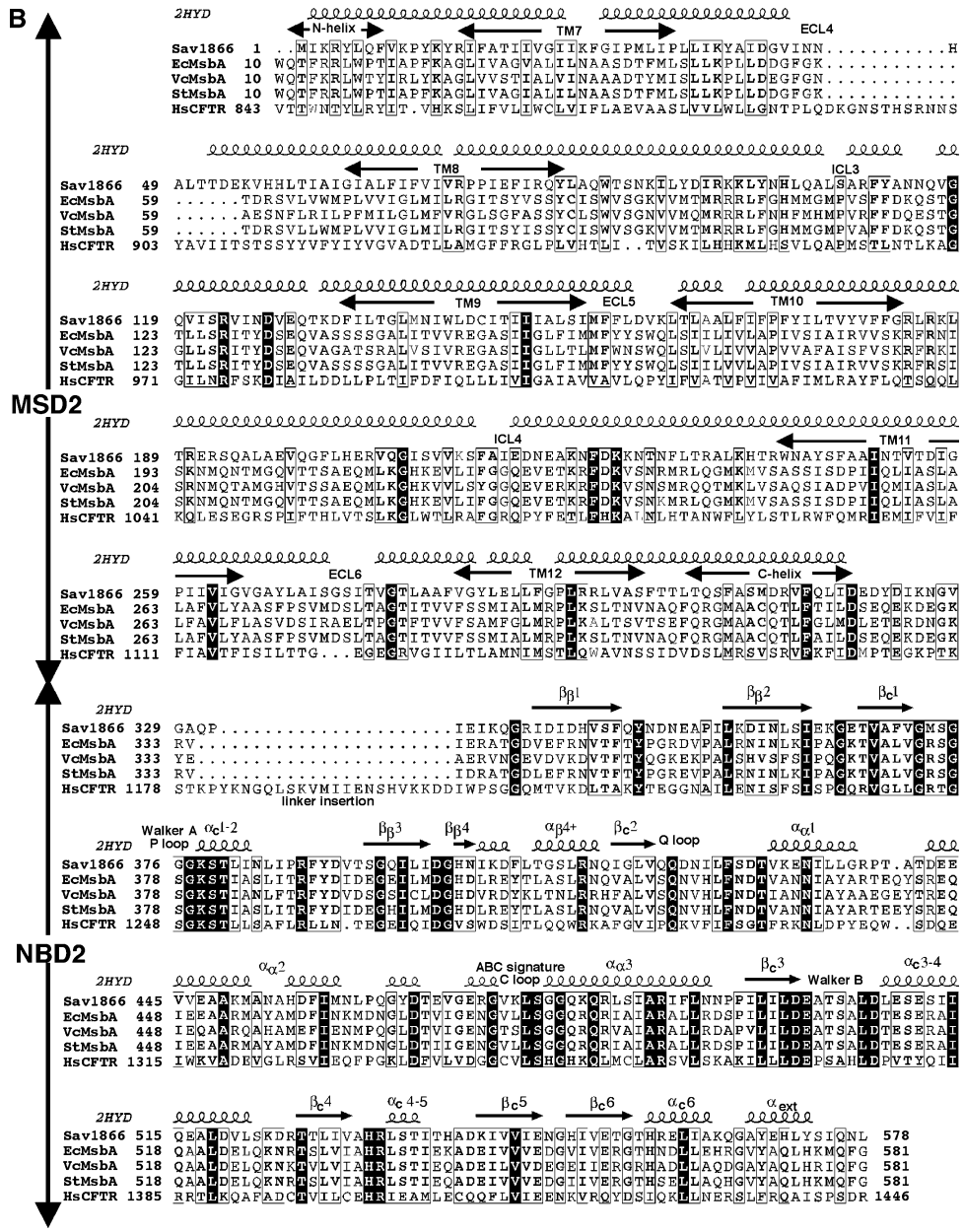


**Fig. 1** continued

during which several gating cycles can occur [1]. Finally, one can also argue that a wide separation of the NBDs may not be required for the functioning of CFTR as a chloride channel, while a large opening resulting from a complete and wide separation of the NBDs may be important for MsbA and other ABC exporters to allow access of very large substrates to the accessibility chambers.

Such a "closed-apo"-based model of CFTR is further supported by electron crystallography data [30], which showed a good agreement of the low resolution electron density of the 2D, ATP-free crystal 1-form of CFTR with the closed-apo VcMsbA structure solved at that time. Although that MsbA structure was incorrect, the whole volume that it occupies is similar to that occupied by the corrected VcMsbA structure which has been published since [25]; this is clearly also in favor of a limited separation of the NBDs for the apo configuration of CFTR (see also "Compatibility of the whole CFTR models with low resolution experimental data", below).

It should be noted that the arrangement of the transmembrane helices in the MsbA-based closed-apo model (see Fig. S1 in Electronic supplementary material, ESM), which pulls them nearly perpendicular to the membrane plane, still does not lead to a complete closing of the channel, in a

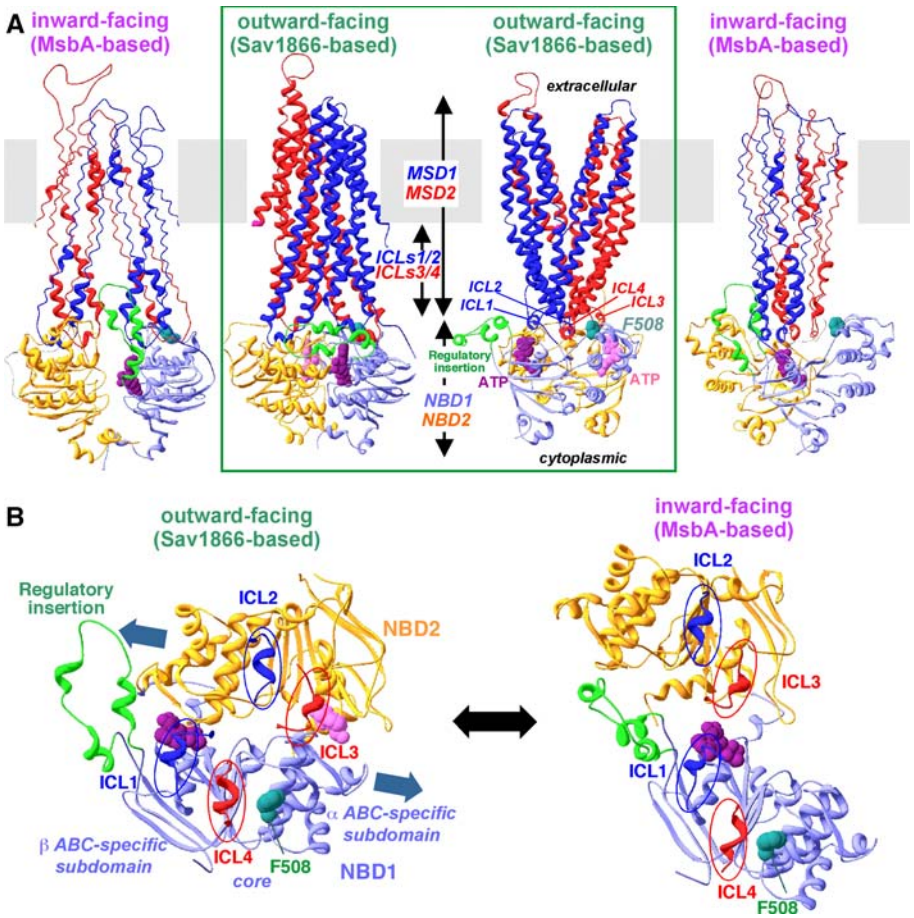


**Fig. 2** The CFTR MSD1:NBD1:MSD2:NBD2 assembly, in the outward- and inward-facing conformations. **a** Overall architecture. Two *orthogonal* views using a *ribbon* representation are shown for each of the two models. MSD1 and NBD1 are colored *dark* and *light blue*, respectively, whereas MSD2 and NBD2 are colored *red* and *orange*, respectively. The NBD1-specific regulatory insertion (RI) is shown in *green*. F508 (in *green*) is shown in a van der Waals representation, as well as the ATP molecules (*dark* and *light pink* for the ATP molecules bound to the non-canonical (NBD1) and canonical (NBD2) sites, respectively). Although the inward-facing conformation of MsbA is free from ATP molecules, we chose to leave one ATP molecule in the non-canonical ATP-binding site in the model of the inward-facing conformation of CFTR (closed channel), as ATP stays for a long time at this site. The position of the membrane lipid bilayer is indicated in *grey*. *MSD* membrane-spanning domain, *NBD* nucleotide-binding domain, *ICL* intracellular loop. **b** The NBD1:NBD2 assembly (viewed from the membrane). The NBD1/NBD2 heterodimer is shown, as well as the coupling helices of the MSDs intracellular loops (ICL1–ICL4) which make contacts with the NBDs. The transition between the two forms of the NBD1:NBD2 assembly is obtained by an almost horizontal rigid-body translation (sliding, see *blue arrows*) of NBD2, the regulatory insertion likely playing the role of a "safety catch" which hinders the complete disassembly of the two NBDs. ATP (*dark pink*) is expected to be bound in the non-conventional binding site (NBD1) in both conformers, whereas the conventional binding site is only occupied by ATP (*light pink*) in the outward-facing model (open channel). However, one can note that in the inward-facing configuration, both ATP molecules can be loaded in the canonical and non-canonical binding sites of the NBDs with no steric hindrance, in close contact with each other, their O3′ atoms being separated by only 4.5 Å. Such a situation might actually occur during the transition between the closed, inward-facing and the open, outward-facing conformations

manner similar to that observed in the open-apo experimental structure of MsbA, in which the NBDs are fully dissociated. Most importantly, however, we observed that a small rigid-body rotation of the two MSDs of our CFTR closed-apo model around the ICL pivots, which are parallel to the membrane plane and parallel to each other (ESM: Fig. S1) could lead, without any further movement of the NBDs, to the full closure of the CFTR channel similar to that observed for the MsbA open-apo conformer (ESM: Fig. S2). This new conformer was energetically refined as the other ones and did not show any steric clash. We thus considered that this "tight closed-apo" conformation (shown in Fig. 2a, left and right panels) may well be representative of the closed state of the CFTR channel.

## General features of the model

Figure 2a shows the resulting whole model of the MsbA-based, inward-facing CFTR MSDs/NBDs assembly, compared to our previous outward-facing model constructed

using the Sav1866 experimental structure as template. Of note, the difference between the large and thin ribbons of the inward-facing model is due to the low resolution of the MsbA template structure (5.5 Å), which often does not provide canonical parameters for α-helices in the MSDs and therefore does not allow their automatic recognition. Figure 2b compares the global NBD dimeric organization (viewed from the MSDs) of the outward- and inward-facing conformations of CFTR and highlights the dimer contact points with the intracellular loops (ICLs). As for the MsbA closed-apo structure [25], the model of the inward-facing configuration of CFTR that we propose here is characterized, when compared to the outward-facing configuration, both by a twisting of the transmembrane helices (Fig. 2a) and by a sliding of the NBDs (Fig. 2b). Interestingly, the contact points of the ICLs with the NBDs (especially ICL2 and ICL4 with NBD2 and NBD1, respectively) appeared to remain unchanged (Fig. 2b), these contacts constituting stable anchors or pivots around which the conformational reorganization of the MSDs and NBDs can occur.

## The NBDs

The experimental 3D structure of the closed-apo, inward-facing configuration of MsbA was observed in the absence of nucleotides [25]. The MsbA-based model of the closed-apo configuration of the human CFTR NBD1/NBD2 assembly is shown on the right in Fig. 2b. In CFTR, in contrast to ABC transporters bearing only canonical ATP-binding sites, an ATP molecule may actually remain bound to the non-conventional binding site (NBD1) during many gating cycles and may thus be located in our closed-apo CFTR model at a possible junction between NBD1 (blue in Fig. 2b) and NBD2 (orange). Such a binding, through a half nucleotide-binding pocket (including the Walker A and Walker B motifs, but not the ABC signature sequence) is supported by the fact that ATP was observed bound to the isolated NBD1 crystal structure [14]. Of note, in such a model (Fig. 2b right), the half nucleotide-binding pockets, formed by the Walker A/Walker B sides, are now in close contact. In contrast, the ABC signature motifs, which in the outward-facing conformation form the other sides of the nucleotide-binding sites, are here rejected outside the NBD heterodimer interface. In our Sav1866-based outward-facing conformation of CFTR, the large regulatory insertion (RI), located between strands $\beta$1 and $\beta$2 of the ABC-specific $\beta$ sub-domain of NBD1 (green ribbon in Fig. 2b), was modeled using as template the isolated CFTR NBD1 experimental structure, with several adaptations toward a "semi-open" conformation in order to avoid steric clashes with NBD2 in the NBD1:NBD2 heterodimer structure (supplementary data 2 in [19]). In the present MsbA-based inward-facing configuration, it is likely that the RI has yet to move upward to a "completely open" conformation, in order to avoid steric hindrance problems with the NBD2 subunit in its novel position (Fig. 2b). Interestingly, this "completely open" RI conformation is likely to make new contacts with the NBD2 non-canonical ABC signature sequence (LSHGH). This potential interaction likely involves the RI region harboring the phosphorylated serine 422 residue, but the detailed atomic characteristics cannot be deduced here because of uncertainties of the model in this region. Nonetheless, this interaction might hinder the complete separation of the NBDs.

## The transmembrane helices and the CFTR channel

One of the most striking features of the Sav1866-based model of the CFTR MSDs in an outward-facing configuration is the intricate association of the two transporter halves. Indeed, the bundles of transmembrane helices diverge into two wings, each containing the two first helices of one half-transporter and the four following ones of the other half, a phenomenon called "swapping" (i.e., for the first and second wings, respectively: TM1 + TM2 + TM9 + TM10 + TM11 + TM12 and TM7 + TM8 + TM3 + TM4 + TM5 + TM6; Fig. 3). Here, in the MsbA-based model of the CFTR inward-facing configuration, the split into two wings is no longer observed, as the transmembrane helices now adopt an orientation that is, on average, perpendicular to the membrane plane (Fig. 3). Even though one can observe a significant spatial rearrangement of the transmembrane helices packing between the outward- and inward-facing conformations, their overall topology (relative location of a helix vs another one) remains similar.

Importantly, amino acid residues, which were previously shown to line the channel pore in our Sav1866-based outward-facing CFTR model [19], kept a similar orientation in the present inward-facing model (Fig. 3, and ESM: Fig. S3). Figure 3 indicates most of the critical residues localized in TM1 (K95, Q98, P99), TM6 (R334, K335, F337, T338, S341) and TM12 (T1134, T1137, N1138, S1141), as well as those recently highlighted in the TM11–TM12 loop (S1118, T1121, T1122, G1127, V1129, I1131, and I1132) [37]. As already predicted by the outward-facing model, residues F337, T338, and S341, which play a crucial role in the chloride channel conductance [38–40], are located in a narrow region of the pore, likely constituting a constriction in the channel. Moreover, this constriction seems to be associated with a bending of the linear ion-conducting pathway, suggesting that it may participate in gating of the channel. R117 and R104 of the extracellular loop 1 (ECL1) may be exposed at the extracellular surface of the channel, at the entry of the pore

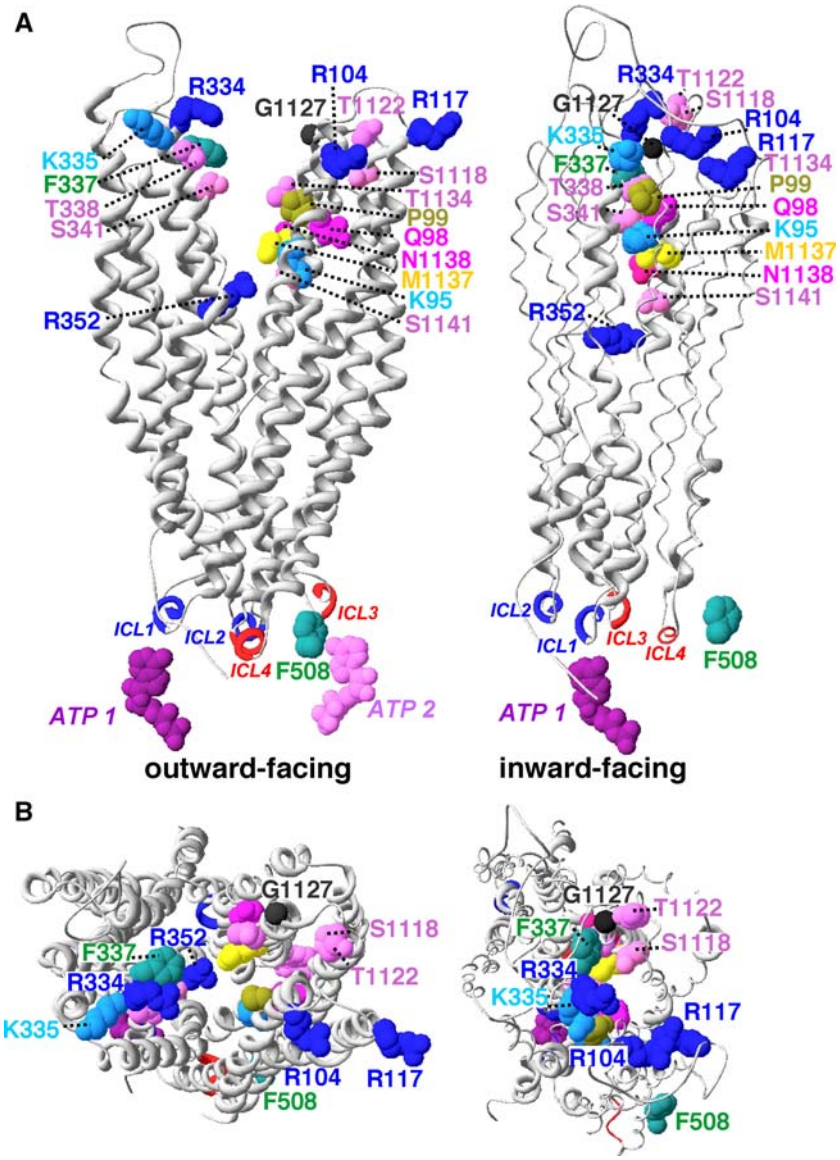


**Fig. 3** Overall architecture of the CFTR chloride channel, in the outward-facing and inward-facing configurations (open and closed channel, respectively). Two *orthogonal* views of the MSDs are shown, with the coupling helices of the four ICLs colored in *blue* (MSD1 ICL1 and ICL2) and *red* (MSD2 ICL3 and ICL4). **a** Longitudinal side views, **b** top views from the extracellular side. As in Fig. 2, two ATP molecules are shown in the outward-facing conformation, whereas only one (in the non-conventional binding site: NBD1) is shown for the inward-facing conformation. Critical amino acid residues lining the channel pore are shown in a van der Waals representation

(Fig. 3, and ESM: Fig. S3). They thus probably participate in the outer mouth, as R334 and K335 or S1118, T1122, and G1127 (TM11–TM12 loop). Altogether, these residues may thus constitute important determinants of anion-binding sites, as experimentally shown for R334 [41] as well as for R104 and R117 [42]. Two other basic residues occupy critical positions in the middle and at the intracellular end of the channel, respectively. K95 (located in the middle of the channel) is involved in the anion selectivity and permeation of the channel, and plays, together with Q98 and P99, a role in unitary conductance and macroscopic current properties [4, 43, 44]. R352, which is also involved in $Cl^-$ channel conductance [45, 46], is located at the cytoplasmic mouth of the pore.

Comparison of the features of the outward-facing and inward-facing conformations (Fig. 3) demonstrated the clear closure of the CFTR channel in the inward-facing conformation (see above), thus supporting the assumption that this conformer may indeed actually lead, at the end of the gating cycle, to the closed state of the channel. However, intermediate structures between this "tight" inward-facing conformation and the initially considered closed-apo structure (ESM: Fig. S1) may transitionally participate in the anion transport.

Finally, it should also be stressed here that an in-depth examination of the CFTR channel in its outward-facing conformation allowed us to highlight novel important structural features within the ICLs which may account for the opening of the channel at the cytoplasmic-side (in contrast to the 3D structures of bacterial exporters). This further supports the assumption that the outward-facing conformation does indeed correspond to the completely

open state of the CFTR channel. Because of its speculative nature, this is discussed in detail in the "Discussion".

## 3D model of the CFTR R domain

### *General considerations*

In contrast to the MSD:NBD assembly, which clearly belongs to the ABC family of exporters (even though sequence identities are often very low), no structural template has yet been described on which the structure of the R domain could be tentatively modeled. Many studies have considered that the regulatory R domain of CFTR (consisting of ~194 amino acids, from residue 646 to 843), which contains multiple phosphorylation sites, is intrinsically disordered [8, 47]. Among many other functions, the CFTR R domain has, from a structural point of view, to link the end of NBD1 (approximately amino acid 645) to the beginning of MSD2 (~residue 841). The shortest distance between these two extremities is ~76 and ~63 Å in the Sav1866-based (outward-facing) and MsbA-based (inward-facing) models, respectively. The segment linking NBD1 to MSD2 has, however, to get around the NBD2 globular domain, increasing thereby the minimal distance involved to ~100 Å. As reported in Table 1 in the ESM, proteins such as human MDR1 (P-gp) and human SUR1, which are related to human CFTR but do not have a CFTR-like regulatory R domain, have linkers whose length ranges from 58 to 93 amino acids. These relatively short linkers contain actually only few strong hydrophobic amino acids (V, I, L, F, M, Y, W) but many residues known to favor loops (P, G, D, N, S) [34]. As a consequence, they mainly lack stable regular secondary structures and thus may well correspond to intrinsically disordered regions. This was indeed reported to be the case in the very recently published experimental structure of mouse P-gp (MDR [22], see "Discussion"). It should be stressed here that the CFTR R domain immediately appears quite different from those minimal linkers, as its sequence is clearly longer and globally contains more strong hydrophobic amino acids, often participating in hydrophobic clusters typical of stable regular secondary structures. Thus, it is probable that, in the full structural context of CFTR, the R domain may be, at least in part, relatively structured. Such a hypothesis is in agreement with recent simulations showing that the R domain may be more compact than had been anticipated [48].

### *The C-terminal part of the R domain: modeling strategy*

The R domain joins the first and second halves of the CFTR protein, each made of a membrane-spanning domain (MSD) and a nucleotide-binding domain (NBD). These two halves likely occurred by duplication of an ancestral MSD–NBD module, as suggested for MDR [49]. According to this duplication hypothesis, we suggest here that the sequences immediately preceding the MSDs [i.e., the CFTR N-terminal sequence (preceding MSD1) and the C-terminal part of the R domain (immediately preceding MSD2 and herein named the "R2" domain, residues ~775–840)] may also share a similar 3D arrangement. This hypothesis is supported by sequence similarities highlighted by hydrophobic cluster analysis (HCA) and by the good agreement observed between the secondary structures predicted for the two segments (ESM: Fig. S4). Moreover, these predictions are in good agreement with the secondary structure propensities of the R2 domain calculated on the basis of experimental NMR data [47] (ESM: Fig. S4). The R2 domain is characterized by 29% strong hydrophobic amino acids (V, I, L, M, F, Y, W) and by a ratio between hydrophobic and loop-forming residues of 1.14; such values are close to those generally found for well-structured domains (e.g., 30–32% and 1.35–1.39, respectively, for the CFTR NBDs). Given the absence of experimental 3D structures that could serve as templates, the R2 domain (~65 amino acids) was actually modeled ab initio, by taking into account the secondary structure propensities reported for the phosphorylated protein [47] and by masking the hydrophobic patches displayed on the surface of the NBD heterodimer in the outward-facing CFTR conformation with the hydrophobic faces of the ab initio modeled segments. According to our hypothesis of a relationship between the N-terminal sequence of CFTR and its R2 domain, both segments should be similarly structured, with the very N-terminus located in the vicinity of the CFTR C-terminal end, which follows NBD2.

### *The N-terminal part of the R domain: modeling strategy*

On the other hand, we suggest here that the N-terminal part of the R domain (called herein the "R1" domain) may be structurally variable, but that a possible structure it may adopt might be related to the structure of regulatory domains located immediately after the NBDs in some bacterial ABC transporters (e.g., the maltose importer MalK [50] and the molybdate/tungstate ABC transporter ModC [51]), with which we identified potential similarities using HCA. These regulatory domains, which are located at the cytoplasmic bottom of the NBD homodimers, are known to be made of two similar, consecutive "TOBE" sub-domains (transport-associated OB, Pfam domain database PF03459, [52], classified in the SCOP structural classification database within the MOP (molybdate/tungstate binding protein) superfamily b40.6]. Such five-stranded TOBE sub-domains consist of four-stranded "Greek key" motifs for which the fourth strand of each

sub-domain is locked by a fifth strand of the other sub-domain (ESM: Fig. S5). For clarity purposes, we herein call "R1a" and "R1b" the first and second "TOBE" sub-domains, respectively. The $\beta$16–$\beta$17 and $\beta$20 strands (TOBE labeling) in the R1b sub-domain appeared as the best markers for a possible link between the CFTR sequence and those of the regulatory domains of other ABC transporters (ESM: Fig. S6); this R1b sub-domain also displayed globally a better general alignment. On another hand, this R1b sub-domain almost perfectly matches the secondary structure propensities deduced from experimental NMR data for the phosphorylated form of the R domain [47] (ESM: Fig. S6). Also, according to the TOBE hypothesis, strand $\beta$20 within the R1b sub-domain locks strand $\beta$14, which corresponds to the fourth strand of the Greek-key motif of the R1a sub-domain. This R1b sub-domain strand $\beta$20 can readily be suggested in the CFTR sequence (ESM: Fig. S6). However, discrepancies appear to exist for the first part of the R1a sub-domain, in which the two small $\beta$12 and $\beta$13 strands may be substituted, according to the published secondary structure propensities, by an $\alpha$-helix (between L662 and F669) (ESM: Fig. S6). These particular sequence features suggest that the N-terminal part of the R1a sub-domain might also be associated with a different fold and that a different location of the whole R1 domain relative to the whole MSD:NBD assembly might also be considered, as discussed below and in ESM: Fig. S7.

The R1 "globular-like" domain was thus tentatively modeled using as templates the structures of Glcv (pdb 1oxs) [53] and PH0194 (pdb 2d62) (Lokanath N. K. et al., unpublished data), the sequences of which are the most similar to the CFTR sequence. The template structures were respected, except for two regions (ESM: Fig. S6): (1) strands $\beta$12 and $\beta$13 ($\beta$-strands numbering according to the labeling of the templates) were replaced by an $\alpha$-helix, and (2) the insertion present in human CFTR between amino acids 680 and 690 was also modeled as an $\alpha$-helix, according to the secondary structure propensities of phosphorylated R [47]. It is also worth noting that the upstream part of the R1 domain contains the so-called regulatory extension (RE) observed in the experimental structure of mouse NBD1 and found to be structured as an alpha helix packing against the NBD1 core. This RE, which contains phosphorylable serine residues (S660, S670), may in fact fold differently in the full length CFTR protein, as it may directly participate in the R1 core domain.

Finally, the sequence linking the R1 and R2 sub-domains of the CFTR R domain (R1–R2 linker, 28 amino acids) was also modeled ab initio, according to the secondary structure propensities reported for the phosphorylated protein [47] and to the same restraints as for the R2 domain. The short C-terminal extension of CFTR was also modeled ab initio. It includes a segment, PSDRVKLFP (1443–1451), which likely adopts an extended configuration ($\beta$ propensity of the corresponding hydrophobic cluster = 1.93 vs $\alpha$ propensity 0.85 [34]), while another segment, extending from residue 1,462 to 1,480, may correspond to an $\alpha$-helix. The extended region of the C-terminal end of CFTR may interact with this helix, as well as with the NBD1 helix following the Walker B motif ($\alpha_{core}$3–4; [12]).

### *3D model of the R domain: results*

Figure 4 (and ESM: Fig. S8) shows the resulting tentative model that we propose here of the R domain and of the N-terminal extension of human CFTR, within the context of the MSDs:NBDs assembly (shown as solvent accessible surfaces). Basically, as described before, the CFTR R domain may thus consist of (1) a globular-like R1 sub-domain exhibiting two TOBE-like structures, and (2) a R2 sub-domain, which similarly to the CFTR N-terminus, is expected to run (from the N-terminal to the C-terminal end) upwards until the MSDs and may consist of a series of alpha helices and extended segments. Figure 4 also shows that the R2 sub-domain and the CFTR N-terminus, which likely occur from a duplication event, may consist of three $\alpha$-helices separated by two extended regions ("ext"). Interestingly, the helix $\alpha$3 of the N-terminus may interact with ICL1 and the link between MSD1 and NBD1, while the extended region of the N-terminus ("ext2") may also interact with this link, as well as with the N-terminus helices $\alpha$1 and $\alpha$2, thereby forming a small sub-domain. Symmetrically, the R2 sub-domain may run around NBD2, with the R2 extended regions "ext1" and "ext2" as well as the R2 helix $\alpha$2 interacting with the large region (in yellow) which links MSD2 and NBD2, all these segments likely forming together a small globular domain. Helix $\alpha$1 of the R2 subdomain offers the transition towards the probable helix located in the R1–R2 link, which may interact with the second TOBE-like subdomain (grey).

The R1–R2 sub-domains were modeled on the Sav1866-based model of CFTR and found to fit to this structure. Interestingly, after we had built the MsbA-based (inward-facing conformation) model of CFTR reported herein, we observed that the proposed R1–R2 path was fully compatible with this new model and that there was no steric clash, except for a small displacement of part of the R2 sub-domain, shown in pink in Fig. 4. Of note, R1 and the first part of R2 (shown in light blue in Fig. 4) contact one side of NBD2 in the outward-facing model (mainly the NBD2 ABC-specific $\alpha$ sub-domain) and the other side in the inward-facing configuration (from Y1219 to 1445, and principally the $\alpha/\beta$ core).

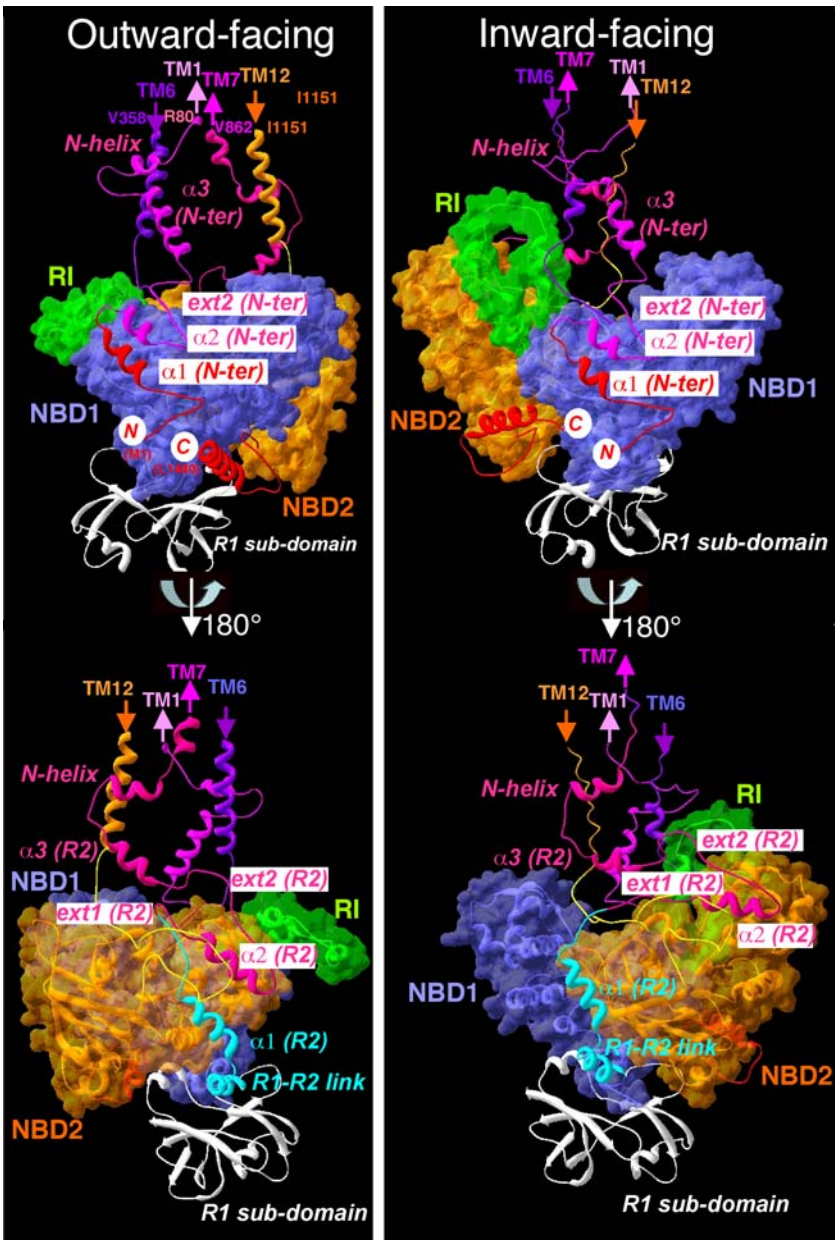


**Fig. 4** The R domain model, in the outward-facing (*left*) and inward-facing (*right*) configuration. Two opposed side views at 180° (*upper panels* vs *lower panels*) of NBD1 (in *blue*, with its regulatory insertion (RI) in *green*) and NBD2 (in *orange*) showing the R1 sub-domain (*white*) as well as the helices and extended strands of R2 sub-domain and of the CFTR N-terminus (which was used as template for modeling the R2 sub-domain, see *text*). The N-terminal and C-terminal ends of CFTR are indicated in *red font* on a *white circle* background

## Compatibility of the whole CFTR models with low resolution experimental data

Recently, a first clear 3D structural reconstruction of CFTR obtained at 20 Å resolution by electron microscopy (EM) was reported [28], offering an experimental validation tool for modeling studies. This work supported a "tail-to-tail" dimeric assembly of the complete mature CFTR molecule (169 kDa + ~8 kDa of associated glycans), in a closed state, which was observed as 162 × 120 × 106 Å ellipsoïdal particles possessing a two-fold axis along their large dimension. Interestingly, the size and the main features of the EM-based reconstructed CFTR are in good agreement with the dimers built by using our MsbA-based, inward-facing conformation model of CFTR (closed channel), as illustrated in Fig. 5 and in ESM: Fig. S9. Mio and colleagues [28] also demonstrated (1) that, as we propose it in the present model, at least a large part of the R domain was located in the observed particles at the bottom of the cytosolic architecture, and (2) that orifices were present beneath the ICLs, a feature also observed in our model.

Even more recently, during the revision of this manuscript, an article by Zhang et al. [29], following a previous work [54], reported the architecture of dimeric CFTR particles using cryo-electron microscopy (see Fig. 4 in [29] which describes the main features of the dimers). According to the dimensions of the Sav1866 structure fitted

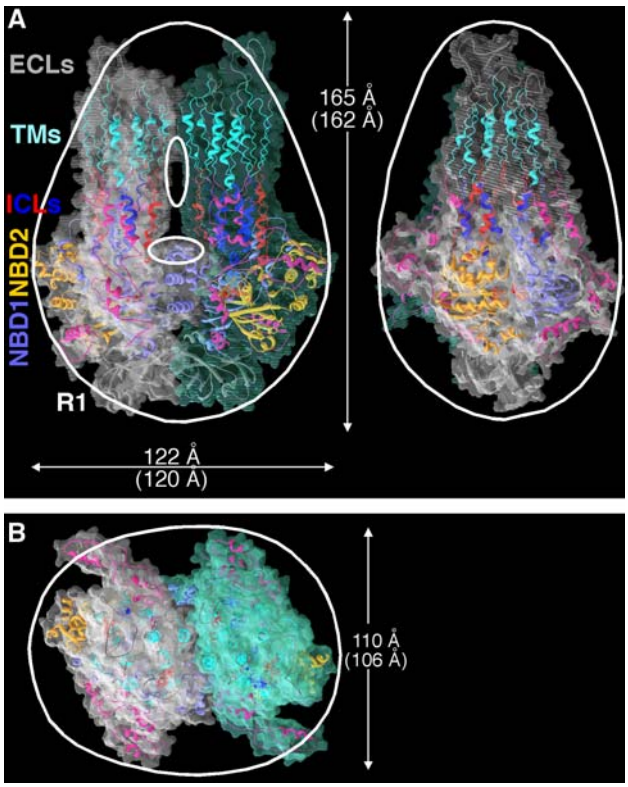


**Fig. 5** Compatibility of a dimeric assembly of the whole CFTR model in its closed, inward-facing configuration with ellipsoidal structures observed for the CFTR dimer by electron microscopy. **a** Two *orthogonal* side views; **b** view from the extracellular side. The contours of the ellipsoïdal particle are shown in *white*, as deduced from Fig. 8 of the article by Mio et al. [28]. The estimated dimensions of a dimeric assembly of our CFTR model are indicated above those observed by electron microscopy, which are shown between *parentheses*. For the two *upper orthogonal* side views, the structures encountered from the *top* to the *bottom* are *colored* as follows: ECLs in *white*, TMs in *light blue*, ICLs in *dark blue* and *red*, NBD1 in *blue* and NBD2 in *orange*. The N- and C-terminal segments, as well as linkers, are in *pink*. At the *bottom*, the R1 domains are shown in *white*. The interface between each monomer mainly includes, from *top* to *bottom*, contacts between TM10 and TM10, $\alpha_2$2 of NBD1 and R2, as well as R1 and R1. Glycans are localized on asparagine residues N894 and N900, in the first ECL of the second MSD (at *top* of the particle)

into the CFTR experimental density map, it appears that this dimeric architecture is closely related to that observed by Mio et al. already discussed above [28]. Indeed, the sizes of the observed particles are similar as regards two of the three dimensions: 123 Å wide versus 120 Å and 106 Å deep versus 106 Å, with a packing of the fitted atomic model very similar to that reported in Fig. 5, as well as similar locations of the orifices. However, there is an obvious difference for the largest dimension along the CFTR pseudo two-fold axis (132 Å in [29] vs 162 Å in [28]). This difference may likely be explained by a different packing of the R region, which appeared located between the NBDs and MSDs near the ICLs region in the work by Zhang et al., whereas it was mainly located at the bottom of the particles in the report by Mio et al. Thus, these two reports described, for the first time, two possible states for the R domain, a feature which is not surprising for a region known as highly mobile. It is, however, noteworthy here that the location of the R domain may be dependent on the experimental conditions, rather than be biologically relevant. On the other hand, the mobility of the R domain is also consistent with an alternative folding of the R1a sub-domain, as proposed herein. Indeed, this region, which corresponds to the first TOBE-like domain and appears to be the less related to cognate TOBE domains (ESM: Fig. S6), might actually be unfolded and thereby cover the distance between the NBD1 end and the ICLs region, where the folded part of the R domain would then be located. As regards this hypothesis, it is worth noting that the yeast CFTR-related Yor1p protein possesses a R region, whose size is limited to that of the first TOBE domain of CFTR, and whose sequence texture is very similar to the CFTR first TOBE-like domain (ESM: Fig. S7). Taken altogether, these considerations support the assumption that the structural features of this part of the CFTR sequence may be highly versatile, being able to switch from relatively well-structured states to extended ones.

The good fit highlighted above between the aforementioned microscopy data and our CFTR models is further supported by the low resolution electron microscopy results obtained with negatively stained, 2D crystals of CFTR [30] (see Fig. 8 in [30], showing the three orthogonal views of CFTR, crystallized in the absence of nucleotides, in which was fitted the structure of the closed-apo conformation of VcMsbA, as solved at that time). Although incorrect, this nowadays obsolete VcMsbA structure has overall dimensions very similar to those of the corrected structure of the same VcMsbA protein reported afterwards [25], which was used here to model the inward-facing configuration of CFTR. Therefore, our model of the inward-facing, closed configuration of CFTR appears consistent with these data. Moreover, the shapes of the MSDs of the two CFTR states (with and without nucleotides) shown in Fig. 9 in [30] are in good agreement with the two models compared herein. In particular, the MSDs of the nucleotide-free crystal form (Fig. 9a in [30]) shows a hexagonal shape, with a central link between two holes. These features match our MsbA-based closed model of CFTR, in which the central link is constituted by helices 6 and 12, and the holes are located between helices 6 and 12 and helices 1 and 11 at one side, and helices 5 and 7 at the other side. The MSDs of the nucleotide-bound form of CFTR are characterized by a triangular shape with a large central hole (Fig. 9b in [30]), a feature which is also found in our Sav1866-based model (open state) of CFTR. Similar features have also been observed by Zhang and colleagues [29].

## Discussion

CFTR belongs to a specific family of ABC transporters, namely ABC exporters, which share typical long intracellular loops (ICLs) providing critical contact with the NBDs, as shown by the recently elucidated structures of bacterial transporters. One of the most striking features of the ABC exporters, highlighted herein by the comparison of the Sav1866 3D structure (nucleotide-bound, outward-facing conformation) and the structures of the MsbA apo forms (nucleotide-free, inward-facing conformations), is the global invariability of the contacts between the ICL coupling helices and the NBDs (Fig. 2a). This is now also supported by the very recent X-ray structure at 3.8 Å of the mouse P-gp/MDR in an inward-facing configuration [22], where these ICLs/NBDs contacts appear to be conserved relatively to those observed in the MsbA (inward- and outward-facing) and Sav1866 (outward-facing) structures. Remarkably, the superimposition of the experimental P-gp/MDR structure with that of our Sav1866-based 3D model of CFTR shows that contacts similar to those we have previously described in CFTR between F508 in NBD1 and the ICL4 coupling helix [19] are also observed for the corresponding residues of mouse P-gp/MDR. In particular, the P-gp/MDR R908 residue, which corresponds to CFTR R1070, can make similar contacts with the main chain atoms of P-gp/MDR NBD1 residues, among which Y486 (corresponding to F508) (Fig. 6). Thus, this very recent structure of mouse P-gp (the first experimental structure of an eukaryotic ABC exporter with long ICLs), which shares with human CFTR ~11% sequence identity in its MSDs (a percentage similar to that observed between CFTR and Sav1866 or MsbA; Fig. 1), strongly supports our CFTR 3D model, especially its use for the prediction of functional features that can be experimentally validated.

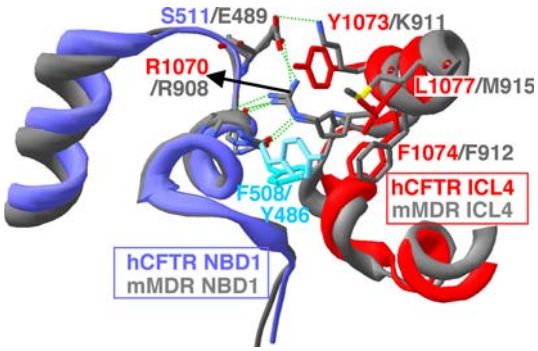


**Fig. 6** Comparison of our outward-facing Sav1866-based 3D model of human CFTR [19] with the experimental 3D structure of mouse P-gp/MDR in its inward-facing configuration [22]: superimposition of the NBD1/ICL4 interface. The 3D crystal structure of mouse P-gp/MDR (pdb 3g5u) was energetically refined in a similar way to our CFTR model. The human CFTR NBD1 and ICL4 structures are shown in *blue* and *red*, respectively, whereas the corresponding segments of mouse P-gp (MDR) are shown in *grey* (*ribbon representation*). In particular, critical bonds between the CFTR ILC4 residue R1070 (NE, NH1, and NH2 atoms) and the CFTR NBD1 residues I507 (CO)/F508 (CO)/S511 (OG1) may also be found in MDR between the side chain atoms of R908 (which corresponds to CFTR R1070) and R485 (CO)/Y486 (CO)/E489 (OE1). An additional bond can be observed in MDR between the ICL4 K911 and the NBD1 E489

Comparison of the outward- and inward-facing conformations of ABC transporters has indicated that the ATP-bound state is generally associated with the most extensive interface between the two NBDs (head-to-tail configuration), whereas the structures of nucleotide-free transporters exhibit conformations characterized by a greater separation between the NBDs [27]. In contrast to other ABC exporters which have symmetrical NBDs, an important particular feature of our human CFTR model is that, upon ATP hydrolysis, the sliding of the two CFTR NBDs along the heterodimer interface (which is coupled to the twisting motion of the MSDs) might be limited, and a complete separation of the NBDs may therefore be hindered, because of the presence of the CFTR-specific RI (green in Fig. 2) and that of an ATP molecule quasi-permanently bound at the non-conventional site (NBD1) and which likely maintains some contact at the NBD1:NBD2 interface. Indeed, by interacting in the inward-facing, closed-apo configuration with the modified LSHGH signature of NBD2 (ESM: Fig. S10), the RI might act as a "safety catch" limiting the NBD sliding motion and the associated twisting of the MSDs. This might constitute a mechanism slowing down the kinetics of the channel closing, in addition to that provided by the non-conventional poorly active ATP-binding site [55]. The presence of ATP at the non-conventional binding site (Fig. 2b) for several minutes, a time during which many open-closed gating cycles occur [1], may favor the association of NBD1 and NBD2 in a situation similar to that observed for the closed-apo conformer shown in Fig. 2b, and may thus also probably influence the channel kinetics. Altogether, the general asymmetry of the CFTR NBDs may allow a longer duration of the open state of the channel and consequently permit the passage of more anions during each gating cycle. It is tempting to speculate that the R domain, at least its globular-like R1 region predicted herein, might act in a similar way, behaving like a brake for the overall movement. Moreover, as shown in ESM: Fig. S10, one can also observe that, during the transition from the inward-facing (closed state) to the outward-facing, ATP-hydrolizable form (open state), the NBD2 domain may rub against the R1 and R2 sub-domains. In addition, the NBD2-specific patches of basic residues [in (1) the linker insertion (H1197, K1199, K1200), (2) the NBD2 main core (R1239, R1385, R1386, K1389, K1429, R1434, R1438) and (3) the C-terminal extension (K1457, K1459, K1461, K1468)] appear well positioned to interact with the R1 and R2 phosphorylated serine residues in the outward-facing conformer

(clearly rather than in the inward-facing one, ESM: Fig. S11). So, it is tempting to speculate that this putative phosphorylation-dependent interaction might favor the outward-facing form and thus the opening of the channel towards the extracellular milieu. Such a molecular mechanism may provide an explanation of the role of phosphorylation in the activation of CFTR (discussed in [56]). This hypothesis is supported by the conclusions of the studies (1) of Rich and colleagues [57], who suggested interactions of the R domain with NBD2, and (2) of Chappe and colleagues [58], who showed that phosphorylation of the R domain regulates the CFTR function by promoting its association with other CFTR domains, rather than by causing its dissociation from an inhibitory site. These considerations also appear to be supported by a recent review [59], highlighting the potential role of MSDs and substrates (also named allocrites) in driving conformational changes, rather than most directly ATP hydrolysis. Here, in the case of the CFTR channel, the size of small anions may not be sufficient for driving large conformational changes, which might thus be favored by additional elements such as the regulatory insertion (RI) and the R domain.

The rubbing of the R1/R2 sub-domains with NBD2 during inter-conversion also appears consistent with their content in strongly hydrophobic amino acids (ESM: Table 1) which is intermediate between fully disordered and canonical, folded globular domains. Indeed, the flexibility allowed thereby appears important for the fine-tuning of the concerted movements of the NBDs with the MSDs.

Finally, the present model of the inward-facing conformation of CFTR may allow some novel insights into the functioning of the ABC transporter CFTR as an anion channel. Indeed, it appears that the region around F337 (Fig. 3) may constitute a constriction of the channel, which, with the eventual help of the extracellular loops and their associated glycans, might form the structural basis of a gate allowing the channel to close. Conversely, our whole CFTR model does not provide an obvious structural basis for a second gate, the existence of which remains thus elusive. Accordingly, as already suggested by others [6], the cytoplasmic-side gate of CFTR may have become atrophied, or uncoupled from the outer gate, and the CFTR channel may thus function as a "broken" ABC transporter.

Here, it is very helpful to analyse the MsbA and Sav1866 structures to get insights into the specificities of CFTR. Indeed, for MsbA and Sav1866 in their outward-facing conformations, it can be observed that the four helices of the bundle formed by the ICLs around the vertical two-fold axis and contacting the top surface of the NBDs, are too close to each other to allow the passage of molecular substrates from the cytosol. In particular, the accurate Sav1866 structure clearly shows that (1) the two symmetric ICL2 histidine (H204) residues (corresponding in CFTR to ICL2 S263 and ICL4 V1056) definitely lock the space, around the two-fold axis, between the ICLs, and (2) that this inner region cannot be accessed from the cytosolic solvent. In contrast, as shown in Fig. 7 and ESM: Fig. S12, the situation appears to be quite different for the Sav1866-based model of CFTR [19] which takes into account the sequence specificity of CFTR in that very same region. First, the opening of the ICL bundle along the two-fold axis appears sufficiently large to allow chloride ions to pass through a hydrophobic ring between the ICLs (Fig. 7a). Next, as shown in Fig. 7b, a long, linear funnel (not observed in Sav1866) appears to cross the structure from one side to another, perpendicularly to the vertical pseudo two-fold axis, and may thus constitute a direct access from the cytosolic solvent to the putative channel pore running along the pseudo two-fold axis. These CFTR-specific features thus support the notion that the outward-facing conformation may well correspond to the open channel, able to transport a relatively large number of small anions. By the way, the four-helix bundle shown in Fig. 7a is reminiscent of that of the M2 proton channel [60], which is essential for Influenza infection and is also the binding site of anti-viral molecules. It should be stressed here that, at the cytoplasmic bottom of the CFTR ICLs channel and below the funnel discussed above, the side chain of ICL2 E267 clearly locks the channel pore and weakly interacts with its symmetrical residue in ICL4 (K1060) (Fig. 7a). As aspartic or glutamic residues are rare in the putative channel, such an observation may be of importance. As illustrated in ESM: Fig. S12C, E267 might, however, jump to another conformer, where it may weakly interact with the side chain of K968, making then the bottom end of the channel accessible from the NBD1:NBD2 interface around the pseudo two-fold axis. This is illustrated in ESM: Fig. S12A, B. The side chain of E267 might thus act as a potential fast gate, which could swing out of the ion pathway, similarly to what is observed in the ClC family [6, 61]. It is, however, tempting to speculate that the funnel described above may constitute a more direct way for small anions, trafficking from (or to) the cytoplasmic solvent, to access the channel pore, which is thus likely to start at the level of the ICLs bundle along the pseudo two-fold axis. According to that hypothesis, the aforementioned hydrophobic ring present in the ICLs bundle (Fig. 7) might function as a "pseudo" gate. It is interesting to note that hydrophobic girdles acting as gates do actually exist in many channels, such as, for example, the nicotinic ACh receptor and $K^+$ selective channel [62]. However, in the case of CFTR, the hydrophobic ring may still allow the passage of a chloride ion from beneath the NBDs, when the glutamic acid E267 swings out of the central conduction pathway.

In conclusion, our results suggest that CFTR, as every ABC transporter, uses for its gating cycles a mechanism

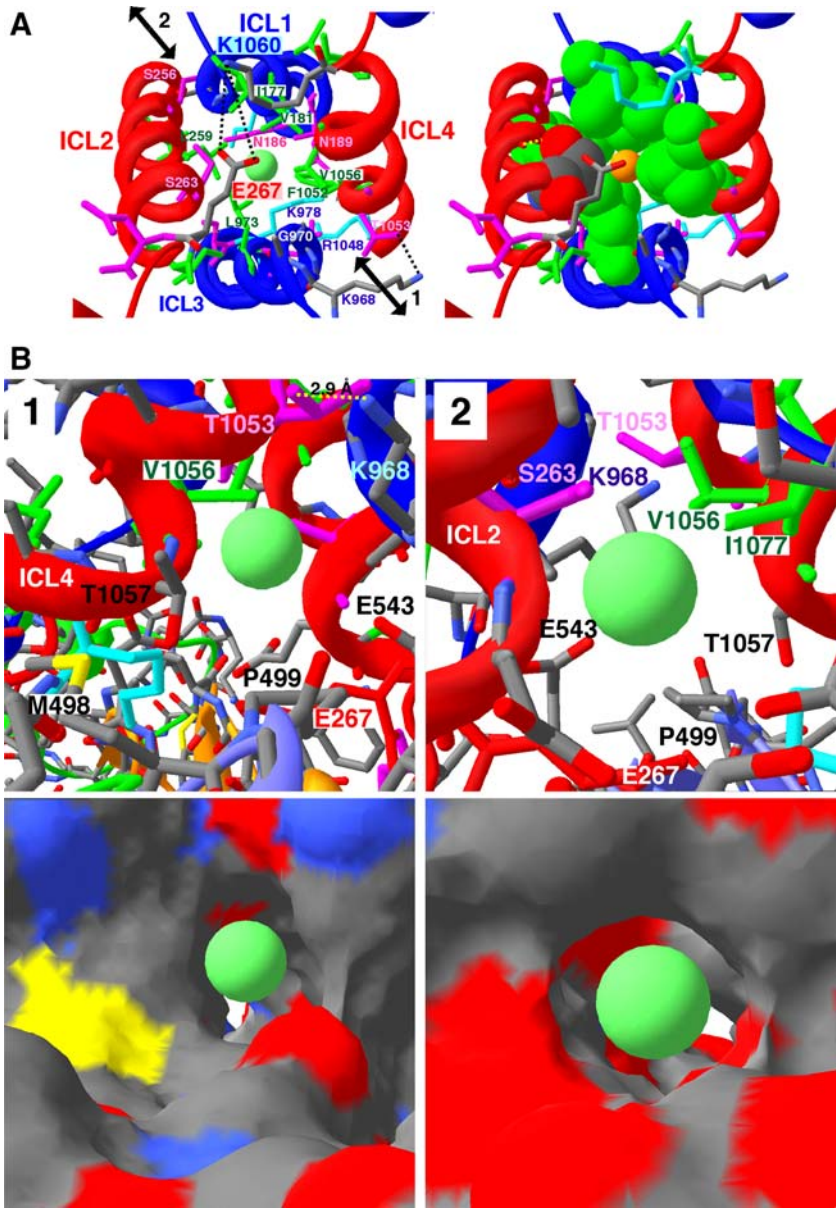


**Fig. 7** A putative cytoplasmic entry to the CFTR channel. **a** Views of the ICLs bundle region seen from the NBD1:NBD2 interface (not shown for sake of clarity) and along the pseudo two-fold axis, with E267 in the foreground. The side chain of this residue may lock the putative CFTR channel at its basis. E267 weakly interacts with its symmetrical amino acid K1060 (distance between the two amino acids ~4.8 Å), nearly at the centre of the pore which has a diameter of 5–8 Å, a size at least equal to (or greater than) that of the selectivity filter of the ClC channel [66]. Lysine and arginine residues are shown in *blue*, hydrophobic residues in *green*, and hydrophilic ones in *pink*. However, as illustrated in ESM: Fig. S11, E267 might jump to another conformer, where it may weakly interact with the side chain of K968. Chloride ions may thus either freely run from the funnel (*arrows 1* and *2*), described in more details in **b**, to the open CFTR channel (and vice versa), or alternatively access with more difficulties the CFTR channel through the ICLs bottom, when this access is not locked. *Left* a chloride ion (*green*) is shown at the level of a hydrophobic ring (I177, V181, L259, S263, G970, L973, F1052, V1056). One can note that the side chain oxygen of S263 binds to the V260 (not indicated) main chain carbonyl, as it is often the case within helices, and thus only exposes its C$\beta$ atom to the channel. The hydrophobic ring is followed by a more hydrophilic one (including S185, N186, K190, N974, S977, K978, K1040), where the channel becomes larger. *Right* hydrophobic surfaces shown in *green* with the chloride ion in *yellow* and S263 in standard colors (oxygen in *red*, carbon in *grey*). **b** Symmetrical views (*1* and *2*) of the putative access [for chloride ions coming from (or moving to) the cytoplasmic milieu] to the beginning of the putative CFTR channel pore. This access, which is located between the helices of the ICLs bundle, consists of an almost linear horizontal funnel (~20 Å long) running at the basis of ICLs and nearly perpendicular to the vertical pseudo two-fold axis of CFTR. The directions of views *1* and *2* are indicated by *arrows* in (**a**). The *top* views show the atomic environment of the funnel, the *bottom* views the corresponding molecular surfaces (*red* oxygen, *blue* nitrogen, *yellow* sulfur, *grey* carbon atoms). A chloride ion (*green ball*) was positioned at both entrances of the funnel. In both *bottom* views, the small *white* areas near the center correspond to the protein exterior at the other side of the funnel

alternating between outward- and inward-facing conformations. However, our model of the outward-facing conformation also reveals the presence, at the level of the MSDs intracellular loops, of an aperture which is sufficiently large to allow the passage of chloride ions; this clearly supports the channel function of CFTR and the assumption that the outward-facing conformation corresponds to the open state of the CFTR channel. The transition from that conformer towards an inward-facing conformation finally leads to the closure of the channel. Indeed, the structural features described above for inward-facing conformation of CFTR, modeled on the "closed-apo" conformer of MsbA, which is characterized by a very small separation of the NBDs, should allow the closure of the channel. However, for CFTR to function as a more classic ABC protein transporting much larger substrates than small ions, it may be expected that it will have to reach an inward-facing configuration similar to the "open-apo" structure described for MsbA, with the two NBDs now fully separated. Such a configuration is otherwise observed for the inward-facing form of P-gp [22]. Remarkably, the concerted movements between the transmembrane helices constituting the channel and the NBDs apparently occur around constant pivots corresponding to the transmission interfaces linking the ICLs coupling helices and the NBDs. Our results strongly emphasize the critical role of the residue F508 at these interfaces, not only for CFTR domain assembly but also for the signaling between the NBDs and MSDs. Finally, the models presented here are in good agreement with various experimental data, including those recently observed by electron microscopy, which support the predicted overall assembly of CFTR dimeric arrangement and the possible location of the R domain, partly below the NBD heterodimer. Moreover, the location of the N-terminal and C-terminal ends of CFTR, close to the R domain, is in good agreement with experimental data suggesting the existence of a physical interaction between them [63, 64]. Although some of the structural features of CFTR predicted herein as well as the above speculations about the CFTR functioning await experimental validation, our models provide helpful support for understanding of how CFTR functions and how mutations lead to dysfunctions, as well as for the rational design of drugs for a CFTR-specific pharmacological treatment of CF.

**Acknowledgments** This work was supported by grants from the association "Vaincre La Mucoviscidose" (Paris, France).

## References

1. Gadsby DC, Vergani P, Csanády L (2006) The ABC protein turned chloride channel whose failure causes cystic fibrosis. Nature 440:477–483
2. Rowe SM, Miller S, Sorscher EJ (2005) Cystic fibrosis. N Engl J Med 352:1992–2001
3. Biemans-Oldehinkel E, Doeven MK, Poolman B (2006) ABC transporter architecture and regulatory roles of accessory domains. FEBS Lett 580:1023–1035
4. Anderson MP, Gregory RJ, Thompson S, Souza DW, Paul S, Mulligan RC, Smith AE, Welsh MJ (1991) Demonstration that CFTR is a chloride channel by alteration of its anion selectivity. Science 253:202–205
5. Bear CE, Li CH, Kartner N, Bridges RJ, Jensen TJ, Ramjeesingh M, Riordan JR (1992) Purification and functional reconstitution of the cystic fibrosis transmembrane conductance regulator (CFTR). Cell 68:809–818
6. Gadsby DC (2009) Ion channels versus ion pumps: the principal difference, in principle. Nat Rev Mol Cell Biol 10:344–352
7. Cheng SH, Rich DP, Marshall J, Gregory RJ, Welsh MJ, Smith AE (1991) Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel. Cell 66:1027–1036
8. Ostedgaard LS, Baldursson O, Welsh MJ (2001) Regulation of the cystic fibrosis transmembrane conductance regulator $Cl^-$ channel by its R domain. J Biol Chem 276:7689–7692
9. Cheng SH, Gregory RJ, Marshall J, Paul S, Souza DW, White GA, O'Riordan CR, Smith AE (1990) Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. Cell 63:827–834
10. Riordan JR (1999) Cystic fibrosis as a disease of misprocessing of the cystic fibrosis transmembrane conductance regulator glycoprotein. Am J Hum Genet 64:1499–1504
11. Ward CL, Omura S, Kopito RR (1995) Degradation of CFTR by the ubiquitin–proteasome pathway. Cell 83:121–127
12. Callebaut I, Eudes R, Mornon JP, Lehn P (2004) Nucleotide-binding domains of human cystic fibrosis transmembrane conductance regulator: detailed sequence analysis and three-dimensional modeling of the heterodimer. Cell Mol Life Sci 61:230–242
13. Eudes R, Lehn P, Férec C, Mornon J-P, Callebaut I (2005) Nucleotide binding domains of human CFTR: a structural classification of critical residues and disease-causing mutations. Cell Mol Life Sci 62:2112–2123
14. Lewis HA, Buchanan SG, Burley SK, Conners K, Dickey M, Dorwart M, Fowler R, Gao X, Guggino WB, Hendrickson WA, Hunt J, Kearins MC, Lorimer D, Maloney PC, Post KW, Rajashankar KR, Rutter ME, Sauder JM, Shriver S, Thibodeau PH, Thomas PJ, Zhang M, Zhao X, Emtage S (2004) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. EMBO J 23:282–293
15. Lewis HA, Zhao X, Wang C, Sauder JM, Rooney I, Noland BW, Lorimer D, Kearins MC, Conners K, Condon B, Maloney PC, Guggino WB, Hunt JF, Emtage S (2005) Impact of the deltaF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. J Biol Chem 280:1346–1353
16. Cui L, Aleksandrov L, Chang XB, Hou YX, He L, Hegedus T, Gentzsch M, Aleksandrov A, Balch WE, Riordan JR (2007) Domain interdependence in the biosynthetic assembly of CFTR. J Mol Biol 365:981–994
17. Du K, Sharma M, Lukacs GL (2005) The DeltaF508 cystic fibrosis mutation impairs domain–domain interactions and arrests post-translational folding of CFTR. Nat Struct Mol Biol 12:17–25
18. Dawson RJ, Locher KP (2007) Structure of the multidrug ABC transporter Sav1866 from *Staphylococcus aureus* in complex with AMP–PNP. FEBS Lett 581:935–938
19. Mornon J-P, Lehn P, Callebaut I (2008) Atomic model of human cystic fibrosis transmembrane conductance regulator: membrane-spanning domains and coupling interfaces. Cell Mol Life Sci 65:2594–2612

20. Serohijos AW, Hegedus T, Aleksandrov AA, He L, Cui L, Dokholyan NV, Riordan JR (2008) Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function. Proc Natl Acad Sci USA 105:3256–3261
21. Loo TW, Bartlett MC, Clarke DM (2008) Processing mutations disrupt interactions between the nucleotide binding and transmembrane domains of P-glycoprotein and the cystic fibrosis transmembrane conductance regulator (CFTR). J Biol Chem 283:28190–28197
22. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL, Chang G (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323:1718–1722
23. Krasnov KV, Tzetis M, Cheng J, Guggino WB, Cutting GR (2008) Localization studies of rare missense mutations in cystic fibrosis transmembrane conductance regulator (CFTR) facilitate interpretation of genotype-phenotype relationships. Hum Mut 29:1364–1372
24. Seibert FS, Linsdell P, Loo TW, Hanrahan JW, Clarke DM, Riordan JR (1996) Disease-associated mutations in the fourth cytoplasmic loop of cystic fibrosis transmembrane conductance regulator compromise biosynthetic processing and chloride channel activity. J Biol Chem 271:15139–15145
25. Ward A, Reyes CL, Yu J, Roth CB, Chang G (2007) Flexibility in the ABC transporter MsbA: alternating access with a twist. Proc Natl Aca Sci USA 104:19005–19010
26. Oldham ML, Davidson AL, Chen J (2008) Structural insights into ABC transporter mechanism. Curr Opin Struct Biol 18:726–733
27. Rees DC, Johnson E, Lewinson O (2009) ABC transporters: the power to change. Nat Rev Mol Cell Biol 10:218–227
28. Mio K, Ogura T, Mio M, Shimizu H, Hwang T, Sato C, Sohma Y (2008) Three-dimensional reconstruction of human cystic fibrosis transmembrane conductance regulator chloride channel revealed an ellipsoidal structure with orifices beneath the putative transmembrane domain. J Biol Chem 283:30300–30310
29. Zhang L, Aleksandrov LA, Zhao Z, Birtley JR, Riordan JR, Ford RC (2009) Architecture of the cystic fibrosis transmembrane conductance regulator protein and structural changes associated with phosphorylation and nucleotide binding. J Struct Biol (in press)
30. Rosenberg MF, Kamis AB, Aleksandrov LA, Ford RC, Riordan JR (2004) Purification and crystallization of the cystic fibrosis transmembrane conductance regulator (CFTR). J Biol Chem 279:39051–39057
31. Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291–325
32. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723
33. Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon JP (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci 53:621–645
34. Eudes R, Le Tuan K, Delettré J, Mornon JP, Callebaut I (2007) A generalized analysis of hydrophobic and loop clusters within globular protein sequences. BMC Struct Biol 7:2
35. Gaboriaud C, Bissery V, Benchetrit T, Mornon JP (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett 224:149–155
36. Chen J, Lu G, Lin J, Davidson AL, Quiocho FA (2003) A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. Mol Cell 12:651–661
37. Fatehi M, Linsdell P (2009) Novel residues lining the CFTR chloride channel pore identified by functional modification of introduced cysteines. J Membr Biol 228:151–164
38. Guinamard R, Akabas MH (1999) Arg352 is a major determinant of charge selectivity in the cystic fibrosis transmembrane conductance regulator chloride channel. Biochemistry 38:5528–5537
39. Linsdell P, Zheng SX, Hanrahan JW (1998) Non-pore lining amino acid side chains influence anion selectivity of the human CFTR $Cl^-$ channel expressed in mammalian cell lines. J Physiol 512:1–16
40. McDonough S, Davidson N, Lester HA, McCarty NA (1994) Novel pore-lining residues in CFTR that govern permeation and open-channel block. Neuron 13:623–634
41. Gong X, Linsdell P (2003) Molecular determinants and role of an anion binding site in the external mouth of the CFTR chloride channel pore. J Physiol 549:387–397
42. Zhou JJ, Fatehi M, Linsdell P (2008) Identification of positive charges situated at the outer mouth of the CFTR chloride channel pore. Pflugers Arch 457:351–360
43. Ge N, Muise CN, Gong X, Linsdell P (2004) Direct comparison of the functional roles played by different transmembrane regions in the cystic fibrosis transmembrane conductance regulator chloride channel pore. J Biol Chem 279:55283–55289
44. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL et al (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245:1066–1073
45. St Aubin CN, Linsdell P (2006) Positive charges at the intracellular mouth of the pore regulate anion conduction in the CFTR chloride channel. J Gen Physiol 128:535–545
46. St Aubin CN, Zhou JJ, Linsdell P (2007) Identification of a second blocker binding site at the cytoplasmic mouth of the cystic fibrosis transmembrane conductance regulator chloride channel pore. Mol Pharmacol 71:1360–1368
47. Baker JM, Hudson RP, Kanelis V, Choy WY, Thibodeau PH, Thomas PJ, Forman-Kay JD (2007) CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nat Struct Mol Biol 14:738–745
48. Hegedus T, Serohijos AW, Dokholyan NV, He L, Riordan JR (2008) Computational studies reveal phosphorylation-dependent changes in the unstructured R domain of CFTR. J Mol Biol 378:1052–1063
49. Gros P, Croop J, Housman D (1986) Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. Cell 47:371–380
50. Diederichs K, Diez J, Greller G, Müller C, Breed J, Schnell C, Vonrhein C, Boos W, Welte W (2000) Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon *Thermococcus litoralis*. EMBO J 19:5951–5961
51. Hollenstein K, Dawson JP, Locker KP (2007) Structure and mechanism of ABC transporter proteins. Curr Opin Struct Biol 17:412–418
52. Koonin EV, Wolf Y, Aravind L (2000) Protein fold recognition using sequence profiles and its application in structural genomics. Adv Protein Chem 54:245–275
53. Verdon G, Albers SV, Dijkstra BW, Driessen AJ, Thunnissen AM (2003) Crystal structures of the ATPase subunit of the glucose ABC transporter from *Sulfolobus solfataricus*: nucleotide-free and nucleotide-bound conformations. J Mol Biol 330:343–358
54. Awayn NH, Rosenberg MF, Kamis AB, Aleksandrov LA, Riordan JR, Ford RC (2005) Crystallographic and single-particle analyses of native- and nucleotide-bound forms of the cystic fibrosis transmembrane conductance regulator (CFTR) protein. Biochem Soc Trans 33:996–999

55. Zhou Z, Wang X, Liu HY, Zou X, Li M, Hwang TC (2006) The two ATP binding sites of cystic fibrosis transmembrane conductance regulator (CFTR) play distinct roles in gating kinetics and energetics. J Gen Physiol 128:413–422
56. Hegedűs T, Aleksandrov A, Mengos A, Cui L, Jensen TJ, Riordan JR (2009) Role of individual R domain phosphorylation sites in CFTR regulation by protein kinase A. Biochim Biophys Acta 1788:1341–1349
57. Rich DP, Gregory RJ, Anderson MP, Manavalan P, Smith AE, Welsh MJ (1991) Effect of deleting the R domain on CFTR-generated chloride channels. Science 253:205–207
58. Chappe V, Irvine T, Liao J, Evagelidis A, Hanrahan JW (2005) Phosphorylation of CFTR by PKA promotes binding of the regulatory domain. EMBO J 24:2730–2740
59. Kos V, Ford RC (2009) The ATP-binding cassette family: a structural perspective. Cell Mol Life Sci (in press)
60. Nishimura K, Kim S, Zhang L, Cross T (2002) The closed state of a $H^+$ channel helical bundle combining precise orientational and distance restraints from solid state NMR. Biochemistry 41:13170–13177
61. Dutzler R, Campbell EB, MacKinnon R (2003) Gating the selectivity filter in ClC chloride channels. Science 300:108–112
62. Doyle DA (2004) Structural changes during ion channel gating. Trends Neurosci 27:298–302
63. Fu J, Kirk KL (2001) Cysteine substitutions reveal dual functions of the amino-terminal tail in cystic fibrosis transmembrane conductance regulator channel gating. J Biol Chem 276:35660–35668
64. Naren AP, Cormet-Boyaka E, Fu J, Villain M, Blalock JE, Quick MW, Kirk KL (1999) CFTR chloride channel regulation by an interdomain interaction. Science 286:544–548
65. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: multiple sequence alignments in PostScript. Bioinformatics 15:305–308
66. Lobet S, Dutzler R (2006) Ion-binding properties of the ClC chloride selectivity filter. EMBO J 25:24–33